

**AN EVALUATION OF THE CORRECTIONAL TRAINING PROGRAM
DELIVERED BY THE CORRECTIONAL SERVICE OF CANADA**

by

Shevaun Corey

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Psychology

Carleton University
Ottawa, Canada

© 2012
Shevaun Corey



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-494-93659-7

Our file Notre référence

ISBN: 978-0-494-93659-7

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

Despite the need for training evaluation, many organizations do not adequately evaluate their training programs. When organizations do evaluate their training, they often rely on Kirkpatrick's (1959a) four-level evaluation framework, which focuses on trainee reactions, learning, behavioural outcomes, and results for the organization. This is one of the most widely used evaluation frameworks despite the fact that it has many limitations, one of the most serious being its narrow focus on just a few training outcomes at each level of the framework. An expanded version of this framework was used to evaluate the Correctional Training Program (CTP), which attempts to provide correctional officers who might be hired by the Correctional Service of Canada (CSC) with the knowledge, skills, and attitudes necessary to effectively manage offenders within institutions. Three specific issues were examined: (1) Does CTP produce positive changes in trainees (e.g., with respect to domain-relevant knowledge, skills and attitudes)?; (2) Do the training outcomes measured at each level of the expanded framework correlate with one another?; and (3) Do barriers to training effectiveness (e.g., lack of supervisor support) moderate the relationships between the training levels? By analyzing the data from 102 CTP trainees, valuable information was obtained concerning the effectiveness of CTP with respect to a wide variety of trainee reactions, learning gains, and the degree to which knowledge and skills transferred to the job. However, the results from this study do not suggest that the training outcomes measured at each level of the expanded framework correlate with one another, nor do they suggest that the training effectiveness factors, as measured in the current study, significantly moderate training outcomes, with the exception of content validity and self-efficacy. The theoretical and practical implications

of these results are discussed, along with a discussion of the suitability of an expanded version of Kirkpatrick's four-level evaluation framework. Recommendations for CSC as well as directions for future research are also suggested.

Acknowledgements

I once read that a mentor is someone who sees more talent and ability within you than you see in yourself, and that they help bring it out in you. Craig, you have certainly been a mentor to me over the past eight years and I would like to thank you for your unwavering support and confidence in my academic pursuits. I have learned many things from you that I will take forward into my own career, such as your objectivity, your dedication to your students, and your pragmatism. I feel so very fortunate to have had the opportunity to work with you and learn from you. My graduate career has been an amazing experience in large part because of you.

I would like to thank all of my committee members for their time and thoughtful insights – Drs. Patti Phillips, Ralph Serin, Shelley Brown, and George Pollard. I would like to express my sincerest gratitude to the staff of the Learning & Development Branch and the Regional Staff Colleges of the Correctional Service of Canada for their assistance, support and encouragement, as well as all of the correctional officer participants. A very special thank you to the members of my evaluation team – Steve Loeb, Jean-Luc Chamailard, Carl Jalbert, Julia Crosier, and Jacques Brien – your expertise, time and support have been invaluable. Also, funding from the Social Science and Humanities Research Council has helped to make this research possible.

For the countless hours spent listening, collaborating with and encouraging me, I must thank my colleagues past and present of the Police Research Lab, as well as Natalie Pressburger, Etelle Bourassa, April McGrath, and Clare Wilkinson. I am grateful to my friends Erin Crawley and John Edmunds who took care of my boys, Jax and Loki, during

my many trips for data collection. For their endless love and support, I have to thank my parents, who have always believed in me. I share this success with you both.

Lastly, but certainly not least, to Marc-Antoine, whose love has kept me going over the past few years. Thank you for your encouragement and support, your patience, and for putting my needs ahead of your own. The boys thank you too for playing with them when I was too busy... Here's to the future and to putting us first!

“Success is not final, failure is not fatal: it is the courage to continue that counts.”

Winston Churchill

Table of Contents

Abstract.....	ii
Acknowledgements	iv
Table of Contents.....	vi
List of Tables.....	xi
List of Figures.....	xv
List of Appendices.....	xvi
Chapter 1. Introduction	1
The Focus of the Research.....	1
A Multi-Level Framework for Training Evaluation.....	4
A Useful Starting Point: Kirkpatrick’s Four-Level Framework.....	5
Correlations Between the Levels of Kirkpatrick’s Framework	8
Potential Reasons for the Lack of Significant Correlations	10
The Need for a Multi-Dimensional, Multi-Level Training Evaluation Framework	10
The Uni-dimensionality of Kirkpatrick’s Levels.....	11
Expanding the Levels: Making Kirkpatrick’s Framework More Psychologically Sound.....	12
Level 1	12
Level 2	14
Level 3	16
The Need to Consider Training Effectiveness Factors.....	19
So Why Does Training Effectiveness Matter?	22
Holton’s Learning Transfer System Inventory	23

The Current Study	29
Hypotheses.....	30
Chapter 2. Methodology.....	33
Participants	35
Measures.....	38
Informed Consent Forms	38
Demographic Questionnaire	38
Evaluation Tools.....	39
Level 1 Measure: Trainee reaction questionnaire.....	40
Level 2 Measures.....	42
Knowledge.....	43
Skills	45
Attitudes.....	51
Level 3 Measures: Trainee and supervisor checklists	53
Factors Influencing Training Effectiveness	55
Debriefing Forms	57
Procedure	57
Chapter 3. Data Screening and Attrition Analyses	61
Data Screening.....	61
Missing values	61
Outliers and normality	63
Attrition Analyses.....	68
Attrition between the recruitment and training phases.....	68

Attrition between the training and transfer phases	70
Chapter 4. An Analysis of Training Outcomes Included in the Expanded Evaluation Framework	72
Hypotheses.....	72
Level 1 Outcome.....	73
Qualitative results	79
Level 2 Outcomes.....	79
Knowledge.....	85
Skills.....	91
Soft skills	91
Hard skills.....	99
Attitudes.....	107
ATCW	107
ATI.....	111
Level 3 Outcomes.....	112
Trainee Behavioural Checklists.....	113
Supervisor Behavioural Checklists.....	122
Chapter Summary	130
Chapter 5. Correlations Between Each Level of the Evaluation Framework	133
Hypotheses.....	133
Level 1 & 2	134
Level 1 & 3	141
Level 2 & 3	144

Chapter Summary	148
Chapter 6. The Influence of Training Effectiveness Factors	150
Hypotheses.....	151
LTSI.....	152
Moderation Effects	159
Supervisor support.....	162
Resistance-Openness to Change.....	169
Content Validity	175
Self-efficacy.....	182
Transfer Design	189
Chapter Summary	194
Chapter 7. Discussion	197
Training Outcomes	198
Level 1	198
Level 2	200
Level 3	209
Correlations Between the Levels	214
Training Effectiveness Factors	220
Summary of Questions Addressed in the Current Study	229
CTP-Specific Recommendations.....	233
General Recommendations for Training Evaluations	235
Additional Study Limitations and Directions for Future Research	236
References	240

Appendices256

List of Tables

Table 1: Kirkpatrick's (1994) Four-Level Framework	6
Table 2: The Transfer of Training Model by Baldwin and Ford (1988)	20
Table 3: Description of the Factors Included in the LTSI (Holton et al., 2000)	25
Table 4: A Brief Outline of the Four Stages of CTP	34
Table 5: Participation Numbers per Region and Study Phase	36
Table 6: Demographic Variables of Participants	37
Table 7: Outline of the Evaluation Measures Used in the Current Study, Categorized According to Level of Evaluation	40
Table 8: Outline of the Evaluation Measures Categorized by Study Phase	40
Table 9: Overview of Outlying Variables	64
Table 10: Significantly Skewed and Kurtotic Variables	66
Table 11: Attrition Between the Study Phases	70
Table 12: Correlation Matrix for the Level 1 Trainee Reaction Outcomes (Including Subsections)	76
Table 13: Descriptive Statistics for the Trainee Reaction Questionnaire	77
Table 14: Correlation Matrix for the Level 2 Outcome Measures During the Recruitment Phase	82
Table 15: Correlation Matrix for the Level 2 Outcome Measures During the Training Phase	83
Table 16: Correlation Matrix for the Level 2 Outcome Measures During the Transfer Phase	84
Table 17: Descriptive Statistics for the Knowledge Outcomes (% Correct)	86

Table 18: Descriptive Statistics for the Soft Skills Measures	93
Table 19: Descriptive Statistics for the Hard Skills Composite Checklist Score	101
Table 20: Descriptive Statistics for the Eight Individual Hard Skills Checklist Scores	102
Table 21: Descriptive Statistics for the Eight Individual Hard Skills Checklist Scores for the Pacific and Ontario Regions Combined ($n = 50$)	107
Table 22: Descriptive Statistics for the Attitudes Measures	109
Table 23: Descriptive Statistics for the Trainee Behavioural Checklist - Frequency Items	114
Table 24: Descriptive Statistics for the Trainee Behavioural Checklist - Quality Items	119
Table 25: Descriptive Statistics for the Supervisor Behavioural Checklist - Frequency Items	124
Table 26: Descriptive Statistics for the Supervisor Behavioural Checklist - Quality Items	127
Table 27: Correlation Matrix for the Level 1 Outcome (Training Phase) and Level 2 Outcomes (Recruitment Phase)	135
Table 28: Correlation Matrix for the Level 1 Outcome (Training Phase) and Level 2 Outcomes (Training Phase)	137
Table 29: Correlation Matrix for the Level 1 Outcome (Training Phase) and Level 2 Outcomes (Transfer Phase)	139
Table 30: Correlation Matrix for the Level 1 and 3 Outcome Measures	143
Table 31: Correlation Matrix for the Level 2 and 3 Outcome Measures	146
Table 32: Descriptive Statistics and Internal Consistency Scores for the Subscales of the	

LTSI ($n = 53$)	155
Table 33: Correlation Matrix for the Training Effectiveness Factors and the Training Outcomes (Levels 1-3)	157
Table 34: Training Effectiveness Variables Included in the MMR Analyses	160
Table 35: Proposed Hierarchical Regression Analyses to Examine the Potential Moderating Effect of Supervisor Support	163
Table 36: Summary of Hierarchical Multiple Regression Analyses for the Moderator of Supervisor Support	165
Table 37: Regression Coefficients for Each Effect and Their Associated Confidence Intervals and <i>T</i> -Tests – Supervisor Support	166
Table 38: Proposed Hierarchical Regression Analyses for the Moderation Effect of Openness to Change	170
Table 39: Summary of Hierarchical Multiple Regression Analyses for the Moderator of Resistance-Openness to Change	171
Table 40: Regression Coefficients for Each Effect and Their Associated Confidence Intervals and <i>T</i> -Tests – Resistance-Openness to Change	172
Table 41: Proposed Hierarchical Regression Analyses for the Moderation Effect of Content Validity	176
Table 42: Summary of Hierarchical Multiple Regression Analyses for the Moderator of Content Validity	177
Table 43: Regression Coefficients for Each Effect and Their Associated Confidence Intervals and <i>T</i> -Tests – Content Validity	178
Table 44: Summary of the Results of a Simple Slopes Analysis Showing the Relationship	

Between Knowledge Test Scores at the Training and Transfer Phases as a Function of Content Validity Ratings	181
Table 45: Proposed Hierarchical Regression Analyses for the Moderation Effect of Self- Efficacy	183
Table 46: Summary of Hierarchical Multiple Regression Analyses for the Moderator of Self-Efficacy	184
Table 47: Regression Coefficients for Each Effect and Their Associated Confidence Intervals and <i>T</i> -Tests – Self-Efficacy	185
Table 48: Summary of the Results of a Simple Slopes Analysis Showing the Relationship Between Knowledge Test Scores at the Training and Transfer Phases as a Function of Self-Efficacy Ratings	188
Table 49: Proposed Hierarchical Regression Analyses for the Moderation Effect of Transfer Design	190
Table 50: Summary of Hierarchical Multiple Regression Analyses for the Moderator of Transfer Design	191
Table 51: Regression Coefficients for Each Effect and Their Associated Confidence Intervals and <i>T</i> -Tests – Transfer Design	192

List of Figures

Figure 1: Plot illustrating the knowledge outcomes for the test subsections across study phases.....	91
Figure 2: Plot illustrating the soft skills outcomes across study phases per region.....	96
Figure 3a: Plot illustrating the soft skills outcomes for the subsections across study phases for the Ontario region.....	97
Figure 3b: Plot illustrating the soft skills outcomes for the subsections across study phases for the Pacific region.....	98
Figure 4: Plot of the significant interaction effect of Study Phase x Skill	105
Figure 5: Plot of the attitudes outcomes across study phases.....	111
Figure 6: Results of the simple slopes analysis, displaying the relationship between knowledge test scores at the training and transfer phases as a function of content validity ratings	180
Figure 7: Results of the simple slopes analysis, displaying the relationship between knowledge test scores at the training and transfer phases as a function of self-efficacy ratings	187

List of Appendices

Appendix A: Overview of courses and subject areas covered in Stage 1 including an indication of the number of formal knowledge-based evaluations.....	256
Appendix B: Overview of the courses and subject areas covered in Stage 3 including an indication of the number of formal knowledge and skills-based evaluations	257
Appendix C: A List of Topics Included in the Stage 4 Institutional Orientation Checklists.....	259
Appendix D: Participation Rates per Region and CTP Cohort for Each Phase of the Study.....	260
Appendix E: Demographic Data for the Training and Transfer Phases	261
Appendix F: INFORMED CONSENT FORM (For the Trainees of CTP)	265
Appendix G: INFORMED CONSENT FORM (For the Supervisors)	268
Appendix H: Demographic Questionnaire	270
Appendix I: Completion Rates of all Evaluation Tools by Participants overall ($N = 102$) and per Region, Presented for each Phase of the Study	272
Appendix J: Level 1 Reaction Evaluation Questionnaire Given to Trainees at the End of Stage 3	273
Appendix K: Hard Skills Selection Questionnaire	278
Appendix L: Definition of the Hard Skills Assessed in the Current Study	283
Appendix M: Hard Skills Scenarios	285
Appendix N: Attitudes Toward Correctional Work Scale (Pre-Training)	290
Appendix O: Attitudes Toward Inmates Scale	292
Appendix P: Attitudes Toward Correctional Work Scale (Four Months Post-	

Training)	294
Appendix Q: Trainee Behavioural Checklist for CTP 2008	296
Appendix R: Supervisor Behavioural Checklist for CTP 2008.....	300
Appendix S: DEBRIEFING FORM (For Trainees).....	304
Appendix T: DEBRIEFING FORM (For Supervisors)	306
Appendix U: Data Collection Schedule.....	308
Appendix V: Descriptive Statistics per Item of the Trainee Reaction Questionnaire Ontario & Pacific Regions Combined	310
Appendix W: Descriptive Statistics per Item of the Trainee Reaction Questionnaire for the Pacific Region	314
Appendix X: Descriptive Statistics per Item of the Trainee Reaction Questionnaire for the Ontario Region	318
Appendix Y: Qualitative Analysis of Trainee Reaction Comments.....	322
Appendix Z: Descriptive Statistics for the Level 2 and 3 Training Outcomes as well as the Training Effectiveness Factors for the Quebec Region.....	331
Appendix AA: Correlation Matrices for the Individual Hard Skills	337
Appendix BB: Responses to the Open-Ended Items from the Trainee Behavioural Checklist	339
Appendix CC: Correlation Matrix for the Scales of the LTSI	345

CHAPTER 1

Introduction

The purpose of this dissertation is to evaluate the Correctional Officer Training Program (CTP) currently being delivered by the Correctional Service of Canada (CSC) to all potential correctional officers and primary workers.¹ The overarching goal behind this evaluation is to address several questions that will provide empirical evidence as to the effectiveness of CTP, while also addressing broader issues that arise in the learning and development literature, such as the suitability of current training evaluation frameworks. The three questions that will be focused on throughout this dissertation research are briefly discussed below, along with an illustration of how they relate to CTP through a review of the literature surrounding an expanded version of Kirkpatrick's training evaluation framework (Kirkpatrick, 1959a).

The Focus of the Research

The first question that will be tackled in this dissertation is: Is CTP an effective training program? In order to determine whether or not CTP is indeed effective, it will be argued that a multi-level approach to training evaluation (that stems from the training objectives) is necessary. Specifically, an expanded version of Kirkpatrick's training evaluation framework will be tested (Kirkpatrick, 1959a). Conducting such a broad training evaluation is not currently the norm in most organizations, where the focus is usually just on reaction surveys that tap into general satisfaction with training and/or multiple choice exams that tap into only a portion of what was taught during training

¹ A primary worker is an individual who is pre-determined to work in a women's institution and thus, does not receive the firearms training component of CTP. A smaller number of Primary Workers are recruited by CSC because there are fewer women's institutions compared to men's. Because it was anticipated that the number of Primary Workers recruited during the current study would be very small, only correctional officers were examined within this dissertation.

(Machles, 2003; Phillips & Stone, 2002). Arguably, these very narrow evaluations offer little useful information to organizations and can even provide misleading results (e.g., training is deemed effective due to high satisfaction ratings when learning outcomes stemming from the training objectives have not been measured). In order for training evaluation to be useful, it will be argued that it must extend beyond satisfaction ratings and measures of declarative knowledge, to include a broader range of reactions, a more well rounded assessment of learning and, most importantly, an examination of how training transfers to the job (Baldwin & Ford, 1988; Holton, 2005; Phillips & Stone, 2002).

Given that a multi-level evaluation framework will be advocated for in this dissertation, the second question that will be addressed is how the different levels in the evaluation framework relate to one another. That is, in a training evaluation framework where reactions, learning, and transfer are measured, for example, what does data collected at one level tell you about data collected at another level? For instance, does the degree to which a trainee enjoyed the training correlate with the amount of learning achieved by the trainee, or the amount of transfer that takes place? While most studies have suggested that the answer to this question is no, this research is very limited and tends to focus only on relationships between general trainee reactions and declarative knowledge (i.e., lower-order training criteria) (e.g., Alliger & Janak, 1989; Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997). In addition, few studies have offered reasons why significant correlations are not found between the different levels of evaluation and even fewer have indicated whether there is anything that can be done to increase the correlations. This research will explore some answers to these questions.

Finally, the last question that will be explored in this dissertation relates to factors that may influence training effectiveness (referred to as barriers and enablers by Phillips and Phillips, 2007) at the various levels of an evaluation. For example, even if trainees react well to training and learn from it, are there factors that may inhibit the transferability of that training? If so, what might these factors be? Although many factors have the potential to impact training effectiveness, such as a lack of organizational support for the training (Baldwin & Ford, 1988), rarely have these factors been examined in a systematic way. This lack of examination occurs despite the fact that leaders in the field strongly outline the necessity to examine variables that either negatively impact or directly contribute to a program's success (Phillips & Phillips, 2007). Even rarer is the assessment of the predictive power of factors that may influence training effectiveness on training outcomes at levels of evaluation beyond reactions and learning. The aim of the current research is to examine the impact of such "training effectiveness" factors on each of the levels proposed in the training evaluation framework in order to determine if they indeed influence the effectiveness of training.

By addressing these three questions we will hopefully be able to determine how training programs should be evaluated in order to maximize an agency's ability to make decisions about the training program. Specifically, the results of the current dissertation will provide evidence as to the suitability of an expanded version of Kirkpatrick's evaluation framework for the purpose of evaluating CTP, and perhaps training programs being offered outside the correctional setting.

A Multi-Level Framework for Training Evaluation

While definitions of training vary, training is generally thought to consist of “instructional experiences provided primarily by employers for employees, designed to develop new skills and knowledge that are expected to be applied immediately upon (or within a short time after) arrival on or return to the job” (Broad & Newstrom, 1992, p. 5). This definition applies well across many training contexts, including that in which CTP is delivered. For example, CTP is aimed at providing potential correctional officers with the knowledge, skills, and attitudes necessary to effectively manage offenders within correctional institutions once they are hired by CSC. This is accomplished by incorporating various learning strategies into the training including an online knowledge component, workbook activities that are completed individually, and in-class sessions to teach practical skill application.

It is of great interest to know whether or not trainees generally like the training they receive and whether they are learning what is being taught to them (Phillips, 1997). However, of equal or even greater importance is determining whether or not the information that is learned is actually retained over time and utilized on the job (Baldwin & Ford, 1988). In other words, it is paramount that trainees completing CTP are able to pass their formal tests at the end of training to show, for example, that they possess relevant knowledge of CSC policy, as well as self defence skills. It is also critical, however, that the trainees remember and implement CSC’s policies, as well as their self defence skills, once they are in the institutions and interacting with offenders.

Given the importance of ensuring that trainees are reacting well to training, learning what is being taught, and transferring their new knowledge and skills to their

job, it is crucial that agencies like CSC carefully evaluate their training programs. Indeed, a sound training evaluation is necessary for a variety of reasons, of which Kirkpatrick (1994) identifies three: (1) to determine how existing training programs can be revised so that they become more efficient and effective at accomplishing their objectives, (2) to aid in decisions regarding the continuation or discontinuation of specific training programs, and (3) to provide an indication of how the training department itself fits in with the broader goals of the organization, thus justifying its existence. For all of these reasons, it is important for agencies like CSC to evaluate their training.

A Useful Starting Point: Kirkpatrick's Four-Level Framework

A variety of training evaluation models currently exist (e.g., Phillips, 1997; Stufflebeam, 1973; Warr, Bird, & Rackham, 1970), however, Kirkpatrick's (1959a) four-level evaluation framework is one of the most widely known (Alliger et al., 1997; Phillips & Phillips, 2007; Salas & Cannon-Bowers, 2001; Tannenbaum & Yukl, 1992; Van Buren & Erskine, 2002). In addition to its use within organizations, this framework has also received a lot of attention from researchers. While the results from this research do not always support the framework, there is arguably more available evidence in support of Kirkpatrick's framework compared to any of its competitors, and many successful modifications to the model have been proposed (e.g., Alliger & Janak, 1989; Alliger et al., 1997; Bates, 2004; Guerra-Lopez, 2008; Phillips & Stone, 2002; Salas & Cannon-Bowers, 2001; Tannenbaum & Yukl, 1992; Werner & DeSimone, 2012).

One of the reasons why Kirkpatrick's (1959a) framework is so popular is that it encourages the evaluation of training at multiple levels, providing the opportunity for both formative (i.e., determining how a training program should be modified) and

summative evaluations (i.e., determining whether a training program should continue or not) (Phillips, 1997). Specifically, Kirkpatrick (1959a, 1959b, 1960a, 1960b) posits that there are four levels to training evaluation, including reaction, learning, behaviour, and results. A more detailed description of what is assessed at each level is provided in Table 1 using the most recent descriptions provided by Kirkpatrick (1994). The higher levels of the framework (i.e., behaviour and results) are clearly more important than the lower levels for establishing the impact of training on an organization, although the task of incorporating higher levels in an evaluation can be quite challenging due to logistical complexities and cost (Kirkpatrick, 1994; Phillips, 1997; Salas & Cannon-Bowers, 2001; Tannenbaum & Yukl, 1992).

Table 1

Kirkpatrick's (1994) Four-Level Framework

Level	Question	Description
Level 1: Reaction	Did the trainee like the training?	Measures satisfaction reactions to the training program.
Level 2: Learning	Did the trainee learn from the training?	Measures changes in participants' knowledge, skills, and/or attitudes that result from training.
Level 3: Behaviour	Did the trainee transfer what was learned in training to the job?	Measures actual changes in behaviour on the job and the specific application of the training material.
Level 4: Results	Did the training impact the organization?	Measures the impact of training with respect to changes in quality, costs, time, productivity, etc.

Despite the need for training evaluation, it is clear that the majority of organizations do not adequately evaluate their training programs (Al-Athari & Zairi, 2002; Blanchard, Thacker, & Way, 2000; Van Buren & Erskine, 2002). Furthermore, when organizations do evaluate their training programs they tend to focus their evaluation in a very narrow fashion (e.g., focusing on general reactions to training, which may only be useful for a superficial formative evaluation). Such is the case at CSC, for example, where historically there has been a huge under-emphasis on training evaluation with most programs being evaluated mainly at Level 1 and sometimes Level 2 (S. Mongrain, personal communication, May 15, 2008). As an organization, CSC is clearly not unique in this regard.

For example, in an annual evaluation of Kirkpatrick's model, the American Society for Training and Development (ASTD; 1996) found that, out of the 276 organizations surveyed, 75% evaluated reactions, 41% evaluated learning, 21% evaluated transfer, and only 11% evaluated whether the training improved business results. The lack of adequate data collection across all levels, especially at the highest, potentially most important levels, prevents researchers from being able to sufficiently examine a range of training issues, such as the power of lower-order training criteria to predict outcomes associated with higher-order training criteria. This is clearly problematic, primarily because agencies like CSC assume that lower-order training criteria (e.g., responses to reaction surveys) can predict higher-order training outcomes (e.g., the degree to which training will transfer) despite the fact that little research exists to support that assumption (because higher-order training outcomes are rarely ever measured).

Correlations Between the Levels of Kirkpatrick's Framework

As stated above, while it is often assumed that there is a hierarchical relationship between the levels of Kirkpatrick's model (i.e., results from Level 1 will predict results from Level 2, and so on) such results are not always observed in empirical research (Alliger et al., 1997; Antheil & Casper, 1986; Dixon, 1987). Thus, a reliance on lower-order training criteria is problematic, in that it does not necessarily provide the organization with much needed information to make decisions about their training programs (e.g., whether or not a training program should be continued based on the degree to which the training enabled the transfer of skills to the job). For a training course like CTP, this means that even if trainees indicate that they liked the training program, or that they felt the training objectives were met, this does not mean that they necessarily learned the skills taught during CTP, or that they were able to apply those skills effectively within their institutions.

Previous research clearly indicates that strong relationships between the various levels of Kirkpatrick's (1959a) model do not exist. For example, in their meta-analysis, Alliger and Janak (1989) found only non-significant correlations between reactions and learning (.07), and between reactions and behaviour (.05). Likewise, in a slightly more recent meta-analysis, Alliger et al. (1997) found only non-significant correlations of .02, .03, and .07 between reactions and immediate learning, behavioural demonstrations, and transfer of skills, respectively. These research findings have led many researchers to conclude that "reaction measures cannot be used as surrogates of other measures" (Alliger et al., 1997, p. 353).

In contrast to reactions, learning measures have been found to correlate with higher-order training outcomes to a somewhat greater degree (Alliger & Janak, 1989; Alliger et al, 1997; Kirkpatrick, 1976; Kraiger, 2002; Tannenbaum, Cannon-Bowers, Salas, & Mathieu, 1993). For example, in the meta-analysis by Alliger and colleagues (1997), they found moderate positive correlations between learning and retention (.35), learning and behaviour (.18), and retention and behaviour (.14). In addition, immediate learning correlated positively with results (Level 4) (.52), although this finding was based on a very limited number of studies. These findings led these researchers to conclude that, compared to reactions to training, learning appears to be a more important predictor of higher-order training outcomes, although learning is not always sufficient to ensure that transfer will occur (at least not in the way it has traditionally been measured via Kirkpatrick's framework and not without examining factors that may influence its application).

Interestingly, despite the fact that transfer is one of the most important goals of training, it is rarely ever assessed by organizations (Alliger et al., 1997). Thus, it is not entirely clear whether transfer measures predict higher-order training outcomes, like organizational results. However, the theoretical link between transfer and results is strong. For example, most researchers argue that if training objectives are tied to organizational objectives, and training produces change in trainees that gets transferred to the job, a positive impact on the organization is likely to occur (Alvarez, Salas, & Garofano, 2004; Burke & Hutchins, 2007; Cheng & Hampson, 2008).

Potential Reasons for the Lack of Significant Correlations

The findings from research examining the relationship between levels of Kirkpatrick's (1959a, 1959b, 1960a, 1960b) framework begs the question: Why do we fail to find significant positive correlations between the levels? There are a variety of answers to these questions, and the current research will explore some of these (to be discussed in later sections). With respect to reaction-level data, one possibility is that learners have a difficult time articulating their reactions to training (i.e., if they were better at describing their reactions, positive correlations might be found; Hofstadter & Dennett, 1981). In addition, reactions to training are often influenced by training factors (e.g., whether the instructor is charismatic, fun, or likeable; Chandler, 1978; Kaplan & Pascoe, 1977) and trainee characteristics (e.g., the trainee's personality, mood, and attitudes; Clement, 1982) that have little to do with whether the training is effective at instilling knowledge or skills that can be transferred to the job. Finally, there is often a disconnect between what is measured at the various levels of evaluation; for example, reaction measures are attitudinal in nature, whereas the measurement of learning, behaviour, and results are primarily cognitive, competency or skills-based (Alliger & Janak, 1989).

The Need for a Multi-Dimensional, Multi-Level Training Evaluation Framework

Given the research findings presented above, if organizations want to know if their training is having a significant impact they will have to either consistently evaluate their training at the highest levels of the evaluation framework, or find significant relationships between lower-order training criteria and higher-order training outcomes that have yet to be found. Given that the first option is unlikely to happen on a large

scale, due to the resources required or the suitability of a training program for a Level 4 evaluation (ASTD, 2003; Phillips, 1997), there is value in exploring some of the reasons why low correlations have been found between the levels of Kirkpatrick's (1959a, 1959b, 1960a, 1960b) framework. Beyond the potential explanations already provided (e.g., learners are unable to articulate their reactions, reactions are based on factors unrelated to training, trainee characteristics measured have little to do with training), one of the key reasons why researchers find small correlations between the levels likely has to do with the uni-dimensional nature of Kirkpatrick's original levels.

The Uni-dimensionality of Kirkpatrick's Levels

When the four levels of evaluation were originally proposed by Kirkpatrick, assessing "reactions" meant quite literally asking the trainee how they liked the training (Alliger et al., 1997). For example, a typical item on a reaction questionnaire was/is, "I found this training program to be enjoyable" (Alliger et al., 1997). In addition, learning was originally indexed by results from traditional tests of declarative knowledge (Alliger et al., 1997). Specifically, this level of evaluation related to the "principles, facts, and techniques understood and absorbed by the [trainees]" (Alliger & Janak, 1989, p. 331).

It is now clear that at least some of Kirkpatrick's levels, particularly reaction and learning, represent multi-dimensional constructs and this opens up the possibility that lower-order training criteria are not predictive of higher-order outcomes because the wrong things are being measured in training evaluations (Kraiger, Ford, & Salas, 1993; Morgan & Casper, 2000). It is particularly likely that this is the case if some of the items being omitted from training evaluations relate directly to the training objectives (e.g.,

practical skills, in addition to knowledge, are being targeted in training) (Kraiger et al., 1993).

When one considers training programs like CTP, it is easy to see why such a uni-dimensional approach to training evaluation might be problematic. As is made clear in the Methods section where CTP is discussed in detail (see below), this training program has a wide range of objectives, many of which would not be captured using Kirkpatrick's original model. Consider a Level 2 evaluation as an example. In addition to increasing declarative knowledge, CTP is also designed to change the attitudes of trainees (towards inmates and correctional work) and to enhance their skill set (ranging from writing reports to the handling of weapons). Clearly, a sole focus on declarative knowledge when conducting a Level 2 evaluation of CTP would hide many important things. Fortunately, efforts by a number of researchers to flesh out the multi-dimensional nature of Kirkpatrick's levels can be used to significantly improve evaluation efforts.

Expanding the Levels: Making Kirkpatrick's Framework More Psychologically Sound

Recent research has indicated that each of the levels of Kirkpatrick's (1959a, 1959b, 1960a, 1960b) four-level framework can be expanded to take into account the range of training objectives that are typical of most modern training programs, including CTP. Research related to this issue at Levels 1-3 will be discussed briefly in the sections that follow. The reason for omitting Level 4 from the current evaluation is also discussed below.

Level 1. As indicated above, Level 1 evaluations typically focus on positive satisfaction ratings (i.e., whether trainees liked the training) even though these reactions

do not appear to predict higher-order training outcomes. However, since Kirkpatrick's model was originally proposed, research has suggested ways of expanding this level.

Most of the recommendations that have been made encourage evaluators to move beyond very general satisfaction ratings that tap into how much trainees enjoyed the training (especially non-content issues like location, media, food, etc.), in order to examine the degree to which trainees are satisfied with other, potentially more important aspects of the results achieved from the training (Phillips & Phillips, 2007). One potential set of reactions that has received some attention in this regard relates to the perceived utility or relevance of training (e.g., To what extent do you think the training gave you skills that are useful for your job?) (e.g., Alliger et al., 1997; Axtell, Maitlis, & Yearta, 1997; Baumgartel, Reynolds & Pathan, 1984; Phillips, & Phillips, 2007; Ruona, Leimbach, Holton, & Bates, 2002; Tan, Hall, & Boyce, 2003; Warr, Allan, & Birdi, 1999). Empirically, this type of reaction often predicts higher-order training outcomes to a greater extent than positive satisfaction ratings (Ruona et al., 2002; Tan et al., 2003; Tannenbaum & Yukl, 1992), leading researchers to conclude that for training to be effective, "learners should perceive that the new knowledge and skills will improve a relevant aspect of their work performance" (Burke & Hutchins, 2007, p. 269).

In addition to utility reactions, the importance of measuring perceived training difficulty has also been raised (e.g., To what extent did you feel overwhelmed by the training?). According to Warr and Bunce (1995), perceived difficulty relates to the cognitive and emotional effort required to perform well in training. However, while this variable is intuitively appealing it has not been studied as thoroughly as general satisfaction or utility reactions (e.g., Alliger & Janak, 1989; Alliger et al., 1997).

Nevertheless, the small amount of research that does exist suggests that perceived difficulty can predict higher-order training outcomes to a greater extent than other reaction measures, with Warr et al. (1999) reporting correlations between perceived training difficulty and higher-order training outcomes in the range of -.26 to -.52. These findings accord well with theories of learning, such as cognitive load theory (Clark, Nguyen, & Sweller, 2006; Chandler & Sweller, 1991; van Merriënboer, 1997), which suggest that cognitive overload during training will invariably lead to low levels of learning and behaviour.

Having made the case for moving Level 1 evaluations beyond very general satisfaction ratings, it must be said that there is still value in collecting this sort of data, not just for the purpose of predicting other training outcomes but for understanding how learners perceive the training experience (Phillips & Phillips, 2007). For example, general satisfaction reactions can play a key role in formative evaluations where an attempt is made to improve training with respect to program content, materials, and delivery (Kirkpatrick & Kirkpatrick, 2005; Phillips & Stone, 2002). These ratings are also useful for demonstrating to trainees that their opinions are valued (Kirkpatrick, 1994).

Level 2. Because not all training programs target the same type of learning, a range of learning outcomes should be measured when evaluating training (Kraiger et al., 1993; Phillips & Stone, 2002). This will require the use of assessment instruments beyond traditional knowledge tests (e.g., multiple choice exams). The selection of what learning outcomes to focus on will depend on the training objectives. Kraiger et al. (1993) have proposed a commonly accepted taxonomy of learning outcomes, which

should guide the evaluation strategy. These include: (1) knowledge outcomes, (2) skills-based outcomes, and (3) affective outcomes, each of which can be sub-divided further.

Knowledge outcomes can be categorized as verbal knowledge, knowledge organization, and cognitive strategies and even these domains can be split (Kraiger et al., 1993). For example, in the case of CTP, verbal knowledge, or knowledge in the form of facts and concepts, is clearly the primary kind that is targeted and the specific types of verbal knowledge focused on consist of declarative knowledge (i.e., information about what), procedural knowledge (i.e., information about how), and tacit knowledge (i.e., information about which, when, and why) (Anderson, 1982; Kraiger et al., 1993; Wagner, 1987). Thus, when tests of knowledge are used to evaluate CTP, these tests need to be constructed so that they tap into these various aspects of the training program (Alliger et al., 1997; Kraiger et al., 1993; Phillips, 1997).

Skills development (in addition to enhanced knowledge) is another key training objective in CTP and like the knowledge domain, this domain can be further sub-divided. From an evaluation perspective, important areas of skills development include: (1) compilation and (2) automaticity (Kraiger et al., 1993). Of primary concern in CTP is compilation, which refers to the development of: (1) domain-specific routines (e.g., ability to identify problematic situations, and take the necessary measures to ameliorate them) that become ever more fluid as expertise develops, (2) an ability to apply those routines to novel situations, and (3) an understanding of when those routines cannot be applied (Anderson, 1982). The optimal evaluation instrument for assessing compilation is behavioural observation of trainees (e.g., by subject matter experts) taking part in role

plays (Kraiger et al., 1993). This assessment technique will have to be incorporated into an evaluation of CTP if it is to be adequate.

Lastly, affective learning outcomes are of interest in the current study, and these too can be sub-divided into: (1) changes in attitudes or (2) changes in motivations. This learning outcome should not to be confused with Level 1 reaction data, which are often related to affective reactions towards training. Here, the change in affect itself – either an attitude or a motivation – is the target of training. CTP clearly attempts to influence the attitudes of trainees (towards correctional work and inmates) as one of its training objectives, and thus these attitudes should be measured in order to evaluate the training program. Indeed, according to Kraiger et al. (1993), relying solely on “behavioral or cognitive measurement [in the context of a training evaluation] at the expense of attitudinal and motivational measurement provides an incomplete profile of learning and the learning process” (p. 318).

Thus, we know that CTP targets a variety of learning-related outcomes, including verbal knowledge, compilation of skills, and attitudes towards correctional work and inmates. By increasing the breadth of measurement at Level 2, a more in-depth examination of the relationships between the levels of Kirkpatrick’s framework will be possible, and a more appropriate examination of the training objectives in CTP can take place.

Level 3. In Kirkpatrick’s original model, Level 3 was referred to as behaviour. Today, Level 3 is commonly referred to as transfer (Alliger et al., 1997) or application and implementation (Phillips & Phillips, 2007). The transfer or application label accurately captures what is being measured at Level 3. Specifically, evaluators are

interested in assessing the degree to which what was “acquired in a training program are applied, generalized, and maintained over some time in a job environment” (Salas & Cannon-Bowers, 2001, p. 488). Due to the lack of research at this level it is not clear what the optimal multi-level evaluation options are. Presumably, an expanded Level 3 evaluation should track all of the items (domains) measured in Level 2 when the trainee is actually on the job (e.g., the implementation of knowledge, skills, and attitudes). This of course will require the use of various assessment instruments.

Typically when implementing Kirkpatrick’s model, transfer is evaluated by having supervisors fill out a survey or checklist describing the trainees’ behaviour on the job (e.g., Hand, Richards, & Slocum, 1973; Peters, O’Connor, & Eulberg, 1985; Tan et al., 1993; Tracey, Tannenbaum, & Kavanagh, 1995). According to Phillips and Stone (2002), supervisors are a rich resource for providing information about the trainees that they directly oversee as they are able to “observe the [trainees] as they attempt to use the knowledge and skills acquired in the [training] program” (p. 145). However, there are well-known problems with supervisor ratings, which suggest that it would be useful to expand the range of assessment instruments used to conduct a Level 3 evaluation. For example, supervisor ratings can be biased for a variety of reasons (Holzbach, 1978; Mann, 1997). In addition, supervisor ratings can provide incomplete information if the supervisor does not have the opportunity to observe trainees in action, if the supervisor is unfamiliar with the training, or if the supervisor is responsible for overseeing too many employees (Jones, Steffy, & Bray, 1991).

Given these concerns, it is now recommended that multiple measures be used to evaluate Level 3 (beyond supervisor ratings). Possible options include the use of self-

report surveys completed by trainees as well as an examination of performance data (i.e., measurable aspects of job performance; e.g., an assessment of written reports completed by trainees on the job to determine whether they are completed correctly) (Phillips & Phillips, 2007). As with supervisor ratings, self-report ratings are subjective in nature, but provide valuable information regarding whether the trainee's themselves perceive that they have transferred their training. Performance data on the other hand provide a way to objectively determine the ways in which training has transferred to the job (Phillips, 1997; Phillips & Stone, 2002).

At least two of the Level 3 evaluation strategies (supervisor checklists and self-report surveys) can be used in the current study to help evaluate whether trainees in CTP transfer their knowledge, skills, and attitudes to their job. Performance data was not collected due to the difficulty in obtaining the data as well as the inconsistencies in reporting that occurred based on institution. For example, when officers write reports they are typically filed under the inmate's name that was involved in the incident. In order to obtain the reports we would have to ask the officers to keep track of all the reports written and then provide the inmate's name in order to retrieve them. This process would not only violate the privacy of the inmates but would also allow officers to omit any reports they felt were not well done. Additionally, due to the high amount of incidents that occur in maximum-security institutions, reports are not always written for each one (it might depend on the severity of the incident), whereas at minimum institutions, fewer incidents occur making it more likely that all incidents are reported. It was determined that these problems associated with performance measures in the current study posed too much of a threat to their validity.

Level 4 measures were not included in the current study mainly due to the nature of the program being evaluated. Recall that CTP is an entry-level, or core, training program that must be completed by all new correctional officers. It is not possible for the job to be done without completing this training program first and thus the impact on the organization is less of a concern and focus for CSC here. The negative impact of failing to implement such a program would be too great. In addition, an extensive needs analysis carried out by the Training Design and Development Division at CSC when updating this program (prior to the development of this evaluation) addressed the needs of the organization. Support for measuring only Levels 1 through 3 in the current study was also provided by Phillips who indicated that Level 4 measures are not always necessary or appropriate for entry-level training programs (J. Phillips, personal communication, April 3, 2011).

The Need to Consider Training Effectiveness Factors

While Kirkpatrick's training evaluation framework addresses many important evaluation issues, it fails to take into account important effectiveness issues, even in its revised and expanded form. To clarify, training evaluation is about measuring training outcomes in order to determine if a training program worked (or how it can be improved). Training effectiveness, on the other hand, is about developing an understanding of the entire learning system, including an attempt to determine why lower-order training criteria do, or do not, predict higher-order training outcomes and why learning and transfer does, or does not, take place (Alvarez et al., 2004). Phillips (1997) advocates for the collection of effectiveness data (referred to as barriers and enablers) along with other Level 3 measures. This data relates to the need to identify "barriers to success, problems

encountered during implementation...obstacles to the actual application... [as well as]...factors that directly contribute to program success” (Phillips & Phillips, 2007, p. 165).

Most of the emphasis in the training effectiveness literature is on understanding the conditions under which training transfer occurs (e.g., Broad & Newstrom, 1992; Holton, 1996; Kozlowski & Salas, 1997; Noe, 1986; Noe & Schmidt, 1986; Thayer & Teachout, 1995). For example, in one representative model of training transfer proposed by Baldwin and Ford (1988), transfer of training is seen to depend not only on the amount of learning that results from training, but on things like the trainee’s motivation to learn, the relevance of the training content, and organizational factors such as supervisor support. From these sorts of models, three broad categories of confounding, or intervening variables have been identified that can impact the likelihood of training transfer: trainee characteristics, training design features, and organizational climate factors (see Table 2).

Table 2

The Transfer of Training Model by Baldwin and Ford (1988)

Categories of Influencing Factors	Specific Examples Within the Category
Trainee Characteristics	Ability and aptitudes Personality Motivation
Training Design Characteristics	Principles of learning Sequencing Training content
Work Environment Characteristics	Supportive organizational climate Freedom to set goals Opportunity to use knowledge and skills

Empirical research does seem to suggest that these variables matter. For example, Noe (1986) conducted a review in order to identify the types of attitudes and personality characteristics possessed by trainees that may influence their motivation to learn as well as the degree to which they transfer what they learn to the job. Using Kirkpatrick's original model as his base, he found, amongst other things, that trainees with an internal locus of control and a high level of self-efficacy are more likely to have a high motivation to learn. In addition, he found that a supportive work climate where co-workers and supervisors provide reinforcement and feedback is more likely to result in transfer of training.

In a similar vein, Noe and Schmidt (1986) conducted a study using path analysis to examine relationships among various trainee characteristics (e.g., pre-training motivation, job involvement, career planning, locus of control, etc.) and several training outcomes (e.g., learning, behaviour change, performance measures, etc.) in trainees taking part in a two-day training program targeting administrative and interpersonal skills. They found, amongst other things, that a higher level of job involvement was significantly related to increased skill acquisition, and that a higher level of job involvement and a greater degree of career planning were significantly related to greater levels of transfer. Similar results have been reported by Mathieu, Tannenbaum, and Salas (1992).

Likewise, Seyler, Holton, Bates, Burnett, and Carvalho (1998) found that in a computer-based training program for a large petrochemical company a variety of individual attitudes (e.g., desire to learn, internal work motivation, organizational commitment, etc.) and environmental variables (e.g., supervisor support and sanctions,

peer support, opportunity to use the training, etc.) significantly correlated with motivation to transfer what was learned in training. The largest correlations were observed between motivation to transfer and opportunity to use training ($r = .58$), peer support ($r = .54$), organizational commitment ($r = .41$), and supervisor sanctions ($r = -.40$).

Along the same lines, Baumgartel and colleagues (1984) conducted two studies to examine the moderating effects of trainee characteristics (e.g., locus of control, need for independence, belief in the value of training, etc.) and organizational climate (e.g., freedom to set performance objectives, use of knowledge gained during training, risk-taking in the organization, etc.) on various outcomes (reactions, learning, and transfer) in week long management development courses. In both studies, they found that personal and organizational climate variables significantly moderated training effects. For example, high internal locus of control ratings coupled with highly favourable work climate ratings significantly moderated the relationship between management rank (i.e., job level of the managers) and the degree to which trainees applied what they learned from the training program on the job.

So Why Does Training Effectiveness Matter?

Within the context of a training evaluation, the important point to make is that, without taking training effectiveness factors into account, incorrect decisions about the value of training programs can be made (Holton, 1996, 2005). For example, when evaluating CTP, if non-significant correlations are found between Level 1 and Level 3, this could be interpreted as being an inherent limitation of the use of reaction-level data to predict higher-order training outcomes. However, it could be the case that reaction-level data would predict transferability to a high degree if it were not for the fact that the

trainees taking part in the training program are not supported by their organization when trying to implement what they learned in training. Related to this is the fact that CTP could be deemed ineffective because high levels of transfer do not occur, when in fact the training program is adequate, but organizational support is lacking. In both scenarios, it is impossible to determine whether the problem lies with CTP and/or how it is evaluated, or in training effectiveness factors (Holton, 1996, 2005). The unfortunate consequence of this is that training programs are often inappropriately targeted for change rather than the organization itself.

Holton's Learning Transfer System Inventory

What is needed is a systematic way of examining the relevance of various training effectiveness factors when evaluating a training program so as to offset the serious limitations of Kirkpatrick's evaluation framework. Until recently, no systematic approach existed for doing this, but since the mid-1990s Holton and his colleagues have dedicated their time to the development of a standardized training effectiveness questionnaire that can be used for this purpose (e.g., Chen, Holton, & Bates, 2005; Holton, 1996, 2005; Holton, Bates, & Ruona, 2000; Holton, Bates, Seyler, & Carvalho, 1997).

Their questionnaire, called the Learning Transfer System Inventory (LTSI) (Holton et al., 2000), systematically measures 16 major factors that have been found to influence the impact of training (see Table 3 for a description). The LTSI provides a way to define and measure a range of issues relating to the individual trainee, the training itself, and the organization or work environment that may impact the degree to which learning through training is transferred to the job. The instrument is premised on the fact that "transfer can only be completely understood...by examining the entire system of

influences” (Holton, 2005, p. 44). This instrument was developed as an extension of the HRD [Human Resource Development] Evaluation and Research and Measurement Model (proposed by Holton, 1996), which was designed to provide “a more comprehensive framework for diagnosing and understanding the causal influences of HRD intervention outcomes” (Holton, 2005, p. 37). The effectiveness factors included in the LTSI were determined based on previous research, which examined a wide variety of relevant issues (e.g., Baldwin & Ford, 1988 (motivation to transfer); Noe & Schmitt, 1986 (job attitudes); Ree & Earles, 1991 (ability); Wexley & Baldwin, 1986 (transfer design), etc.).

Table 3

Description of the Factors Included in the LTSI (Holton et al., 2000)

Factor	Definition	Sample Item
Learner Readiness	The extent to which individuals are prepared to enter and participate in training.	Before the training I had a good understanding of how it would fit my job-related development.
Motivation to Transfer	The direction, intensity, and persistence of effort toward utilizing the skills and knowledge learned in a work setting.	I get excited when I think about trying to use my new learning on my job.
Positive Personal Outcomes	The degree to which applying training on the job leads to outcomes that are positive for the individual.	Employees in this organization receive “perks” when they utilize newly learned skills on the job.
Negative Personal Outcomes	The extent to which individuals believe that not applying skills and knowledge learned in training will lead to outcomes that are negative.	If I do not utilize my training I will be cautioned about it.
Personal Capacity for Transfer	The extent to which individuals have the time, energy, and mental space in their work lives to make changes required to transfer learning to the job.	My workload allows me the time to try the new things I have learned.
Peer Support	The extent to which peers reinforce and support use of learning on the job.	My colleagues encourage me to use the skills I have learned in training.
Supervisor Support	The extent to which supervisors/managers support and reinforce use of training on the job.	My supervisor sets goals for me that encourage me to apply my training on the job.
Supervisor Opposition	The extent to which individuals perceive negative responses from supervisors/managers when applying skills learned in training.	My supervisor opposes the use of the techniques I learned in training.

Factor	Definition	Sample Item
Perceived Content Validity	The extent to which trainees judge training content to reflect job requirements accurately.	What is taught in training closely matches my job requirements.
Transfer Design	The degree to which (1) training has been designed and delivered to give trainees the ability to transfer learning to the job, and (2) training instructions match job requirements.	The activities and exercises the trainers used helped me know how to apply my learning on the job.
Opportunity to Use	The extent to which trainees are provided with or obtain resources and tasks on the job enabling them to use training on the job.	The resources I need to use what I learned will be available to me after training.
Transfer Effort-Performance Expectations	The expectation that effort devoted to transferring learning will lead to changes in job performance.	My job performance improves when I use new things that I have learned.
Performance-Outcomes Expectations	The expectation that changes in job performance will lead to valued outcomes.	When I do things to improve my performance, good things happen to me.
Resistance-Openness to Change	The extent to which prevailing group norms are perceived by individuals to resist or discourage the use of skills and knowledge acquired in training.	People in my group are open to changing the way they do things.
Performance Self-Efficacy	An individual's general belief that he is able to change his performance when he wants to.	I am confident in my ability to use newly learned skills on the job.
Performance Coaching	Formal and informal indicators from an organization about an individual's job performance.	After training, I get feedback from people about how well I am applying what I learned.

The 16 different factors that the LTSI taps into relate to the three categories of training effectiveness factors proposed by Baldwin and Ford (1988). For example, in terms of trainee characteristics, the LTSI includes variables such as performance

expectations and performance self-efficacy. Performance expectations can be conceptualized as an individual's perceived locus of control, which influences an individual's beliefs about their ability to gain knowledge and skills (Spector, 1982). Those individuals with an internal locus of control, who believe that job performance is under their own control, are more likely to exhibit higher levels of motivation to learn, achieve greater learning outcomes, and display greater levels of transfer (Noe, 1986). Self-efficacy is defined as "people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performance" (Bandura, 1977, p. 391). Higher levels of self-efficacy often lead to increased participation during training, greater levels of knowledge and skill acquisition, and a higher degree of transfer (Ford, Quinones, Segó, & Sorra, 1992; Gaudine & Saks, 2004; Noe & Wilk, 1993).

With respect to training design characteristics, the LTSI taps into variables such as perceived content validity and transfer design. Research has shown that the more the training content is related to the work environment (i.e., the greater the relevance), the more likely the content will be transferred to the job (Axtell et al., 1997). For example, a high level of similarity between what is taught during training and what is expected on the job has been associated with greater levels of knowledge and skill acquisition, and greater retention (Bhatti & Kaur, 2010; Gagne, Baker, & Foster, 1950; Holton et al., 2000; Thorndike & Woodworth, 1901; Underwood, 1951; Velada, Caetano, Michel, Lyons, & Kavanagh, 2007). With respect to transfer design, the greater the degree to which training has been designed and delivered to give trainees the ability to transfer learning to the job (e.g., as determined by the degree to which training instructions match

job requirements), the greater the chance that what is learned during training will transfer to the job (Holton, 2005).

Lastly, regarding work environment characteristics, the LTSI measures things such as supervisor support, the opportunity to use new knowledge and skills, positive personal outcomes, supervisor sanctions, and performance coaching. Research has shown that increased perceptions of support, both by peers (i.e., coworkers) and supervisors often results in greater transfer (Foxon, 1997). In addition, the more opportunities there are to use newly acquired knowledge and/or skills the more likely it is that transfer will occur (Clarke, 2002). Related to both of these points is the degree to which trainees are reinforced for their use of knowledge and skills gained during training. In particular, recognition, feedback, rewards and punishments by supervisors, coworkers, and subordinates have a large influence on whether or not transfer will take place (House, 1968; Noe, 1986).

Thus, there is clear support from the general research literature that the factors included in the LTSI can have an impact on the likelihood that things learned in training will transfer to one's job. Unfortunately, in studies where the LTSI has been used to illustrate the impact of training effectiveness factors on evaluation outcomes, a limited number of training outcomes have been focused on (e.g., learning; Seyler et al., 1998), or the focus of the study has been on delineating the impact of one particular LTSI factor (e.g., motivation to transfer; Seyler et al., 1998). Rarely are studies conducted where the LTSI is used to help explain or predict multiple outcomes from multiple levels of Kirkpatrick's evaluation framework, or the correlations that exist between the levels. This study will help to address this gap.

The Current Study

Currently, CSC offers CTP to all potential correctional officers and primary workers. It is a training program designed to impart the knowledge, skills, and attitudes needed to effectively manage offenders within correctional institutions. The goal of the current study is to conduct a large scale evaluation of CTP (focusing solely on correctional workers) using an enhanced version of the first three levels of Kirkpatrick's (1959a) four-level framework. Thus, the training program will be evaluated at the levels of reaction, learning, and behaviour, but each level will be expanded in line with the recommendations presented throughout the literature review (see the Methods section for details of what exactly will be evaluated at each level). The LTSI will also be completed by each of the trainees being evaluated in the study to examine how various training effectiveness factors affect training outcomes.

Collectively, the data emerging from this training evaluation (and the training effectiveness surveys) will be used to answer the three questions posed at the beginning of this dissertation: (1) Is CTP effective in meeting its objectives of providing trainees with the knowledge, skills, and attitudes that are necessary to perform well in institutions?; (2) How do the various levels of Kirkpatrick's framework relate to one another, and to what degree can lower-order training criteria predict higher-order training outcomes?; and (3) How is the effectiveness of CTP influenced by training effectiveness factors that are included in the LTSI (and the demographic questionnaire)?

It is expected that the results of this study will further our knowledge about the utility of the expanded version of Kirkpatrick's evaluation framework as well as the potential influence of training effectiveness factors. In addition, the results will be

important to CSC in terms of demonstrating the effectiveness of this core training program. The results will also provide CSC with valuable information that they can use to inform potential modifications to CTP.

Hypotheses

The following hypotheses for the analyses relating to the training outcomes (Levels 1-3), as well as the relationships between levels and the training effectiveness factors were examined in the current study:

- (1) For the Level 1 outcome, positive ratings of the training program were expected regardless of where the training took place (i.e., regardless of region).
- (2) For the Level 2 outcomes, significant increases in test scores were expected across each of the training outcomes between the recruitment and training phases (i.e., pre-post tests), indicating that learning had occurred. It was expected that this would be the case across each of the training outcomes (regardless of test subsections or individual skills), regardless of region. In addition, it was expected that most of what was learned would be retained during the transfer phase (post-transfer tests); though a small degree of forgetting might be expected. It was also expected that at each study phase the different Level 2 measures would be positively correlated with each other (e.g., knowledge and skills, etc.) and that for those Level 2 outcomes with subsections (e.g., knowledge and soft skills), the subsections would be positively correlated with the overall score. Lastly, it was expected that the same Level 2 measures would be positively correlated across study phases (i.e., that high levels of knowledge acquisition at the training phase would be related to high levels of knowledge retention at the transfer phase, etc.).

- (3) For the Level 3 outcomes of frequency and quality of skill use on the job, it was expected that all skills targeted in CTP would be used within the institutions and that they would be used in a way that was consistent with how CTP trains its trainees to use the skills (i.e., frequency ratings and quality ratings will be relatively high, regardless of who is providing the ratings). Regional differences were not expected. In addition, it was expected that the trainee and supervisor ratings would be positively correlated.
- (4) In general, significant positive correlations between the outcome measures at one level (e.g., Level 1) and the outcome measures at another level (e.g., Level 2) are expected. This is based on the inherent assumption of Kirkpatrick's evaluation framework, that results from one level will predict results from the next. This applies in the current study for relationships between measures at Level 1 and 2, Level 1 and 3, and Level 2 and 3.
- (5) Based on the existing literature that has examined the psychometric properties of the LTSI, it is expected that high reliability levels will be found for the majority of the 17 subscales.
- (6) In addition, based on relationships found in the existing literature, positive correlations are expected between the training outcomes and the majority of the LTSI subscales (with the exception of supervisor opposition which is measured in a negative direction).
- (7) In addition, it is expected that the majority of the subscales will be rated at moderate to high levels by trainees.

(8) Despite expectations with regard to subscale reliability, only those with reliability scores of .80 or greater will be used for the moderator analysis. At this time it is not known which scales will meet that criterion. If highly reliable scales are identified, moderator analyses will be carried out and depending on the specific moderators identified; the predictor and outcome variables of interest will be outlined (including training outcome variables from the various study phases). In general, it is hypothesized that each moderator will strengthen the relationship between the predictor and outcome variables.

CHAPTER 2

Methodology

The training program evaluated in the current study was the Correctional Training Program (CTP). CSC currently offers CTP to individuals who successfully complete the recruitment process in order to become correctional officers. CTP is designed to provide trainees with the theoretical knowledge, attitudes, and practical skills that are necessary to carry out their job successfully. While some form of correctional officer training has been in place within CSC since the 1970s, the current conceptualization of CTP has been in existence since 1999 (D. Bally, personal communication, November 3, 2008). However, CTP was recently updated, with design and development beginning in 2005. An on-line component (Stage 1) was added to the existing training program and the newest version of CTP, including the on-line component, was first delivered in 2008.

Currently, CTP is comprised of four stages, delivered over 6 months. For a description of each stage see Table 4 below. Stages 1 and 3 of CTP are further broken up into various courses, each covering a specific subject area. Stage 2 is essentially a preparatory stage for Stage 3 while Stage 4, completed on the job, involves a skills checklist. For an outline of the various courses and topics covered in Stages 1, 3, and 4 see Appendices A through C.

Table 4

A Brief Outline of the Four Stages of CTP

Stage	Description	Delivery Time
1	This portion of CTP is delivered online and consists of 10 different courses. The emphasis is mostly on the theoretical or knowledge-based information needed to move through each training course. The training activities include matching activities, review questions, and quizzes.	8 weeks
2	This portion of CTP is completed by trainees at home and is comprised of knowledge-based workbook activities designed to prepare trainees for the first few weeks of Stage 3 (Courses 1 through 9). It also includes a formal theory test before starting Stage 3.	2 weeks
3	This portion of CTP is delivered in-class at a Regional Correctional Staff College with an emphasis on both theory and skills application. The in-class sessions consist of several types of training activities including class discussions and role play scenarios, as well as formal theory and skills testing.	12 weeks
4	This portion of CTP involves on the job training. It is conducted within institutions and is not governed by the trainers of CTP. This stage is not carried out or evaluated consistently across institutions. It is comprised of skills checklists to be completed at the institution over a two-week period. Some of the areas covered include the ability to conduct inmate counts, proper use of communication equipment, etc. ²	2 weeks

Data collection for the current project took place from June 2009 to March 2011.

During this time a total of 17 different cohorts of trainees (i.e., 17 different CTP's) were involved in the study and training was delivered at Staff Colleges in three different regions. Specifically, there were five cohorts (or separate CTP's) from the Pacific region,

² This portion of CTP was not heavily focused on in the current study since it is governed by the individual institutions and is targeted for updating and further development by the Learning and Development Branch at CSC.

eight from Ontario and three from Quebec.³ Due to changes in the recruitment procedures towards the end of the data collection period, one trainee was recruited in Ontario but trained in the Prairie region.⁴

Participants. One hundred and nineteen potential correctional officers, out of a possible 216, were recruited to participate in the study. This represents a completion rate of 55%. Of those 119 trainees, 17 were excluded from the study. There were two reasons trainees were excluded: (1) they were not successful during CSC's recruitment process or withdrew themselves from the process (i.e., they did not receive or accept the invitation to complete a CTP) ($n = 13$); or (2) they were selected for a Primary Worker position ($n = 4$). Recall that Primary Workers work only in women's institutions and the training they receive is different from the training received by correctional officers working in men's institutions. The final sample consisted of 102 trainees.

The number of trainees involved in the current study varied across the different phases of analysis (explained in more detail below in the Procedure section). This was due largely to expected attrition. Specifically, some trainees did not proceed all the way through CTP (i.e., some withdrew from the program and others failed formal tests), whereas others were not hired immediately following the completion of the program.⁵

See Table 5 below for specific details on the overall participation rates per region based

³ The updated CTP was rolled out at different times across the country starting with Pacific, followed by Ontario, Quebec, the Prairies and Atlantic. At the start of the current evaluation only Pacific and Ontario were delivering the program. Quebec came on board towards the end of the data collection period. This roll out schedule, coupled with the fact that the number of CTP's delivered in each region depends on the needs of that region, explain why there are differing numbers of cohorts (CTP's), and thus participants, from each region in the current study.

⁴ Due to being re-located to the Prairies, the only data collected for this participant was at the recruitment phase. Thus it was coded and included as data from the Ontario Region.

⁵ This only occurred in the Quebec region. In this region, successful CTP graduates are kept in a pool awaiting employment. By the time the current study had concluded, these participants were still awaiting a job offer.

on the phase of the study (the same information for each separate cohort of CTP can be found in Appendix D). Analyses were conducted to examine differences between those who completed each phase and those who did not. These findings are provided below in the Results section.

Table 5

Participation Numbers per Region and Study Phase

Region	Recruitment Phase % (n)	Training Phase % (n)	Transfer Phase % (n)
All Regions	100 (102)	67 (68)	52 (53)
Pacific	100 (37)	92 (34)	70 (26)
Ontario	100 (51)	53 (27)	47 (24)
Quebec	100 (14)	50 (7)	21 (3)

An overview of the demographic variables of all trainees (per region) is presented in Table 6. The data provided below is from the trainees who completed the recruitment phase (i.e., all 102 trainees). As the study progressed and less trainees completed the latter two phases (e.g., training and transfer), the demographic data obviously changed. This information can be found in Appendix E.

Table 6

Demographic Variables of Participants

	All Regions % (n)	Pacific % (n)	Ontario % (n)	Quebec % (n)
	N = 102	n = 37	n = 51	n = 14
Gender				
Female	33 (34)	38 (14)	36 (18)	14 (2)
Male	67 (68)	62 (23)	64 (32)	86 (12)
Ethnicity^a				
Caucasian	76 (77)	73 (27)	74 (37)	86 (12)
Asian	7 (7)	5 (2)	10 (5)	-
African	3 (3)	5 (2)	2 (1)	-
Aboriginal	4 (4)	8 (3)	2 (1)	-
European	3 (3)	8 (3)	-	-
Other	1 (1)	-	-	1 (1)
Age				
< 20 years	1 (1)	-	-	7 (1)
20-25 years	23 (23)	27 (10)	22 (11)	7 (1)
26-30 years	28 (29)	22 (8)	30 (15)	43 (6)
31-35 years	16 (16)	19 (7)	14 (7)	14 (2)
36-40 years	10 (10)	11 (4)	6 (3)	21 (3)
> 40 years	23 (23)	22 (8)	28 (14)	7 (1)
Marital Status				
Single	32 (33)	38 (14)	26 (13)	36 (5)
Married/common law	59 (60)	51 (19)	66 (33)	57 (8)
Separated/divorced	9 (9)	11 (4)	8 (4)	7 (1)
Previous Work Experience				
No experience	40 (41)	43 (16)	28 (14)	71 (10)
<5 years related experience	42 (43)	41 (15)	50 (25)	21 (3)
5-10 years related experience	14 (14)	16 (6)	14 (7)	7 (1)
>10 years related experience	4 (4)	-	8 (4)	-

	All Regions % (n)	Pacific % (n)	Ontario % (n)	Quebec % (n)
	N = 102	n = 37	n = 51	n = 14
Previous Volunteer Experience^b				
No experience	70 (71)	70 (26)	64 (32)	86 (12)
<5 years related experience	25 (25)	30 (11)	24 (12)	14 (2)
5-10 years related experience	3 (3)	-	6 (3)	-
Educational Background^c				
High School	31 (32)	32 (12)	26 (13)	50 (7)
Unrelated College Diploma	17 (17)	16 (6)	18 (9)	14 (2)
Related College Diploma	28 (29)	19 (7)	40 (20)	14 (2)
Unrelated University Degree	10 (10)	16 (6)	2 (1)	21 (3)
Related University Degree	10 (10)	16 (6)	6 (3)	-

Note. ^an = 7 missing (all regions), n = 6 missing (Ontario), n = 1 missing (Quebec). ^bn = 3 missing (all regions), n = 3 missing (Ontario). ^cn = 4 missing (all regions), n = 4 missing (Ontario).

Measures

Informed Consent Forms. A copy of the informed consent form that was signed by CTP trainees is provided in Appendix F. This form was provided to trainees at the outset of the study, during the recruitment phase. Supervisors of trainees who were successful in CTP and were hired also received an informed consent form before they were asked to evaluate trainees regarding their on the job behaviours (see Appendix G). This form was provided to the supervisors via email immediately before they completed the behavioural checklist (i.e., four-six months after the trainee had completed CTP).

Demographic Questionnaire. Trainees were asked to fill out a demographic questionnaire at the outset of the study during the recruitment phase (i.e., before training commenced). As seen in Table 6, the following information was obtained: (1) gender, (2)

ethnicity, (3) age, (4) marital status, (5) work experience, (6) volunteer experience, and (7) educational background (see Appendix H). This information was used to describe the sample under observation (for replication purposes; see Table 6 above).

Evaluation Tools. A brief summary of each of the evaluation tools used in the current study, categorized by Level of evaluation, is provided in Table 7. In addition, a table outlining when each of the measures was administered to trainees, categorized by study phase, is provided in Table 8. It is important to note that Levels 1, 2 and 3 of the evaluation framework do not necessarily equate to the study phases (recruitment, training, and transfer) indicating when evaluation measures were administered. For example, although trainee reactions are a Level 1 measure it is difficult to collect reaction measures to a program before the program actually begins. Instead, more meaningful information may be gleaned during the training program or immediately following its completion. Likewise, although learning outcomes are considered a Level 2 measure, in order to assess change in knowledge or skills it is important to administer these tools both before and after training. A detailed description of the completion rates per evaluation tool (and study phase) can be found in Appendix I. In the following section a description of each measure is provided. The translation department at CSC translated the English version of each measure into French to accommodate any trainees who were French speaking.⁶ The measures are described based on their corresponding Level in the evaluation framework.

⁶ The French measures were also reviewed by bilingual subject matter experts in order to ensure that there were no translation errors, as is frequently the case with specialized training materials.

Table 7

Outline of the Evaluation Measures Used in the Current Study, Categorized According to Level of Evaluation

Level 1 (Reaction)	Level 2 (Learning)	Level 3 (Application/Transfer)
Trainee reaction questionnaire	Pre-tests (Knowledge, Skills & Attitudes)	Trainee self-report ratings on behavioural checklist
	Post-tests (Knowledge, Skills & Attitudes)	Supervisor ratings on behavioural checklist
	Transfer tests (Knowledge, Skills & Attitudes)	Training effectiveness factors, demographic variables & LTSI

Table 8

Outline of the Evaluation Measures Categorized by Study Phase

Recruitment Phase	Training Phase	Transfer Phase
Demographic variables (Level 3)	Trainee reaction questionnaire (Level 1)	Transfer-tests (Knowledge, Skills & Attitudes) (Level 2)
Pre-tests (Knowledge, Skills & Attitudes) (Level 2)	Post-tests (Knowledge, Skills, & Attitudes) (Level 2)	Trainee and supervisor behavioural checklists (Level 3)
		Learning Transfer System Inventory (LTSI) (Level 3)

Level 1 Measure: Trainee reaction questionnaire. Trainee reactions, measured in the form of a self-report questionnaire, were administered after the completion of Stage

3 of CTP (see Appendix J).⁷ The 41-item questionnaire utilized five-point Likert scales ranging from 1 (strongly disagree) to 5 (strongly agree), with the option of 6 (not applicable). The questions related to (1) trainees' reactions to the training materials overall (9-items), (2) the training program overall (11-items), (3) the stages of CTP (15-items; 6 were related to Stage 1, 1 was related to Stage 2, 7 were related to Stage 3, and 1 was related to all stages), and (4) the facilitators overall (6-items).

Given the multi-dimensional nature of reaction-level data (Alliger et al., 1997), and the fact that different reactions appear to possess different levels of power for predicting higher-order training outcomes (Alliger et al., 1997; Warr & Bunce, 1995; Warr et al., 1999), the reaction questionnaire included several types of items that appear to be important (Alliger et al., 1997): general satisfaction with training (25-items), satisfaction with training utility (11-items), and satisfaction with training difficulty (5-items). The questionnaire was developed specifically for this study and therefore no psychometric data are available for it; however, internal consistency scores are provided in the Results section.

In addition to the quantitative items, there were five open-ended questions included in the reaction questionnaire asking about additional issues of concern or suggestions for improvement for each of the stages in CTP as well as for CTP overall.

The results from the open-ended items were analyzed qualitatively and were included to

⁷ Although it is often recommended to collect reaction measures periodically throughout training, especially if it is a lengthy training program (Phillips & Phillips, 2007), this was problematic in the current study. It was simply not possible for the author to attend the in-class training to ensure that reaction measures were administered and completed. In addition, the training program is very intensive, and would have included 36 separate reaction surveys if such measures were collected after every in-class course. It was expected that many participants would choose not to complete these measures on their own, or if they did complete them, they would likely not have completed them in a reliable fashion. Therefore, to ensure the most meaningful assessment of reactions to the program, balanced with ease of implementation, the reaction questionnaire was administered as soon as participants finished the in-class portion of the training program.

provide formative information for CSC and to provide a context for the quantitative responses to the reaction questionnaire. Specifically, content analysis was used to identify themes in the trainees' reactions related to issues and concerns (Burnard, 1996; Jackson & Trochim, 2002; Hsieh & Shannon, 2005). Since the purpose of gathering these comments in the present study was to categorize the responses of trainees, an inductive approach to the analysis was used. Broadly, this approach consisted of three stages: (1) Preparation (determining units of analysis); (2) Organization (open-coding and classification of the units of analysis; and (3) Reporting (Graneheim & Lundman, 2004). The author coded all of the comments. Another independent rater who had been trained to follow the coding procedures coded a random sample of 25% of the data. In order to measure the level of inter rater reliability, the kappa statistic was computed. A kappa statistic of .84 was found, which reflects very good inter rater reliability (Juurlink & Detsky, 2005). The raters discussed any discrepancies and 100% agreement was reached before proceeding with the final analysis.

Level 2 Measures. The Level 2 measures included in the current study relate to the knowledge, skills and attitudes targeted in CTP. For each of these learning areas, baseline levels were obtained via pre-tests administered before training, during what is referred to in the current study as the recruitment phase. Post-training levels were obtained both during CTP (via the formal knowledge-based theory tests completed throughout the program) and immediately following CTP (via a post-test package assessing knowledge, attitudes and skills), during what is referred to as the training phase in the current study.⁸ Lastly, retention levels were obtained with follow-up testing four to

⁸ The majority of the evaluation measures were administered immediately following the completion of Stage 3 (i.e., the in-class portion) of CTP before recruits went into their institutions to complete their on the

six months post-CTP during what is referred to as the transfer phase. The post-tests mirrored the pre-tests given to trainees. A comparison of the test scores across the various phases indicates the degree of learning that occurred with respect to knowledge, skills, and attitudes (these findings will be provided in the Results section). Unfortunately, there are no psychometric data available for the knowledge or skills measures. Limited psychometric data are available for the attitude measures (see below).

Knowledge. There are a total of 17 formal knowledge tests administered throughout CTP (six in Stage 1, one in Stage 2, and 10 in Stage 3) with a total of 376 questions. The knowledge pre-test includes 95 randomly selected questions (26 from Stage 1, 12 from Stage 2, 57 from Stage 3) from all 17 tests (multiple choice, true or false, and fill in the blank). This equates to approximately 25% of the entire question pool. While constructing the knowledge pre-test in this way is not ideal (e.g., some knowledge domains might not be adequately covered), it was not logistically possible to administer a pre-test including all 376 questions as this would have taken too much time. However, great care was taken to ensure that each of the content domains covered and tested during CTP was represented in the knowledge test used for the evaluation.

Currently within CTP, the majority of all formal tests have two versions (A and B). Each version of the tests assesses the same content domains, but they differ in two ways: (1) the order of some questions varies across the versions, and (2) not all questions are exactly the same (although there is a large degree of overlap in the questions contained in each version). Thus, two versions of the pre-tests were also created (A and

job training. The only measures collected throughout the training program were the knowledge tests, which were compiled from all the formal theory tests completed by trainees. Therefore, although this phase of the study is referred to as the "training phase", Level 2 measurement took place immediately following the in-class training portion of CTP.

B). All trainees received a score on the pre-test, with a higher score indicating a greater degree of baseline knowledge. Because the curriculum designers at CSC do not wish to have the knowledge pre-tests released into the public domain due to their protected nature (so as not to provide potential candidates with copies of the questions), these tests cannot be provided in their entirety here. However, several example questions are provided below:

1. The main purpose of a Sally Port is to:
 - a. Search visitor's bags coming in and out of the institution.
 - b. Prevent the escape of an offender.
 - c. Allow vehicles access to the inside of the institution.
2. Within an institution drugs are the number one commodity supplied by criminal organizations.
 - a. True
 - b. False
3. Controlled breathing extends your air supply by lowering your breathing rate. You do this by (fill in the blanks):

Inhaling through your _____ and exhaling through your _____.

The knowledge post-test (versions A and B) was comprised of the same items as the knowledge pre-test; however, the trainees encountered the test items as they progressed through CTP (i.e., when the 17 formal knowledge tests were completed as scheduled). Specifically, once the trainees were finished CTP, their 17 formal tests were examined and a total score for the knowledge post-test was tabulated (by extracting relevant questions from the tests and scoring them). A higher score on the knowledge post-test indicates a greater degree of knowledge acquisition. A larger positive difference between the pre- and post-test indicates a greater degree of learning due to CTP.

In order to measure knowledge retention, a transfer test (versions A and B) was utilized, which was identical to the pre- and post-tests. This knowledge transfer test was administered to trainees four to six months after the completion of CTP. A higher score on the knowledge transfer test indicates a greater degree of knowledge retention. Differences between the post- and transfer test indicates the degree to which learning obtained from CTP was maintained, enhanced, or lost.

Skills. Both soft and hard skills are focused on in CTP and thus both were measured in the current study.⁹ Fortunately, there was an option that already existed at CSC for the pre-test of soft skills. Indeed, the Skills and Abilities Assessment for measuring soft skills is already part of the final phase of recruitment for CTP. It is administered by the recruitment team at CSC as part of the interview process. The trainee completes an 8-minute role play whereby they need to utilize their problem solving and communication skills to deal with an upset individual. Afterwards they are required to write a report outlining the details of the role play. This measure is designed to assess the following soft skills: assessing problematic situations, identifying possible solutions and taking necessary actions; motivating and influencing individuals; communicating effectively both orally and in writing; and summarizing information.¹⁰ While clearly not a comprehensive test of all the soft skills taught in CTP, all of these skills are considered critical abilities for correctional officers and are assumed to form the basis of many skills

⁹ Hard skills are defined as those technical skills related to the core business of an organization and are typically considered easy to train, observe, quantify and measure (Coates, 2006). Soft skills are thought to complement hard skills and enhance interactions on the job (Ramesh & Ramesh, 2010). They are typically hard to train, observe, quantify and measure (Coates, 2006). Soft skills are considered to be interpersonal in nature and encompass communication skills, conflict resolution and negotiation, personal effectiveness, creative problem solving, strategic thinking, team building, influencing skills, etc. (Coates, 2006; Ramesh & Ramesh, 2010).

¹⁰ Due to the protected nature of this assessment tool and its current use in the recruitment of correctional officers by CSC, a copy of the document cannot be shared.

that are targeted in CTP. Thus, these skills should improve as a result of taking part in CTP. All trainees received a test score on the soft skills pre-test (out of 25), determined by the trained recruitment team, with a higher score indicating a greater baseline level of soft skills.

The same tool that was used for the soft skills pre-test (i.e., the Skills and Abilities Assessment) was used for the soft skills post-tests. The post-tests were completed at the end of CTP (i.e., after Stage 3 was completed) and four to six months on the job. It was administered and scored by trained members of the evaluation team (the evaluation team received the same training as the recruitment team in the administration of this tool).¹¹ A higher score on the soft skills post-test indicates a greater degree of skill acquisition. A larger positive difference between the pre- and post-test indicates a greater degree of learning. A higher score on the soft skills transfer test indicates a greater degree of skill retention. Comparisons of the post- and transfer test indicates whether or not the level of skill obtained from CTP was maintained, enhanced, or lost.

As indicated above, although soft skills are an important component of CTP training, so are the hard skills (defined above). The in-class portion of CTP (Stage 3) is designed with a heavy focus on hard skill development (e.g., Frisk Search, Handcuffing, etc.) with hands-on practice and thus, the inclusion of a hard skills assessment was desirable by CSC. The optimal way of assessing practical skills (i.e., hard skills) is by testing them through role play situations rather than surveys or paper-and-pencil tests (Phillips, 1997). However, the implementation of such tests is often difficult and

¹¹ An evaluation team was assembled specifically for the purposes of the CTP Evaluation to assist with the assessment of skills. The evaluation team was comprised of six people plus the author. Of those six people, three were certified CTP instructors, one was an Emergency Response Team instructor and member of the Union and two were members of the Training Design and Development Team.

expensive, and thus they are typically forgone (Phillips, 1997). This would have been the case in the current study if an attempt was made to design a comprehensive and realistic pre-test that assessed the full range of skills covered in CTP. Not only would the time and cost of such a pre-test be prohibitive, but given the types of skills that are taught in certain CTP courses, the test would also raise ethical concerns. For example, to test a trainee's ability to defend himself from physical attack, use appropriate arrest and control maneuvers, or show suitable weapon handling techniques before he/she had actually been trained to do these things would be irresponsible and unsafe. Therefore in the current study, hard skills were assessed after the completion of CTP, at the training and transfer phases only.

It was decided that a similar approach to that taken with the knowledge test would be undertaken for the hard skills, such that a sample of the skills would be examined. Due to the large number of skills covered in Stage 3 (over 60), a small selection of eight hard skills (approximately 13%) was assessed immediately following the completion of CTP and four to six months on the job. Again, it is acknowledged that constructing the hard skills post-test in this way is not ideal (e.g., a number of skill domains could not be covered). However, it was not possible to assess all of the skill areas covered in CTP due to the time that would be needed for testing.

The hard skills that we focused on were selected by the evaluation team in a multi-step process. First, the members were consulted to determine which of the 60 hard skills should be included in the assessment and the length of time needed for assessment (see Appendix K for a copy of the questionnaire sent to the evaluation team members). The responses to the questionnaire were tallied and a list of all skills indicated for

inclusion by all the team members was compiled (resulting in 41 skills). The frequency with which each skill was chosen was recorded and only those skills selected by at least 4 out of the 5 members were retained. This resulted in a total of 11 skills. The list was then sent back to the evaluation team members and the results were discussed as a group to narrow the number down (below 10) and to determine the timing and method of assessment. A consensus was reached easily by removing skills that were similar to one another (e.g., frisk search and strip search) and combining those that naturally occurred together (e.g., both the application and removal of a body belt were chosen as separate skills and it was determined that both skills are taught together and thus were viewed as part of the same skill set). The final list of eight hard skills selected for assessment was: (1) Frisk Search, (2) OC Projector, (3) Two-Way Radio, (4) Twistlock Handcuffing, (5) Low Profile Rear Wristlock Escort, (6) High Profile Rear Wristlock Escort, (7) Application and Removal of Leg Irons and (8) Application and Removal of Body Belt (see Appendix L for a definition of each skill). It was determined that it would take approximately 45 minutes to assess the eight skills and that a scenario-based role play format would be most consistent with the methods used in CTP.

Two scenarios, one that naturally follows the other, were designed for the purposes of the evaluation by the evaluation team. The scenarios mimicked those used in CTP and were designed simply to ensure that each of the skills was displayed by trainees as naturally and realistically as possible. Copies of the scenarios are provided in Appendix M. Because the scenarios were designed as a way to have trainees display the skills, if a trainee chose another skill or an alternate way of handling the scenario, they were given a “time-out” and redirected to display the skill of interest. The instructions

provided to the trainee and the inmate (role played by a member of the evaluation team) are outlined in the scenarios provided in Appendix M.

The assessment tools that are currently used to assess these skills in CTP were also used for the purposes of this study. The tools have two components. The first component is a checklist of the steps required for each skill (i.e., each critical step is listed and trainees get a “check” under one column if they performed it, an “X” if they missed it, or “not applicable” if the particular scenario did not warrant the step, and the check marks are then tallied up). The second part of the assessment is an overall assessment in the form of a rubric. Each of the key assessment criteria for a particular skill are provided and if the skill is not executed effectively, the elements that are missed result in a loss of points. An overall loss of more than five points, or a loss of more than three points in a single key assessment criterion, results in a fail. These assessment tools are also protected by CSC and cannot be shared; however, an example of a few of the items included on the Frisk Search checklist, as well as the overall assessment rubric, are as follows:

RECOMMENDED SEQUENCE OF SKILL PERFORMANCE (Situational factors may require minor modifications of the following recommended sequence)	✓	X	N/A
Check search area and make it safe.			
Instruct offender to remove seasonal clothing (i.e., scarf, hat, parka, etc.).			
Check waistband area. Check outside of front pockets.			
Total # (add up the total number in each column):			

KEY ASSESSMENT CRITERIA	Minus 2 points	Minus 1 point
Performs the frisk search in a systematic and reasonable sequence.	<ul style="list-style-type: none"> • Fails to check the search area and make it safe. 	<ul style="list-style-type: none"> • Unnecessarily repeats search steps. • Search conducted out of sequence.

The two forms of assessment are essentially measuring the same thing – one in the form of points gained on a checklist (i.e., when the components of the skills are executed trainees get check marks, which are added up) and the other in the form of points lost according to a rubric (i.e., if certain aspects of the skill are not performed correctly or steps are missed this results in points deducted). Therefore, in the current study, only the checklist was relied on as the measure of hard skill acquisition and retention. The number of items on each checklist differed for each of the eight skills. Thus, once the check marks were tallied up for each trainee, their scores were calculated in the form of the percentage correct for each skill. The scores for all eight skills were combined to create a composite hard skills score, which was a percentage indicating an overall hard skills score.

Following the selection of the hard skills, a training session with all of the evaluation team members took place to establish consistent rating practices and to ensure a high level of inter rater reliability. During the training session the team members viewed videos of skill demonstrations created by the author. Each of the hard skills was represented by actors that included former correctional officers working in the Security Branch of CSC. One acted as the inmate and the other as the correctional officer demonstrating the application of the skill. The team members viewed the videos as a

group and then completed the assessment tools independently. The responses were then shared and any discrepancies were discussed until a consensus was reached. The intra-class correlation coefficient (ICC) was calculated for the initial responses with a two-way mixed effects model (appropriate when not generalizing to a random sample of raters from the general population; McGraw & Wong, 1996; Shrout & Fleiss, 1979) with a focus on the average measures (used when using multiple raters during data collection as well as for the inter rater reliability assessment; MacLennan, 1993). The average measure ICC in the current study was .95, indicating a high level of agreement among the raters (Shrout & Fleiss, 1979). As mentioned, any discrepancies were discussed resulting in 100% agreement at the end of the training session.

In addition to ensuring consistent rating practices, the evaluation team members reviewed the scenarios and practiced implementing them. A video was made with several of the evaluation team members acting out the scenarios and were kept on hand for reference and review before data collection sessions.

As indicated above, the checklists were used for the hard skills test. This test was completed at the end of CTP (i.e., after Stage 3 is completed) and four to six months on the job. They were administered and scored by at least two members of the evaluation team. A higher score on the hard skills checklists indicates a greater degree of skill acquisition and retention, and any differences between the test at the training and transfer phases indicates whether or not the skill level obtained from CTP was maintained, enhanced, or lost.

Attitudes. CTP also targets trainees' attitudes towards correctional work and inmates themselves. For example, in Stage 1, courses such as Corrections in Canada,

Officer Roles and Responsibilities, Workplace Issues, and The Offender's World strive to educate trainees about what a career in corrections entails, public service values, what to expect on the job, and the general characteristics of offenders. Given this focus on attitudes it was important to assess whether CTP changes the attitudes of trainees.

The attitudes pre-test is comprised of two scales designed to measure attitudes towards correctional work and inmates: the Attitudes Towards Correctional Work scale (ATCW; Robinson, Porporino, & Simourd, 1992) and the Attitudes Towards Prisoners (Inmates) scale (ATI; Melvin, Gramling, & Gardner, 1985) (see Appendices N and O). The ATCW was developed by Robinson et al. (1992) in order to determine general interest in the field of corrections. It currently consists of 12 items (it was originally a 9-item scale) that measure correctional officers' attitudes towards their occupation (e.g., Being involved in the field of corrections gives me a personal sense of pride and accomplishment), as well as their perception of how the public views correctional work and its associated challenges (e.g., For good reasons, the type of work we do in corrections has a bad image with the public). The scale is comprised of both negative and positive items rated as either true (1) or false (0). Higher scores indicate a more positive attitude towards correctional work. In terms of the psychometric data available on this measure, Robinson, Porporino, and Simourd (1993; 1996) report internal consistency scores of .76 for the original 9-item scale and Simourd (1997) reports internal consistency scores of .79 for the 12-item scale.

The ATI was developed by Melvin et al. (1985) and is a 36-item scale with each item rated on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Higher scores indicate a more positive attitude towards inmates. The scale was

modified slightly by Bensimon (2005), whereby the use of the term prisoner was changed to inmate. In the current study the word inmate will also be used. Psychometric data is reported by Melvin et al. (1985) who state that the split-half reliability of the scale is between .84 and .92 in five different samples and the test-retest reliability (pre- and post-test) is .82. In addition, Ortet-Fabregat, Perez, and Lewis (1993) report high levels of internal consistency with alpha levels ranging from .90 to .95.

The attitudes pre-test (ATCW and ATI) was given to trainees during the recruitment phase when they completed the informed consent form, the demographic questionnaire, and the knowledge and soft skills pre-tests, in order to determine baseline levels. The same two attitude scales were presented to the trainees immediately after training (i.e., a post-test after Stage 3 was completed) as well as four to six months on the job (i.e., a transfer test). The transfer test was slightly modified from that used for the pre- and post-test so that the wording of items was changed from prospective to retrospective (see Appendix P). As with the pre-test, higher scores on these scales indicate a greater degree of positive attitude towards correctional work and prison inmates, and larger positive differences between the pre- and post-test indicate a greater degree of positive attitude change. In turn, any differences between the post-test and transfer test indicates whether or not the attitude levels obtained after CTP were maintained, enhanced, or lost.

Level 3 Measures: Trainee and supervisor checklists. Transfer of learning to the job was determined by behavioural checklists completed by trainees and their supervisors four to six months after CTP, once the trainees had been on the job and presumably had a chance to apply what was learned (see Appendices Q and R). These checklists include the behaviours (i.e., skills) that should be observed on the job once

CTP is completed (behaviours that are based largely on the knowledge and skills targeted in CTP). The specific items included in this checklist were determined by examining the critical skills targeted in CTP. Recall that from the soft skills measure these include: (1) assessing problematic situations, identifying possible solutions, and taking necessary measures, (2) motivating and influencing individuals, (3) communicating effectively orally, (4) communicating effectively in writing, and (5) summarizing information. In addition, the broad categories of hard skills covered in the formal skills evaluations conducted within Stage 3 of CTP were also included (e.g., arrest and control skills).

The self-report and supervisor checklists included two components. The first was a measurement of how often the behaviour (i.e., skill) had been exhibited by the trainee (i.e., frequency per month), which is consistent with the method used by Tan et al. (2003). The frequency of behaviour was measured on a five-point scale ranging from 1 (never/no opportunity to use) to 5 (more than 15 times). The second component included in the checklists was a measurement of the quality of each behaviour, in terms of consistency (i.e., whether the behaviour was executed in a way that was consistent with what was taught during CTP). Quality of behaviour was measured on a three-point scale ranging from 1 (below average) to 3 (above average); with an option to indicate that there was no opportunity to use the skill.

The scales were identical for both the trainees and their supervisors with the exception that, for supervisors, an option was included in the frequency and quality scales for “not observed”. Frequency scores were determined by calculating the median and mode of the responses and quality scores were calculated by averaging the scores, excluding all items where there was no opportunity to use/observe the skill. Larger scores

indicate greater use and quality of the skills. There was no psychometric data available for these measures as they were designed for the current study. Internal consistency scores are provided in the Results section below.

In addition to the frequency and quality items, there were a number of other open-ended questions included on the checklist provided to the trainees. These items capture a range of potentially important issues that contribute to or inhibit the application of knowledge and skills (e.g., What barriers, if any, have you encountered that have prevented you from using the skills or knowledge gained in CTP?; Phillips, 1997). The results from these items were analyzed qualitatively in the same way described above for the open-ended comments provided on the trainee reaction questionnaire (Level 1 outcome) and were included mainly to provide formative information to CSC and to provide context to the results found with the behavioural checklists. The author and a single rater followed the content analysis procedures described above with a random sample of 25% of the data. In order to measure the level of inter rater reliability, the kappa statistic was computed. A kappa statistic of .77 was found, which reflects good inter rater reliability (Juurink & Detsky, 2005). The raters discussed any discrepancies and 100% agreement was reached before proceeding with the final analysis.

Factors Influencing Training Effectiveness. As indicated in the literature review, many researchers posit that various factors, including trainee, training, and workplace characteristics can influence the degree to which learning transfers to the workplace (e.g., Axtell et al., 1997; Baldwin & Ford, 1988, Holton, 2005; Tracey et al., 1995). The Learning Transfer System Inventory (LTSI) was used to examine how various factors relating to the trainee, the training program, and the organizational climate

influence training effectiveness (Holton et al., 2000). Trainees completed this questionnaire during the transfer phase, four to six months on the job, at the same time as the post-tests. This time lag was necessary so that the respondents could reflect on various organizational factors that may have impacted their application of the training material, such as supervisor support.

The version of the LSTI used in the current study consisted of 93 items that measured 17 factors (see Table 3 for a list of the factors measured by the LSTI; please contact the author for a copy of the LSTI).¹² Each item was rated on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The LSTI is broken up into two sections, with 67 items measuring 12 factors related to the specific training program (e.g., motivation to transfer, supervisor support, etc.) and 26 items measuring five factors related to training in general (e.g., openness to change, performance self-efficacy, etc.).

Much of the research conducted with the LSTI has focused on expanding and refining the constructs that comprise it, and assessing its psychometric properties. Given its short history, the LSTI has been relatively well tested. Holton (2005) reports individual scale reliabilities with alpha levels ranging from .63 to .91, strong evidence of construct validity (Bookter, 1999; Holton et al., 2000), some evidence of criterion validity (Bates, Holton, Seyler, & Carvalho, 2000; Ruona et al., 2002; Seyler et al., 1998), and some evidence of cross-cultural validity (Chen, 2003; Khasawneh, 2004;

¹² The LSTI was updated since the start of the data collection period, which is why there are 17 scales instead of the 16 outlined in Holton et al. (2000). The individual items of the LSTI are not provided in the literature, only the scale names. Only after contacting Dr. Holton to obtain permission to use this tool, were the items released. Information relating each item to the respective scale was never provided. The data, once collected, was sent directly to Reid Bates, Holton's colleague, for the preliminary analysis as part of the agreement for using this tool. Bates provided the raw data scale scores as well as the reliability statistics. Dr. Bates did not provide any information with respect to the individual items that make up each of the 17 scales.

Yamhill, 2001). The internal consistency scores for the current study are provided in the Results section.

Debriefing Forms. The debriefing form provided to trainees, which reiterates the purpose of the study and provides reasons why the research is important, was provided to trainees at the end of the study (see Appendix S). The debriefing form that was provided to the supervisors is provided in Appendix T.

Procedure

The current study was carried out over several phases, which have been outlined above in the materials section (see Table 8 above). A more detailed description of what occurred during each of those phases is provided below.

The *recruitment phase* is when all potential trainees go through a rigorous and time consuming selection process that determines whether or not they will be invited to attend a CTP. The recruitment phase includes many steps, most of which are not of concern in the present study. During this phase, the interviewees were given an information sheet with an overview of the study. If the potential trainees indicated an interest in participating in the current study they were given the informed consent form, the demographic questionnaire, and the knowledge and attitudes pre-tests.¹³ It took approximately 1-1.5 hours for the potential trainees to complete this phase of the study. All measures other than the soft skills pre-test were administered in a face-to-face format by the author or an approved CSC staff member after potential trainees completed their recruitment interview. The Skills and Abilities Assessment (i.e., the soft skills pre-test)

¹³ The two versions (A and B) of the knowledge tests were randomly distributed to trainees and Staff Colleges were informed so as to ensure they received the corresponding version throughout CTP.

was administered by CSC's recruitment team during the interview and the scores were provided to the author.

Recruiting for CTP is always on-going and happens in waves, depending on organizational needs. Therefore, the pre-testing occurred in waves, in all three regions, until 119 trainees completed the process.¹⁴ The data collection period spanned approximately one and a half years (from June 2009 to March 2011). For a detailed schedule of the data collection process please see Appendix U.

Those trainees who successfully made it through the recruitment phase received an invitation from CSC to attend a CTP and therefore began the *training phase*. Each CTP cohort consisted of a maximum of 24 trainees (attendance typically ranged anywhere from 18-24) and the timing of the delivery of CTP's in each region depended again on organizational needs. Thus, data during the training phase was also collected in waves as each new cohort began and completed their training course (see Appendix U). The author, along with the evaluation team, collected most of the data during this phase in a face-to-face format, which took approximately 2.5-3 hours for the trainees to

¹⁴ Due to logistical constraints it was not always possible to have trainees complete the pre-test evaluation tools during the recruitment phase before receiving an invitation to start CTP (e.g., some trainees indicated an interest in participating in the study but did not have time to stay and complete the pre-test materials at the time of their recruitment interview). Therefore, due to scheduling challenges, some trainees completed the pre-test evaluation tools once they had already begun the online component of CTP (Stage 1). This occurred in a relatively large portion of the sample (approximately 50%, $n = 54$). Therefore, it was necessary to ensure that this did not negatively impact the knowledge outcomes in terms of limiting the learning gains observed between pre- and post-tests. To examine this, the overall knowledge test scores for those trainees who completed the pre-test materials as intended were compared with those who had already started Stage 1 before completing them. Small, but significant differences were found on the test scores at the recruitment phase, such that those completing the pre-test materials after starting Stage 1 scored slightly higher than those who were not exposed to any of the training material (trainees tested before CTP, $M(SD) = 58.09(8.23)$; trainees tested during Stage 1, $M(SD) = 62.34(9.79)$; $t = 2.24$, $df = 100$, $p = .027$, Cohen's $d = -.47$). Of greater interest, however, was whether or not there were any differences in learning gains across the phases of the study (i.e., between the recruitment phase and the training phase as well as between the training phase and the transfer phase) for those who had started CTP before completing the pre-test evaluation tools versus those that did not. No significant differences were found between these two groups of individuals for either of the follow-up comparisons (pre-post, post-transfer). Thus, all the data was retained and analyzed as is.

complete.¹⁵ The measures collected by the author and evaluation team included the attitudes post-test, the Skills and Abilities Assessment (i.e., soft skills post-test) and the hard skills test. At the end of each CTP, the Staff Colleges (i.e., staff or instructors) ensured that the trainee reaction questionnaires were distributed, collected and forwarded to the author. Lastly, copies of the knowledge post-tests (formal tests from Stages 1, 2 and 3) were provided to the author by the Staff Colleges.¹⁶

The *transfer phase* began four to six months after the trainees completed CTP. The main reason for the span of 2 months (i.e., four *to* six months post-CTP) was due to the logistics associated with the co-ordination of both the trainee's schedules as well as the evaluation team member's schedules. Again, the data collection occurred in waves due to the fact that cohorts completed CTP at different times (see Appendix U). During this phase, the author and the evaluation team members administered the transfer tests, which took approximately 3-3.5 hours to complete. The measures that were collected included the knowledge test, the soft and hard skills tests, the attitudes surveys, as well as

¹⁵ There were usually at least 2 members of the evaluation team present for the majority of the data collection during the training and transfer phases. The team members who attended depended on the region and their availability. For some of the sessions towards the end of the study, only one member of the evaluation team accompanied the author due to scheduling difficulties.

¹⁶ Upon receiving the knowledge post-tests it was determined that there were problems with how the test versions (A or B) were sometimes administered during CTP. For example, trainees did not always receive test versions at the Colleges that corresponded to the version they completed as part of the pre-test evaluation, despite the fact that the author was assured that the versions would be matched. In addition, different Colleges used different procedures with regard to the distribution of the different versions of the knowledge test throughout CTP. For example, some trainees only received B versions if they failed a test and had to re-test. Fortunately there was a large overlap in the questions used in each version of the knowledge tests (approximately 65%) and the questions included in each version were similar in terms of difficulty, style, etc. To ensure that this problem would not pose difficulties in this study, the test outcomes for each version of the knowledge tests were compared at each phase. No significant differences were found between them. In addition, to be sure that version discrepancies did not impact the knowledge outcomes, independent samples *t*-tests were run between those who had the same test versions and those who had mixed versions at the training phase (i.e., those who had the same test version for the recruitment and training phases versus those who had differing versions) and the transfer phases (i.e., those who had the same test versions at all three phases versus those who did not). Again, no significant differences were found between those with matching test versions and those with mixed test versions at either the training or transfer phase. Thus, the data for all trainees was retained.

the trainee (self-report) behavioural checklist and the LTSI. The Scheduling Managers at each of the relevant institutions were contacted to coordinate testing times for the trainees and to identify their supervisors. The self-report supervisor behavioral checklist was sent via email to the supervisors that were identified along with an informed consent form. Once all phases of the study were complete, debriefing forms were sent to all trainees and supervisors.

CHAPTER 3

Data Screening and Attrition Analyses

Data Screening

As indicated above, originally there were 119 trainees in the study; however, 17 of those trainees did not go on to complete a CTP and thus were dropped from the analysis. All data were examined for data entry errors as well as missing data. All errors due to data entry procedures were corrected before the missing value analysis (MVA) was completed.

Missing values. Due to the fact that the data was collected over three phases and that all trainees did not complete each phase, the MVA was carried out for each phase separately (pre-CTP, post-CTP, four to six months post-CTP). The MVA revealed that overall, 45% of the sample (46 of the 102 trainees) had some missing data (i.e., 7 trainees from the recruitment phase, 4 from the training phase, and 35 from the transfer phase). The range of missing data for the recruitment phase was from 1.4% (1 item) to 18.8% (13 items). The range for the training phase was .7% (1 item) to 5.8% (8 items). The range for the transfer phase was .3% (1 item) to 9.1% (31 items).

Upon further investigation it was determined that many of the trainees were missing demographic variables, random items, or whole study measures (27 whole measures were missing, mainly consisting of the soft skills measure at the training phase). Most of the whole study measures that were missing resulted from logistical circumstances surrounding data collection (e.g., training from National Headquarters for the evaluation team on the administration of the soft skills evaluation was delayed

beyond the author's control and the data collection process was already underway; thus this measure could not be administered to some trainees).

Further analyses were conducted to determine if the missing data was missing completely at random (MCAR; Allison 2001). A Little's MCAR test was conducted to determine if a pattern could be detected in the missing data. Results for all phases were not significant indicating that the data, at the very least, was missing at random (MAR), or possibly MCAR (Recruitment phase, $\chi^2 = .39$, $df = 1$, $p = .532$; Training phase, $\chi^2 = 11.84$, $df = 9$, $p = .222$; Transfer phase, $\chi^2 = 11.44$, $df = 11$, $p = .407$). In either case, with MAR or MCAR data, the missing data can be ignored (Tabachnick & Fidell, 2007). In addition, *t*-tests for all of the Level 1 through 3 measures (e.g., trainee reactions, knowledge tests, attitudes measures, etc.) revealed no significant differences between those who were and were not missing data. Finally, the patterns of correlations that emerged (across measures, both within and between levels) did not differ across these two groups.

Given the finding that the data was determined to be MAR (at the least), as well as the fact that most trainees (99 %) were not missing more than 10% of their data (only 1 trainee had 18.8% missing data due to the fact that they skipped the back page of one survey), mean substitution was used in order to minimize any loss of data (Tabachnick & Fidell, 2007). Mean substitution involves replacing each missing data point with the mean for that particular variable (Tabachnick & Fiddell, 2007). Although mean substitution can lead to a reduction in variance, regional group means (e.g., for the Pacific region) were used in the current study because it is possible that there were slight differences in the delivery of CTP (or the institutional environments) across regions.

Using group means is not as conservative as using the overall mean. Therefore, the loss in variance associated with mean substitution more generally is not as great in this study (Tabachnick & Fidell, 2007).

Outliers and normality. All continuous variables were screened for univariate outliers. These variables included the Level 1 through 3 training outcomes (including their subsections) across all phases of the study, in addition to the training effectiveness scales of the LTSI. Twenty-two univariate outliers were found for 19 different variables (see Table 9). Outliers were defined as being more than three standard deviations from the mean of a particular variable (i.e., standardized scores greater than ± 3.29 ; Tabachnick & Fidell, 2007). In order to reduce the impact of these items, their raw scores were changed to one unit smaller or larger (depending on the direction of the standardized score) than the next most extreme score for the particular variable of concern (Tabachnick & Fidell, 2007). Grouped data (i.e., by region) was not analyzed separately for outliers as not all proposed analyses were performed using grouped data (e.g., correlational and regression analyses) (Tabachnick & Fidell, 2007).

Table 9

Overview of Outlying Variables

Variable	Study Phase	Number of Outliers
Trainee Reaction Overall	Training	1
Trainee Reaction (Training Materials Overall Subscale)	Training	1
Trainee Reaction (Training Program Overall Subscale)	Training	2
Trainee Reaction (Facilitators Overall Subscale)	Training	1
Trainee Reaction (General Satisfaction Item Types)	Training	1
Trainee Reaction (Difficulty Item Types)	Training	1
Knowledge Test Score (Stage 3 Subsection)	Recruitment	1
Knowledge Test Score	Training	1
Knowledge Test Score	Transfer	1
Hard Skills (OC Checklist)	Training	1
Hard Skills (Radio Checklist)	Training	2
Hard Skills (Body Belt Checklist)	Training	1
Hard Skills (OC Checklist)	Transfer	1
Hard Skills Composite Score	Transfer	1
ATCW	Recruitment	1
ATCW	Training	1
ATCW	Transfer	1
LTSI (Supervisor Opposition)	Transfer	1
LTSI (Performance Coaching)	Transfer	1

Note. All z scores for the outliers were negative in direction except for the LTSI (Supervisor Opposition) subscale, which was positive.

Next the data were examined for multivariate outliers using the Mahalanobis distance test. The Mahalanobis distance was compared with the critical value of $\chi^2 = 124.84$ ($df = 87, p < .001$). Using this critical value no multivariate outliers were identified.

All continuous variables were also examined for normality via the skewness and kurtosis values. These analyses were conducted separately for each region. Both the skewness and kurtosis values were changed to standardized z scores and examined for significance. Several variables were significantly skewed and kurtotic at the $p < .01$ and $p < .001$ levels and thus were transformed. These variables are listed in Table 10. The proposed analyses of t -tests, mixed Analysis of Variance (ANOVAs) and correlations were conducted using the transformed and untransformed variables. No differences were found across these analyses. Therefore, for ease of interpretation, non-transformed variables were used for all analyses.

Table 10

Significantly Skewed and Kurtotic Variables

Variables	Study Phase	Skewness	Transformation	Kurtosis
Pacific				
Trainee Reaction (Training Program Overall Subscale) ^a	Training	4.32***	Reflect & Square Root ^b	5.43***
Trainee Reaction (Facilitators Overall Subscale) ^a	Training	5.75***	Reflect & Logarithm ^c	8.30***
Knowledge Test Score (Stage 1 Subsection) ^a	Recruitment	2.77**	Reflect & Square Root	2.85**
Knowledge Test Score (Stage 3 Subsection) ^a	Recruitment	2.84**	Reflect & Square Root ^b	-
Soft Skills (Role Play Subsection)	Transfer	2.72***	Reflect & Square Root ^b	-
ATCW ^a	Recruitment	3.70***	Reflect & Square Root ^b	2.81**
ATCW ^a	Training	5.00***	Reflect & Logarithm ^c	5.91***
ATCW ^a	Transfer	4.11***	Reflect & Square Root ^b	5.20***
Trainee Behavioural Checklist (Frequency Subsection)	Transfer	2.72***	Square Root	-
LTSI (Performance Coaching)	Transfer	3.04***	Reflect & Square Root ^b	-
Ontario				
Trainee Reactions Overall ^a	Training	9.60***	Reflect & Logarithm ^c	22.82***
Trainee Reaction (Training Materials Overall Subscale) ^a	Training	10.06***	Reflect & Inverse ^d	25.38***
Trainee Reaction (Training Program Overall Subscale) ^a	Training	8.67***	Reflect & Logarithm ^c	18.76***
Trainee Reaction (Stages Overall Subscale) ^a	Training	10.42***	Reflect & Inverse ^d	25.38***
Trainee Reaction (Facilitators Overall Subscale) ^a	Training	9.21***	Reflect & Logarithm ^c	21.00***
Trainee Reaction (General Satisfaction Item Type) ^a	Training	9.57***	Reflect & Logarithm ^c	22.56***
Trainee Reaction (Utility Item Type) ^a	Training	10.46***	Reflect & Inverse ^d	25.96***
Trainee Reaction (Difficulty Item Type) ^a	Training	8.06***	Reflect & Logarithm ^c	16.60***
Hard Skills Composite Score	Training	2.60**	Reflect & Square Root ^b	-
ATCW	Recruitment	4.11***	Reflect & Square Root ^b	-

Note. One variable was not skewed but was leptokurtic in the Pacific region: Knowledge Test, Stage 2 Subsection, Recruitment Phase (2.94**). One other variable was not skewed but was leptokurtic in the Ontario region: Supervisor Opposition subscale of the LTSI (3.43***). All variables were negatively skewed, except one - Behavioural Checklist (Skills Frequency), which was slightly positively skewed.

^aIndicates variables that were also significantly leptokurtic; ^bindicates slight skewness; ^cindicates moderate skewness; ^dindicates severe skewness.

** $p < .01$; *** $p < .001$.

Typically the last step in the data screening process is to assess multicollinearity. Since the final set of variables that will be used in the regression analyses will be determined by the results of the outcome analyses (i.e., findings from the correlation matrices), these results will be presented below.

Attrition Analyses

As indicated above, the size of the sample varied between the different phases of the study. Recall that there were 102 trainees who completed the recruitment phase, 68 who completed the training phase and 53 who completed the transfer phase (see Table 5). Attrition analyses were conducted to determine if there were any differences in outcomes and effectiveness factors between those who did and those who did not complete the different phases of the study.

Attrition between the recruitment and training phases. As indicated above, 68 trainees completed both the recruitment and training phases, compared to 34 who only completed the recruitment phase (see Table 5). Independent samples *t*-tests (for continuous variables) as well as Chi Square tests (for categorical variables) were run (each with appropriate Bonferroni corrections) to determine if there were any differences between these groups on the pre-test evaluation measures and the demographic variables. Specifically, the measures examined included: Region, Gender, Ethnicity, Age, Marital Status, Education, Work Experience, Knowledge Test Score, Soft Skills Score and Attitudes (ATCW and ATI). The only significant finding was for Region, $\chi^2(2, N = 102) = 16.69, p = .000$. Specifically, significantly more trainees in the Pacific Region went on to complete the training phase than the other two regions (see Table 11).

There are likely two contributing factors for the regional differences in attrition rates between the recruitment and training phases of the study. One main factor was the timing between Stage 3 (in-class) and Stage 4 (on the job training) for evaluation purposes. CSC's scheduling process, for a large portion of the CTP's that were part of the evaluation, left very little, if any, time between stages to complete the post-testing evaluation tools (for the study only). For example, Stage 3 would end on a Friday at 4:30pm and Stage 4 would begin on the following Monday, which proved difficult for trainees to stay and participate in the study (e.g., some trainees had to travel home to collect their belongings before taking up their job). Related to this first factor was the lack of College-level support, which was found mostly in the Ontario Region, for dealing with scheduling issues related to the evaluation. In light of these difficulties, attempts were made to have trainees complete the evaluation session for the training phase during the last week of Stage 3 instead of the last day (e.g., during breaks, down time, etc.), or during the first few days of Stage 4 at the institutions. This procedure was supported more readily in the Pacific Region, resulting in the retention of more trainees at the training phase.

Table 11

Attrition Between the Study Phases

Region	Number of Trainees Lost Between the Recruitment and Training Phases % (n)	Number of Trainees Lost Between the Training and Transfer Phases % (n)
Pacific	1 (3)	24 (8)
Ontario	47 (24)	11 (3)
Quebec	50 (7)	57 (4)
Total	33 (34)	22 (15)

Attrition between the training and transfer phases. As indicated in Table 5, 53 of the 68 trainees who completed the training phase went on to complete the transfer phase (also see Table 11). Independent samples *t*-tests as well as Chi Square tests (with appropriate Bonferroni corrections) were run to determine if there were any differences between those 53 trainees who completed all three phases (recruitment, training and transfer) and those 15 who stopped before the transfer phase (i.e., those who only completed the recruitment and training phases). Specifically, the measures examined included the demographic variables (Region, Gender, Ethnicity, Age, Marital Status, Education, and Work Experience) and several evaluation outcomes from the recruitment and training phases (Knowledge Test Scores, Soft Skills Scores, Hard Skills Score and Attitudes (ATCW & ATI).

No significant differences were found after the Bonferroni correction was applied. Region was the only variable that would have reached significance, $\chi^2(2, N = 68) = 6.94$, $p = .031$, with less attrition observed this time in the Ontario Region. One possible reason that more attrition occurred in the Pacific Region between these study phases was a slight

issue in terms of the institutional climate at one institution. Specifically, at one institution in the Pacific Region, other officers (as well as the local union members representing correctional officers) appeared to be pressuring trainees to cease their involvement in the evaluation because it was rumored to be a “rat-line” to National Headquarters. According to some of the trainees, this sentiment was also passed to another institution located in close proximity to the institution where these rumors originated. Efforts were made by the author (with the assistance of the Staff College) to meet with the Regional Union President to quell any misconceptions regarding the purpose of the evaluation. This seemed to help and the loss of trainees was minimized.

CHAPTER 4

An Analysis of Training Outcomes Included in the Expanded Evaluation Framework

The first set of analyses in the current study examined the training outcomes associated with CTP across all three levels of the Kirkpatrick's expanded training evaluation framework. Recall that Level 1 (reactions) was measured using the trainee reaction questionnaire; Level 2 (learning) was measured using the knowledge, skills and attitudes tests at the recruitment, training and transfer phases; and Level 3 (behaviour) was measured using the trainee and supervisor behavioural checklists. To begin the process of accumulating psychometric data for these measures, and to inform further analyses, internal consistency scores for any applicable measures are reported, as are the correlations between each of the measures within each level. In addition, descriptive statistics are provided for each of the measures designed to evaluate Levels 1 through 3.

Hypotheses

Recall from Chapter 1 that the following hypotheses were tested with respect to training outcomes:

- (1) For the Level 1 outcome, positive ratings of the training program were expected regardless of where the training took place (i.e., regardless of region).
- (2) For the Level 2 outcomes, significant increases in test scores, regardless of test subsections, individual skills or regions, were expected across each of the training outcomes between the recruitment and training phases (i.e., pre-post tests), indicating that learning had occurred. In addition, it was expected that most of what was learned would be retained during the transfer phase (post-transfer tests);

though a small degree of forgetting might be expected. It was also expected that at each study phase the different Level 2 measures would be positively correlated with each other (e.g., knowledge and skills, etc.) and that for those Level 2 outcomes with subsections (e.g., knowledge and soft skills), the subsections would be positively correlated with the overall score. Lastly, it was expected that the same Level 2 measures would be positively correlated across study phases (i.e., that high levels of knowledge acquisition at the training phase would be related to high levels of knowledge retention at the transfer phase, etc.).

- (3) For the Level 3 outcomes of frequency and quality of skill use on the job, it was expected that all skills targeted in CTP would be used within the institutions and that they would be used in a way that was consistent with how CTP trains its trainees to use the skills (i.e., frequency ratings and quality ratings will be relatively high, regardless of who is providing the ratings). Regional differences were not expected. In addition, it was expected that the trainee and supervisor ratings would be positively correlated.

Level 1 Outcome

The number of trainees who completed the trainee reaction questionnaire was 67 (Pacific = 34, Ontario = 33). Recall that this measure is completed at the end of CTP, during the training phase. Some trainees chose not to complete this measure ($n = 21$) while those in the Quebec region were not given the opportunity ($n = 14$). The reason no data is available from the Quebec region is because the instructors in that region failed to hand out the questionnaires to the trainees after the completion of CTP.

Bivariate correlations (Pearson's r) were calculated between (1) the total trainee reaction score, (2) the six subsections of the reaction questionnaire (Training Materials Overall, Training Program Overall, Stages of CTP – Stage 1 and Stage 3, Facilitators Overall), and (3) the three item types (General Satisfaction, Training Utility, Training Difficulty). These correlations are presented in Table 12. All of the variables were significantly correlated at the $p < .001$ level, with the majority of comparisons resulting in high positive correlations ($>.80$). Although each subscale is intended to measure a different aspect of the training program, the high correlations between subscales suggest that they are multicollinear (i.e., that they are all potentially tapping into the same construct) and are best examined as a composite (i.e., overall) score (Tabachnick & Fidell, 2007). Similarly, for item types, the high positive correlations suggest that further analyses should not separate items by type.

The internal reliability scores were very high for the trainee reaction questionnaire (overall) as well as for the subsections (George & Mallery, 2003), with scores ranging from .91 to .99. The items representing each of the subsections and item types are highlighted in Appendix J. These scores, along with the descriptive statistics for each section of the questionnaire, are provided in Table 13 (for the average responses at the individual item level please see Appendices V-X).

Consistent with our hypothesis, overall, the mean reaction scores (out of 5) were quite high indicating that, for those who completed this measure, trainees reacted very positively to the training materials, the program, the various stages of CTP, and the facilitators. In terms of the item types, trainees had high levels of general satisfaction with the training, they felt the training material was very useful, and they did not perceive

that the training was too difficult. Also as expected, an independent samples *t*-test revealed that there were no significant regional differences in trainee reaction ratings overall, $t(65) = -.03, p = .976, d = -.02, 95\% \text{ CI } [-.21, .02]$.¹⁷

¹⁷ An examination of regional differences between the various subscales and item types was also going to be undertaken using multivariate analyses of variance (MANOVA). However, due to the high positive correlations (reported in the next paragraph) between all of the dependent variables (DVs) these analyses were not considered appropriate. The information provided by one DV appears to be essentially the same as that provided by one or more of the other DVs. This would make the interpretation of such analyses ambiguous (Tabachnick & Fidell, 2007). Similarly, univariate tests are misleading with such variables and thus, Tabachnick and Fidell (2007) recommend that a composite score (or the most reliable DV) be analyzed using univariate analysis. In the current study, the composite score was used (trainee reactions overall).

Table 12

Correlation Matrix for the Level 1 Trainee Reaction Outcomes (Including Subsections)

Variables	1	2	3	4	5	6	7	8	9	10
1. Trainee Reactions Overall	-									
2. Training Materials Overall	.93	-								
3. Training Program Overall	.90	.82	-							
4. Stages of CTP Overall	.86	.79	.72	-						
5. Stage 1	.75	.70	.82	.73	-					
6. Stage 3	.88	.83	.90	.75	.81	-				
7. Facilitators Overall	.81	.68	.86	.61	.70	.80	-			
8. General Satisfaction Items	.98	.90	.90	.91	.78	.89	.81	-		
9. Utility Items	.82	.74	.82	.72	.70	.77	.70	.82	-	
10. Difficulty Items ^a	.90	.86	.90	.77	.72	.80	.79	.89	.77	-

Note. With a Bonferroni correction = $.05/34 = .0015$, all correlations are significant at the corrected alpha level.

^aRecall that difficulty items were reverse scored so higher scores correspond to ratings of less difficulty.

Table 13

Descriptive Statistics for the Trainee Reaction Questionnaire

<i>Measure</i>	<i>Overall (n = 67)</i>			<i>Pacific (n = 34)</i>			<i>Ontario (n = 33)</i>		
	<i>M (SD)</i>	<i>Range</i>	<i>Alpha</i>	<i>M (SD)</i>	<i>Range</i>	<i>Alpha</i>	<i>M (SD)</i>	<i>Range</i>	<i>Alpha</i>
Trainee Reactions Overall	4.59 (.41)	3.00-5.00	.98	4.59 (.37)	3.00-5.00	.97	4.59 (.46)	3.00-5.00	.99
Training Materials Overall	4.57 (.40)	3.50-5.00	.93	4.58 (.39)	3.50-5.00	.94	4.57 (.41)	3.50-5.00	.93
Training Program Overall	4.55 (.52)	2.50-5.00	.95	4.54 (.51)	2.50-5.00	.95	4.54 (.54)	2.50-5.00	.95
Stages of CTP Overall	4.62 (.45)	3.20-5.00	.96	4.53 (.45)	3.20-5.00	.95	4.71 (.42)	3.20-5.00	.96
Stage 1	4.37 (.63)	1.00-5.00	.95	4.25 (.80)	1.00-5.00	.96	4.48 (.48)	3.00-5.00	.91
Stage 3 ^a	4.58 (.45)	2.86-5.00	.91	4.55 (.50)	2.86-5.00	.91	4.61 (.45)	3.29-5.00	.92
Facilitators Overall	4.67 (.47)	2.75-5.00	.95	4.72 (.45)	2.75-5.00	.94	4.62 (.50)	2.75-5.00	.95
General Satisfaction Items	4.59 (.43)	3.00-5.00	.97	4.57 (.39)	3.00-5.00	.96	4.61 (.47)	3.00-5.00	.97
Training Utility Items	4.64 (.42)	3.18-5.00	.92	4.58 (.42)	3.18-5.00	.91	4.71 (.42)	3.18-5.00	.96
Training Difficulty Items ^b	4.51 (.51)	2.29-5.00	.94	4.56 (.43)	2.29-5.00	.92	4.45 (.58)	2.29-5.00	.92

Note. Recall trainee reactions were measured on a five-point Likert scale, with 5 being more favourable.

^aAs outlined in Chapter 2, only 1 item on the questionnaire pertained to Stage 2, therefore, this item for this Stage was included in the Stages of CTP Overall but not analyzed as a separate category. ^bThe difficulty items were reverse scored such that higher scores indicate less perceived difficulty.

Qualitative results. Following the multiple-choice questions on the reaction questionnaire, trainees were asked about their experiences using open-ended questions. Specifically, trainees were asked to elaborate on any issues/concerns and suggestions for improvement for each of the three stages of CTP as well as CTP overall. This information was gathered and categorized for formative purposes only. Since no additional analyses were planned with the qualitative results for the current study, a breakdown of the coding categories (as well as the specific comments provided by trainees) is provided in Appendix Y. Overall, based on the content analysis described above, answers to the open-ended questions reflected what was found for the quantitative items: trainees indicated a high level of satisfaction with most elements of the training program including the program overall and the instructors. Several suggestions for improvement were provided including more time for the online portion (Stage 1) and more skills practice during the in-class portion (Stage 3).

Level 2 Outcomes

Level 2 outcomes included all of the knowledge, skills and attitudes measures administered across the three study phases. The descriptive statistics for each outcome are presented below, but first the bivariate correlations (Pearson's r) between the Level 2 outcome measures are presented. The correlations were calculated separately for each study phase (i.e., recruitment, training, and transfer) because the learning outcomes are examined separately for each study phase.

The correlation matrix for the recruitment phase is provided in Table 14 (recall that the hard skills were not evaluated in this phase). After the Bonferroni correction was applied, some significant correlations were still found. Interestingly, and in contrast to

expectations, no significant correlations were found between the different Level 2 measures during the recruitment phase (e.g., knowledge, skills and attitudes) suggesting that they are in fact measuring different constructs. As expected, the subsections of the knowledge pre-test (stages 1 through 3) were moderately to highly correlated with the overall score on the knowledge pre-test. Stage 1 of the knowledge pre-test was also correlated with Stage 2. Similarly, the subsections of the soft skills measure (the skills and abilities assessment) were moderately to highly correlated with the overall measure of soft skills.

The correlation matrices for the training and transfer phases are provided in Tables 15 and 16, respectively. Like the recruitment phase, and in contrast to expectations, no significant correlations were found between the different Level 2 measures (e.g., knowledge, skills and attitudes) in either the training or the transfer phase. Similar to the results found for the recruitment phase, and in-line with expectations, the subsections of the knowledge test (Stages 1 through 3) were highly correlated with overall scores for the knowledge tests in both the training and transfer phases. The subsections of the soft skills measure (the skills and abilities assessment) were also highly correlated with the overall measure of soft skills in both the training and transfer phases. In addition, in the training phase, the subsections of the soft skill measure were moderately correlated with each other.

Also of interest are the correlations among the Level 2 measures across the training phases (i.e., are baseline levels of knowledge, skills, and attitudes at the recruitment phase related to acquisition levels at the training phase, and are acquisition levels at the training phase related to retention levels at the transfer phase). For the sake

of brevity and space, the results of the correlations among Level 2 measures across study phases is provided below in Table 30, within Chapter 5 where results of correlations between levels (e.g., Level 2 and 3, etc.) are presented. Surprisingly, only two significant correlations were found. Specifically, significant positive relationships were found between knowledge test scores at the training and transfer phases ($r = .54$) and between attitudes towards inmates at the training and transfer phases ($r = .51$).

Table 14

Correlation Matrix for the Level 2 Outcome Measures During the Recruitment Phase

Variables	1	2	3	4	5	6	7	8	9
1. Knowledge Test	-								
2. Knowledge Test (Stage 1)	.47*	-							
3. Knowledge Test (Stage 2)	.53*	.40*	-						
4. Knowledge Test (Stage 3)	.91*	.11	.27	-					
5. Soft Skills Assessment	.10	.13	.03	.07	-				
6. Soft Skills Assessment (Role Play)	.08	.00	-.03	.10	.89*	-			
7. Soft Skills Assessment (Written)	.09	.27	.10	-.02	.67*	.25	-		
8. ATCW	.04	-.07	.03	.08	-.16	-.10	-.19	-	
9. ATI	-.01	-.21	-.21	.09	.13	.15	.04	.17	-

Note. Bonferroni correction = $.05/36 = .0014$.

*Correlations significant at the corrected alpha level.

Table 15

Correlation Matrix for the Level 2 Outcome Measures During the Training Phase

Variables	1	2	3	4	5	6	7	8	9	10
1. Knowledge Test	-									
2. Knowledge Test (Stage 1)	.49*	-								
3. Knowledge Test (Stage 2)	.35*	.01	-							
4. Knowledge Test (Stage 3)	.80*	.05	.01	-						
5. Soft Skills Assessment	-.24	-.14	-.27	-.13	-					
6. Soft Skills Assessment (Role Play)	-.20	-.04	-.32	-.12	.92*	-				
7. Soft Skills Assessment (Written)	-.20	-.25	-.09	-.10	.76*	.43*	-			
8. ATCW	.12	.16	-.22	.14	.01	.01	.02	-		
9. ATI	-.16	.01	-.17	-.11	-.03	-.05	.00	.15	-	
10. Hard Skills	-.08	-.18	.01	-.01	.20	.14	.22	-.02	-.02	-

Note. Bonferroni correction = $.05/45 = .001$.

*Correlations significant at the corrected alpha level.

Table 16

Correlation Matrix for the Level 2 Outcome Measures During the Transfer Phase

Variables	1	2	3	4	5	6	7	8	9	10
1. Knowledge Test	-									
2. Knowledge Test (Stage 1)	.57*	-								
3. Knowledge Test (Stage 2)	.47*	.21	-							
4. Knowledge Test (Stage 3)	.89*	.32	.33	-						
5. Soft Skills Assessment	-.00	.04	-.03	.00	-					
6. Soft Skills Assessment (Role Play)	-.07	-.00	-.09	-.08	.89*	-				
7. Soft Skills Assessment (Written)	.11	.09	.07	.13	.72*	.31	-			
8. ATCW	-.03	-.04	.20	-.06	.18	.14	.16	-		
9. ATI	-.02	.03	.13	.00	.02	-.06	.14	.29	-	
10. Hard Skills	.22	-.03	.08	.23	-.00	-.02	.03	-.17	-.07	-

Note. Bonferroni correction = $.05/45 = .001$.

*Correlations significant at the corrected alpha level.

Knowledge. The descriptive statistics for the knowledge outcomes (test scores presented as percentage correct) for each of the three study phases (recruitment, training and transfer) are presented in Table 17.¹⁸ The results are presented for the test as a whole as well as for each subsection of the test. The subsections correspond to content related to each of the three stages of CTP (recall that Stage 1 is completed online, Stage 2 is completed at home, and Stage 3 is completed in-class).

For the most part, the expected findings were observed with the knowledge tests, with a lower average score at the recruitment phase (61%) and a higher average score at the training phase (85%), representing an average knowledge gain of approximately 24%. In terms of the average baseline level of knowledge coming into CTP, the score is relatively high (> 50%). As indicated above, this is likely due to the fact that approximately half of the sample had started Stage 1 of CTP before completing the pre-test package.

Also as expected, when looking at retention levels, the average scores on the knowledge test decreased between the post-test (85%) and the transfer test completed four to six months post-CTP (80%). While some degradation of knowledge is expected over time, significant losses are of course not desirable. Further analyses were carried out to determine if these gains and losses over the phases of the study were indeed significant. The results from these analyses are presented below.

¹⁸ The Level 2 descriptive statistics for the Quebec region are provided in Appendix AA. The Quebec region was not included in any regional inferential analyses due to the very low sample size of this group. Without equal sample sizes across groups for analysis of variance (ANOVA) the assumption of homogeneity of variance is violated (Field, 2009).

Table 17

Descriptive Statistics for the Knowledge Outcomes (% Correct)

Study Phase	Knowledge Outcome	Overall			Pacific			Ontario		
		<i>n</i>	<i>M</i> (<i>SD</i>)	<i>Range</i>	<i>n</i>	<i>M</i> (<i>SD</i>)	<i>Range</i>	<i>n</i>	<i>M</i> (<i>SD</i>)	<i>Range</i>
Recruitment Phase	Knowledge Test	102	60.75 (9.43)	30.53- 78.95	37	62.79 (8.55)	40.00- 75.79	51	60.21 (9.09)	36.84- 78.95
	Stage 1	102	77.83 (12.25)	38.46- 100.00	37	78.17 (11.96)	38.46- 96.15	51	75.56 (12.89)	46.15- 100.00
	Stage 2	102	73.53 (12.60)	33.33- 100.00	37	75.00 (12.73)	33.33- 100.00	51	72.71 (13.34)	50.00- 91.67
	Stage 3	102	47.99 (12.33)	8.84- 68.42	37	50.43 (11.40)	22.95- 68.42	51	48.38 (10.94)	15.79- 68.42
Training Phase	Knowledge Test	83	84.91 (3.72)	74.76- 93.94	34	86.04 (4.70)	74.76- 93.94	39	84.69 (2.35)	79.80- 90.91
	Stage 1	83	89.66 (6.55)	69.23- 100.00	34	92.99 (5.56)	80.77- 100.00	39	88.06 (6.59)	69.23- 100.00
	Stage 2	83	85.86 (9.62)	66.67- 100.00	34	86.76 (11.44)	66.67- 100.00	39	85.09 (8.79)	66.67- 100.00
	Stage 3	83	82.73 (4.54)	70.00- 91.80	34	83.02 (5.57)	70.00- 91.80	39	83.18 (3.34)	75.41- 90.16
Transfer Phase	Knowledge Test	53	80.10 (6.05)	63.16- 88.42	26	79.11 (7.55)	63.16- 88.42	24	81.97 (3.06)	76.84- 87.37
	Stage 1	53	85.85 (6.89)	65.38- 96.15	26	85.50 (8.31)	65.38- 96.15	24	86.54 (5.29)	76.92- 96.15
	Stage 2	53	81.86 (7.65)	65.67- 91.67	26	82.65 (7.51)	65.67- 91.67	24	82.20 (7.08)	66.67- 91.67

	Overall			Pacific			Ontario		
	<i>n</i>	<i>M</i> (<i>SD</i>)	<i>Range</i>	<i>n</i>	<i>M</i> (<i>SD</i>)	<i>Range</i>	<i>n</i>	<i>M</i> (<i>SD</i>)	<i>Range</i>
Stage 3	53	77.77 (6.98)	56.14- 87.72	26	76.52 (8.27)	56.14- 87.72	24	80.14 (4.31)	68.42- 87.72

Note. Data for the Quebec region is included in the columns reflecting the sample overall, data for the region itself is provided in Appendix Z.

In order to determine whether or not there were regional differences in knowledge outcomes during each of the phases, as well as to determine the degree of learning that took place across the study phases, a 2 x 3 mixed ANOVA was carried out. The between subjects factor was Region (Pacific and Ontario) and the within subjects factor was Study Phase (recruitment, training and transfer). The dependent variable in this analysis was overall knowledge test scores.

Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(2) = 24.25, p < .001$, therefore degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity ($\epsilon = .71$).¹⁹ As expected, the results show that there was a significant main effect of Study Phase such that knowledge scores significantly differed based on when the test was completed, $F(1.43, 68.42) = 237.99, p = .000$, partial $\eta^2 = .83$, Obs. Power = 1.00. The main effect of Region ($F(1, 48) = .00, p = .966$, partial $\eta^2 = .83$, Obs. Power = 1.00) as well as the Region x Study Phase interaction ($F(2, 96) = 2.59, p = .080$, partial $\eta^2 = .05$, Obs. Power = .51) was not significant, meaning that the test scores were generally consistent regardless of the region where CTP was completed.

Post hoc tests using the Bonferroni correction revealed that the test scores at the recruitment phase ($M = 62.55, SD = 9.26$) were significantly lower than the scores obtained at the training ($M = 85.47, SD = 3.88$) and transfer phases ($M = 80.48, SD = 5.97$) ($p = .000$). In addition, the scores obtained at the training phase were significantly higher than the scores obtained at the transfer phase ($p = .000$). In other words, significant learning gains were observed at the training phase, while significant losses were observed at the transfer phase.

¹⁹ When the sphericity assumption is violated, typically the Greenhouse-Geisser is the recommended correction for estimated epsilon values less than .75 and the Huynh-Feldt is recommended when estimated epsilon values are greater than .75. However, they both produce very similar corrections (Field, 2009).

Given that the subsections of the knowledge test (Stage 1, 2 and 3) are not independent from the overall score, they had to be examined separately. Therefore, in order to determine whether or not the same pattern of learning was found for each of the subsections of the test across study phases, as well as to examine whether or not there were differences in scores amongst the subsections within each study phase, a $2 \times 3 \times 3$ mixed ANOVA was carried out. The between subjects factor was Region (Pacific and Ontario) and the two within subjects factors were Study Phase (recruitment, training and transfer) and Test Subsection (Stage 1, 2 and 3). The dependent variable was the knowledge test scores.

As expected, the main effect of Region was not significant, $F(1, 48) = .02, p = .887$, partial $\eta^2 = .00$, Obs. Power = .05. While a significant main effect of the Study Phase ($F(1.52, 72.80) = 119.26, p = .000$, partial $\eta^2 = .71$, Obs. Power = 1.00²⁰) as well as Test Subsection ($F(2, 96) = 87.78, p = .000$, partial $\eta^2 = .65$, Obs. Power = 1.00) was observed. The majority of the two-way interactions including Region x Study Phase ($F(2, 96) = 1.32, p = .272$, partial $\eta^2 = .03$, Obs. Power = .28), Region x Test Subsection ($F(2, 96) = .13, p = .881$, partial $\eta^2 = .00$, Obs. Power = .07), as well as the three-way interaction of Region x Study Phase x Test Subsection ($F(4, 192) = 2.26, p = .064$, partial $\eta^2 = .05$, Obs. Power = .65), were not significant. Only one significant two-way interaction effect was found. This was for Study Phase x Test Subsection, $F(3.30, 158.57) = 47.39, p = .000$, partial $\eta^2 = .50$, Obs. Power = 1.00.²¹ As illustrated in Figure 1, this

²⁰ Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(2) = 18.04, p < .001$, therefore degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity ($\epsilon = .76$).

²¹ Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(9) = 23.26, p = .006$, therefore degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity ($\epsilon = .83$).

indicates that some of the subsection test scores differed from one another for at least one of the phases of the study.

To examine this interaction further, simple effects analyses were carried out, whereby differences in test scores across each subsection of the test were examined for each phase separately. Although subsection differences were not expected, these analyses indicated that the interaction between study phase and test subsection can largely be explained by the significant difference in test scores found at the recruitment phase, $F(2, 174) = 188.00$, $p = .000$, partial $\eta^2 = .68$, Obs. Power = 1.00, particularly the significantly lower scores obtained at Stage 3 ($M = 50.23$, $SD = 11.08$) compared to Stage 1 ($M = 78.85$, $SD = 13.49$) and 2 ($M = 75.33$, $SD = 13.78$; p 's = .000). The differences between the three stages during the other phases were far less obvious, though these differences were also significant. Specifically, the results at the training phase, $F(1.70, 122.07) = 18.01$, $p = .000$, partial $\eta^2 = .20$, Obs. Power = .99, indicated differences between Stage 1 ($M = 90.15$, $SD = 7.17$) and 2 ($M = 85.50$, $SD = 10.76$; $p = .006$) and between Stage 1 and 3 ($M = 83.63$, $SD = 4.37$) ($p = .000$).²² The results at the transfer phase, $F(2, 98) = 20.44$, $p = .000$, partial $\eta^2 = .29$, Obs. Power = 1.00, also indicated significant differences between Stage 1 and 2 ($p = .022$), Stage 1 ($M = 86.00$, $SD = 6.97$) and 3 ($M = 78.26$, $SD = 6.86$; $p = .000$), and Stage 2 ($M = 82.43$, $SD = 7.24$) and 3 ($p = .003$).

²² Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(2) = 14.06$, $p = .001$, therefore degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity ($\epsilon = .85$).

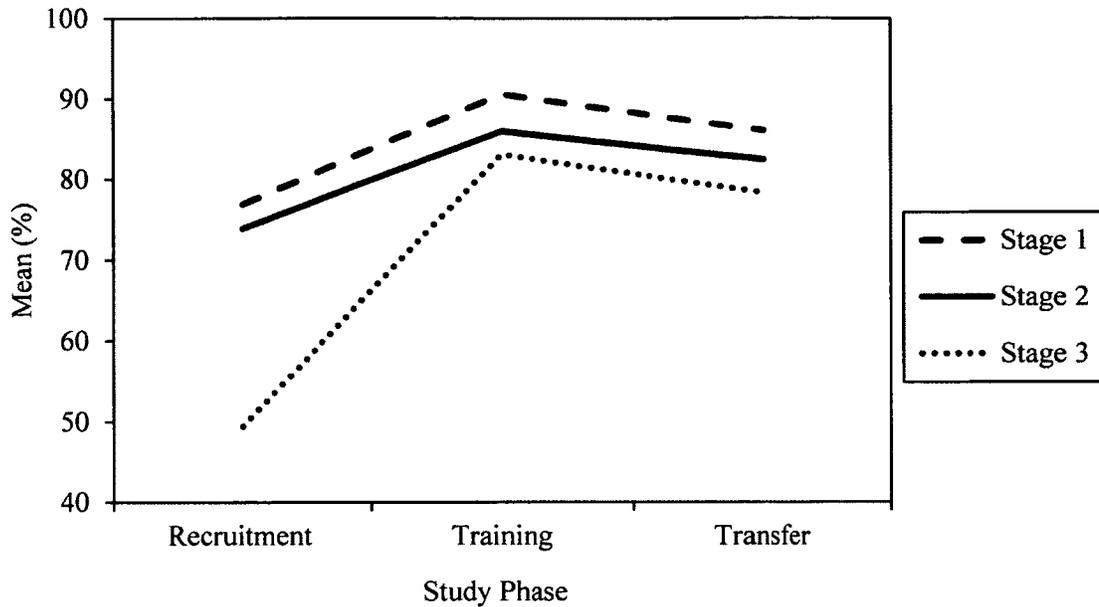


Figure 1. Plot illustrating the knowledge outcomes for the test subsections across study phases.

Skills. Recall that for the skills outcomes there were measures for both soft skills and hard skills gathered through the role play scenarios used to assess each type of skill. The results are presented here.

Soft skills. The descriptive statistics for the soft skills assessment at each of the study phases (recruitment, training and transfer) are presented in Table 18. The results are presented for the test as a whole (out of a maximum of 25), as well as for each subsection, including the role playing component (out of a maximum of 15) and the written component (out of a maximum of 10). All of the results are presented as the percentage correct for ease of interpretation. Generally speaking the results were as expected. In general, the soft skills scores were moderately high, with average scores for the composite measure ranging from 73%-80%.

In terms of soft skill level gains, there were small but steady increases in overall skill levels demonstrated by the sample as a whole (with the exception of the written component), but mainly this occurred between the training and transfer phases. This is illustrated by the increases in the average scores from the recruitment training phases (73%) to the transfer phase (80%). In order to determine whether or not these increases were significant and to determine if there were any regional differences in soft skill outcomes at each phase, a 2 x 3 mixed ANOVA was carried out. The between subjects factor was Region (Pacific and Ontario) and the within subjects factor was Study Phase (recruitment, training and transfer). The dependent variable was the soft skill test score.

Table 18

Descriptive Statistics for the Soft Skills Measures

Study Phase	Soft Skills Outcome	Overall			Pacific			Ontario		
		<i>n</i>	<i>M %</i> (<i>SD</i>)	<i>Range</i>	<i>n</i>	<i>M %</i> (<i>SD</i>)	<i>Range</i>	<i>n</i>	<i>M %</i> (<i>SD</i>)	<i>Range</i>
Recruitment Phase	Skills & Abilities Assessment Overall	102	72.55 (9.39)	60-92	37	72.97 (10.13)	60-92	51	71.92 (9.28)	60-92
	Role Play	102	73.73 (12.01)	60-100	37	73.51 (12.62)	60-100	51	74.51 (12.13)	60-100
	Written Component	102	70.78 (11.14)	60-100	37	72.16 (12.50)	60-100	51	68.04 (9.17)	60-90
Training Phase	Skills & Abilities Assessment Overall	68	73.27 (15.93)	32-100	34	65.92 (15.83)	32-92	27	78.85 (12.80)	52-100
	Role Play	68	76.20 (19.16)	33.33-100	34	67.91 (18.93)	33.33-100	27	83.18 (17.11)	40-100
	Written Component	68	68.90 (17.83)	20-100	34	62.97 (16.48)	20-100	27	72.39 (16.85)	10-100
Transfer Phase	Skills & Abilities Assessment Overall	53	79.68 (12.55)	52-100	26	75.84 (12.00)	52-96	24	76.00 (13.08)	52-100
	Role Play	53	80.76 (15.29)	40-100	26	81.33 (15.78)	40-100	24	78.84 (15.18)	46.67-100
	Written Component	53	70.55 (15.38)	40-100	26	67.60 (14.77)	40-90	24	71.74 (15.51)	40-100

Note. Data for the Quebec region is included in the columns reflecting the sample overall, data for the region itself is provided in Appendix Z.

The main effects of Region ($F(1, 48) = 1.93, p = .170$, partial $\eta^2 = .04$, Obs. Power = .28) and Study Phase ($F(2, 96) = 1.89, p = .157$, partial $\eta^2 = .04$, Obs. Power = .38) were not significant; however, as illustrated in Figure 2, the results of this analysis show that the Region x Study Phase interaction was significant, $F(2, 96) = 5.80, p = .004$, partial $\eta^2 = .11$, Obs. Power = .86. In other words, the pattern of soft skills scores across the three study phases differed depending on the region where the tests were being conducted.

To examine this interaction further, simple effects analyses were carried out, whereby differences in soft skills scores between the regions were examined for each phase separately. These analyses indicated that the interaction between region and study phase can largely be explained by the significant difference in soft skills scores exhibited by trainees in the Pacific and Ontario region during the training phase, $F(1, 59) = 11.85, p = .001$. In this case, the scores exhibited by trainees in the Pacific region were significantly lower than the scores of trainees in the Ontario region ($p < .001$) (see Table 18 for the regional means and standard deviations). There were no significant differences in soft skills scores between regions at the recruitment ($F(1, 86) = .25, p = .615$) or transfer phases ($F(1, 48) = .00, p = .964$).

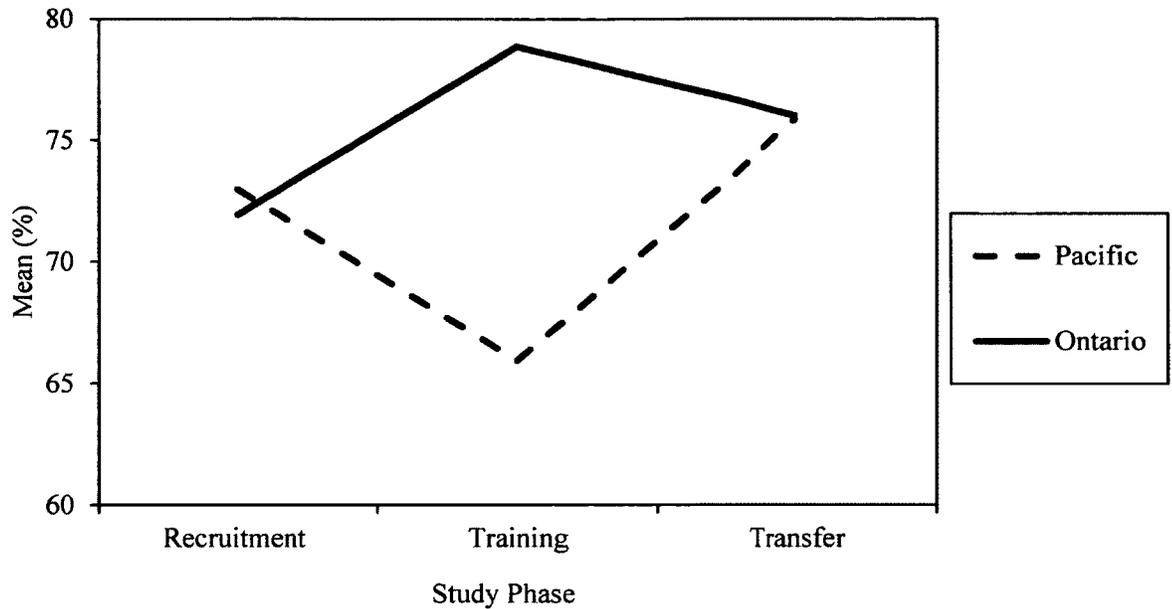


Figure 2. Plot illustrating the soft skills outcomes across study phases per region.

Given that the subsections of the soft skills assessment (role play and written) are not independent from the overall score, they had to be examined separately. Therefore, in order to determine whether or not the same pattern of skill acquisition was found for each of the subsections of the soft skills assessment across study phases and regions, a 2 x 3 x 2 mixed ANOVA was conducted. The between subjects factor was Region (Pacific and Ontario) and the within subjects factors were the Study Phase (recruitment, training and transfer) and Test Subsection (role play and written). The dependent variables were test scores (percentage correct).

The main effects of Region ($F(1, 48) = .20, p = .654, \text{partial } \eta^2 = .00, \text{Obs. Power} = .07$) and Study Phase ($F(1.78, 85.57) = 1.77, p = .180, \text{partial } \eta^2 = .04, \text{Obs. Power} = .34^{23}$) were not significant, whereas a significant main effect of the Test Subsection was found, $F(1, 48) = 22.71, p = .000, \text{partial } \eta^2 = .32, \text{Obs. Power} = 1.00$. None of the two-

²³ Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(2) = 6.10, p = .047$, therefore degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity ($\epsilon = .89$).

way interactions were significant (Region x Study Phase - $F(2, 96) = 1.59, p = .209$, partial $\eta^2 = .03$, Obs. Power = .33; Region x Test Subsection - $F(1, 48) = 1.97, p = .167$, partial $\eta^2 = .04$, Obs. Power = .28; Study Phase x Test Subsection - $F(2, 96) = 1.70, p = .189$, partial $\eta^2 = .03$, Obs. Power = .35).

Only one significant interaction effect was found. This was for the three-way interaction between Region, Study Phase, and Test Subsection, $F(2, 96) = 6.10, p = .003$, partial $\eta^2 = .11$, Obs. Power = .88. In contrast to expectations, this indicates that the mean scores obtained on the two subsections of the test differed as a function of Study Phase and Region. To examine this interaction further, simple effects analysis was carried out and involved running two separate two-way ANOVA's for each region (see Figures 3a and 3b). Thus, a 3 x 2 within subjects ANOVA was run for each region separately. The within subjects factors were Study Phase (recruitment, training and transfer) and Test Subsection (role play and written). The dependent variable was soft skill test score.

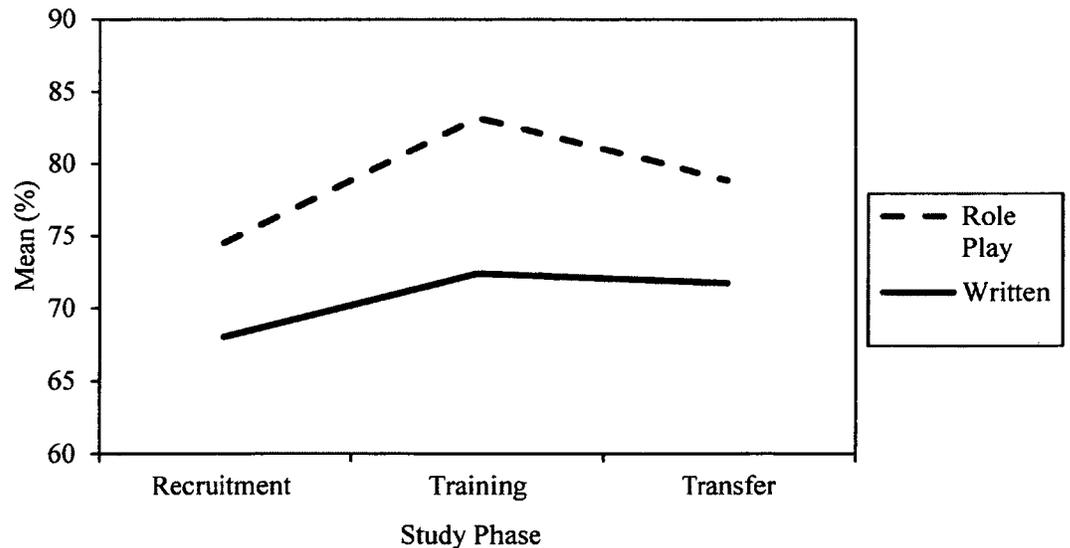


Figure 3a. Plot illustrating the soft skills outcomes for the subsections across study phases for the Ontario region.

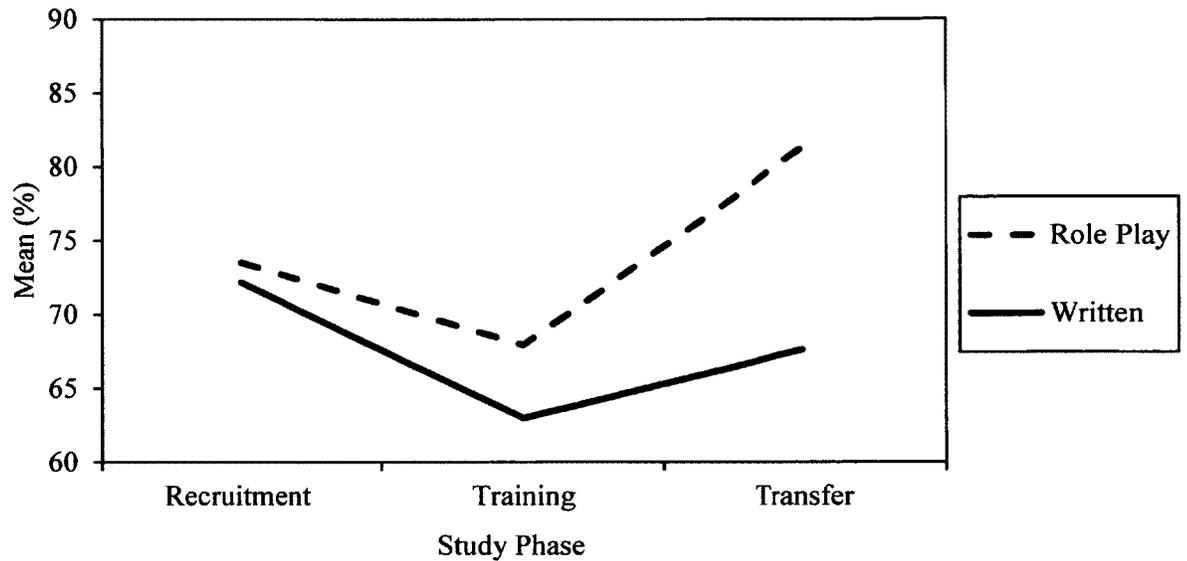


Figure 3b. Plot illustrating the soft skills outcomes for the subsections across study phases for the Pacific Region.

In the Ontario region, both the main effects of Study Phase and Test Subsection were significant ($F(2, 46) = 3.45, p = .040, \text{partial } \eta^2 = .13, \text{Obs. Power} = .62$; $F(1, 23) = 283.57, p = .000, \text{partial } \eta^2 = .93, \text{Obs. Power} = 1.00$, respectively). The two-way interaction of Study Phase x Test Subsections was not significant, $F(2, 46) = .78, p = .462, \text{partial } \eta^2 = .03, \text{Obs. Power} = .18$. In the absence of an interaction effect, the significant main effects were interpreted. Further examination of the pairwise comparisons for the main effect of Study Phase indicate that significantly lower soft skills scores are observed at the recruitment phase compared to the training phase, regardless of subsection (see Table 18 for the means and standard deviations for the Ontario region). However, no differences are observed between the training and transfer phases nor the recruitment and transfer phases. Further examination of the pairwise

comparisons for the main effect of Test Subsection indicate that significantly higher scores are observed for the role play component compared to the written component, regardless of study phase.

As made clear in Figure 3b, this pattern of results is not what was found in the Pacific region. In this case, both the main effects of Study Phase ($F(2, 50) = 4.25, p = .020$, partial $\eta^2 = .15$, Obs. Power = .72) and Test Subsection ($F(1, 25) = 247.33, p = .000$, partial $\eta^2 = .91$, Obs. Power = 1.00) were significant along with the Study Phase x Test Subsection interaction ($F(2, 50) = 7.60, p = .048$, partial $\eta^2 = .14$, Obs. Power = .59). In other words, the pattern of soft skills scores across the three study phases differ depending on test subsection – role play or written.

To examine this interaction further, simple effects analyses were carried out, whereby differences in soft skills scores across the two subsections were examined for each phase separately with one-way ANOVA's. Significant differences in soft skill test scores between the role play and written subsections were found for each of the study phases (recruitment, $F(1, 36) = 138.49, p = .000$, partial $\eta^2 = .79$, Obs. Power = 1.00; training, $F(1, 33) = 87.18, p = .000$, partial $\eta^2 = .73$, Obs. Power = 1.00; transfer, $F(1, 25) = 117.18, p = .000$, partial $\eta^2 = .82$, Obs. Power = 1.00). An examination of the means (see Table 18) shows that significantly lower scores were obtained on the written component compared to the role play component for all study phases in the Pacific region.

Hard skills. The descriptive statistics for the hard skills assessment for the training and transfer phases are presented in Table 19 (recall that hard skills were only assessed at the latter two phases). The results in this table are presented as the percentage

correct on the checklists for all eight skills combined. As expected, in general, the percentage correct was quite high across both study phases with scores during the training phase being consistently around 90% and scores during the transfer phase being consistently about 80%. The descriptive statistics for each of the eight skills are provided in Table 20. The correlation matrices for the individual hard skills (per study phase) are provided in Appendix AA.

To examine hard skill acquisition and retention, as well as any regional differences, a 2 x 2 mixed ANOVA was carried out with Region (Pacific and Ontario) as the between subjects factor and Study Phase (training and transfer) as the within subjects factor. The dependent variable was the hard skills score. The main effect of Region was not significant ($F(1, 48) = 2.98, p = .091, \text{partial } \eta^2 = .06, \text{Obs. Power} = .40$); however, a significant main effect of Study Phase was found, $F(1, 48) = 33.51, p = .000, \text{partial } \eta^2 = .41, \text{Obs. Power} = 1.00$. The Region x Study Phase interaction was not significant ($F(1, 48) = .41, p = .524, \text{partial } \eta^2 = .01, \text{Obs. Power} = .10$); therefore, the significant main effects found for Study Phase were interpreted. An examination of the means revealed that hard skills scores were significantly higher during the training phase ($M = 90.31, SD = 7.41$) compared to the transfer phase ($M = 82.80, SD = 8.30; p = .000$), in other words there was a significant loss of hard skill ability between training and four to six months on the job regardless of Region.

Table 19

Descriptive Statistics for the Hard Skills Composite Checklist Score

Study Phase	Overall			Pacific			Ontario		
	<i>n</i>	<i>M % (SD)</i>	<i>Range</i>	<i>n</i>	<i>M % (SD)</i>	<i>Range</i>	<i>n</i>	<i>M % (SD)</i>	<i>Range</i>
Training	68	90.83 (7.20)	69.88-99.26	34	88.29 (7.51)	69.88-97.30	27	92.02 (6.00)	78.66-98.61
Transfer	53	83.05 (8.23)	64.74-97.83	26	81.73 (8.84)	65.74-97.83	24	83.96 (7.68)	64.74-97.22

Note. Data for the Quebec region is included in the columns reflecting the sample overall, data for the region itself is provided in Appendix Z.

Table 20

Descriptive Statistics for the Eight Individual Hard Skills Checklist Scores

Skill	Overall			Pacific			Ontario		
	<i>n</i>	<i>M % (SD)</i>	<i>Range</i>	<i>n</i>	<i>M % (SD)</i>	<i>Range</i>	<i>n</i>	<i>M % (SD)</i>	<i>Range</i>
Training Phase									
Frisk Search	68	81.80 (15.05)	33.33- 100.00	34	80.60 (14.69)	55.56- 100.00	27	78.60 (14.43)	33.33- 100.00
OC Spray	68	96.66 (6.99)	74.00- 100.00	34	96.76 (6.73)	75.00- 100.00	27	95.66 (8.02)	74.00- 100.00
Radio	68	97.03 (5.54)	83.62- 100.00	34	95.19 (6.67)	83.62- 100.00	27	98.58 (3.72)	84.62- 100.00
Handcuffing	68	89.04 (10.11)	62.50- 100.00	34	84.19 (9.64)	62.50- 100.00	27	92.55 (8.57)	68.75- 100.00
Low Profile Escort	68	89.71 (28.07)	0.00- 100.00	34	83.82 (34.20)	0.00- 100.00	27	94.44 (21.18)	0.00- 100.00
High Profile Escort	68	90.51 (14.54)	42.86- 100.00	34	86.06 (17.41)	42.86- 100.00	27	94.18 (9.91)	71.43- 100.00
Application/Removal of Leg Irons	68	90.38 (9.29)	58.82- 100.00	34	88.37 (10.34)	64.71- 100.00	27	92.16 (10.06)	58.82- 100.00
Application/Removal of Body Belt	68	92.54 (8.13)	74.00- 100.00	34	93.16 (7.67)	74.00- 100.00	27	90.28 (8.90)	75.00- 100.00
Transfer Phase									
Frisk Search	53	65.14 (21.35)	22.22- 100.00	26	63.56 (22.12)	22.22- 100.00	24	65.28 (21.56)	33.33- 100.00

Skill	Overall			Pacific			Ontario		
	<i>n</i>	<i>M</i> % (<i>SD</i>)	<i>Range</i>	<i>n</i>	<i>M</i> % (<i>SD</i>)	<i>Range</i>	<i>n</i>	<i>M</i> % (<i>SD</i>)	<i>Range</i>
OC Spray	53	98.73 (3.81)	86.50- 100.00	26	98.38 (4.29)	86.50- 100.00	24	98.95 (3.53)	87.50- 100.00
Radio	53	89.40 (10.48)	69.23- 100.00	26	85.49 (11.17)	69.23- 100.00	24	92.31 (8.49)	76.92- 100.00
Handcuffing	53	83.37 (10.54)	56.25- 100.00	26	81.00 (9.92)	56.25- 93.75	24	85.42 (11.31)	62.50- 100.00
Low Profile Escort	53	77.89 (37.23)	0.00- 100.00	26	78.00 (37.63)	0.00- 100.00	24	77.08 (38.95)	0.00- 100.00
High Profile Escort	53	71.72 (27.52)	0.00- 100.00	26	72.57 (24.88)	0.00- 100.00	24	69.05 (31.43)	0.00- 100.00
Application/Removal of Leg Irons	53	90.83 (7.86)	70.59- 100.00	26	90.59 (8.15)	70.59- 100.00	24	91.91 (7.32)	76.47- 100.00
Application/Removal of Body Belt	53	86.96 (11.24)	56.25- 100.00	26	85.25 (12.23)	56.25- 100.00	24	89.79 (9.23)	62.50- 100.00

Note. Data for the Quebec region is included in the columns reflecting the sample overall, data for the region itself is provided in Appendix Z.

In order to determine whether the same pattern of skill acquisition was found for each of the eight individual hard skills across study phases and regions, a 2 x 2 x 8 mixed ANOVA was conducted. The between subjects factor was Region (Pacific and Ontario) and the within subjects factors were Study Phase (training and transfer) and Skill (Frisk Search, OC, Radio, Handcuffing, Low Profile Escort, High Profile Escort, Application/Removal of Leg Irons and Application/Removal of Body Belt).

The main effect of Region was not significant ($F(1, 48) = 2.46, p = .123$, partial $\eta^2 = .05$, Obs. Power = .34). A significant main effect of Study Phase was found for the hard skills checklist scores ($F(1, 48) = 35.06, p = .000$, partial $\eta^2 = .42$, Obs. Power = 1.00) as well as Skill, $F(2.36, 113.22) = 19.07, p = .000$, partial $\eta^2 = .28$, Obs. Power = 1.00²⁴). The majority of the two-way interactions, including Region x Study Phase ($F(1, 48) = .46, p = .502$, partial $\eta^2 = .01$, Obs. Power = .10) and Region x Skill ($F(7, 336) = .67, p = .695$, partial $\eta^2 = .01$, Obs. Power = .29), as well as the three-way interaction of Region x Study Phase x Skill ($F(7, 336) = 1.52, p = .158$, partial $\eta^2 = .03$, Obs. Power = .64) were not significant. Only one significant interaction effect was found. This was for the two-way interaction of Study Phase x Skill, $F(2.87, 137.95) = 6.56, p = .000$, partial $\eta^2 = .12$, Obs. Power = .96.²⁵ As illustrated in Figure 4, this indicates that, in contrast to expectations, the scores obtained for the individual hard skills differed based on the phase of the study.

²⁴ Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(27) = 254.85, p = .000$, therefore degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity ($\epsilon = .34$).

²⁵ Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(27) = 226.84, p = .000$, therefore degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity ($\epsilon = .41$).

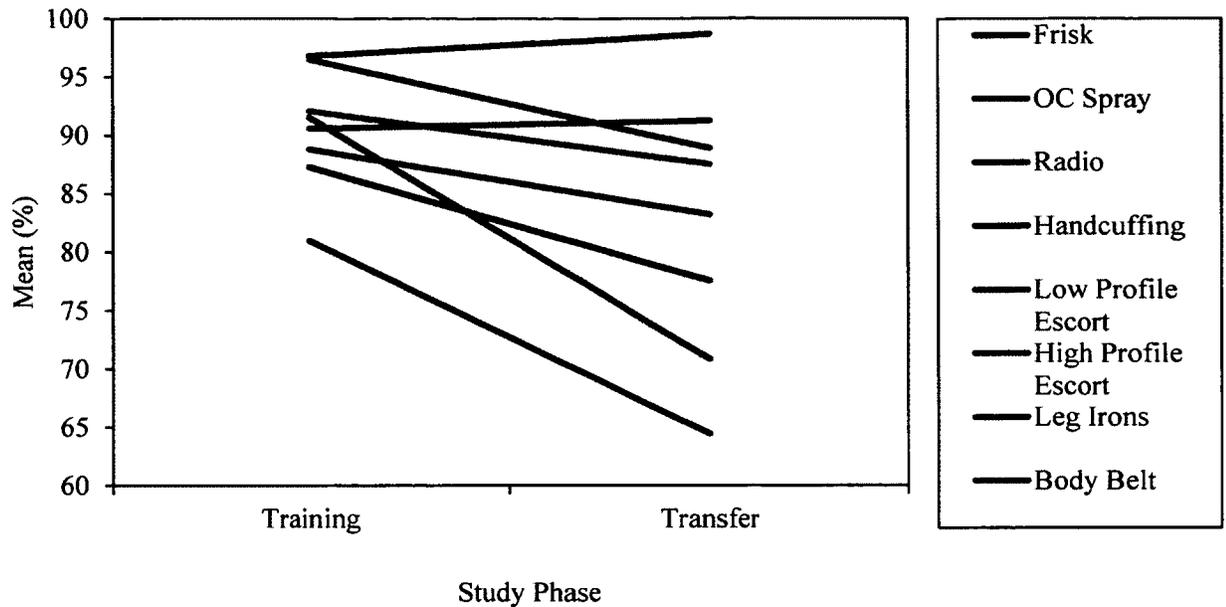


Figure 4. Plot of the significant interaction effect of Study Phase x Skill.

To examine this interaction further, simple effects analyses were carried out, whereby differences in scores for each skill were examined separately for each phase via one-way ANOVAs. These analyses indicated that the interaction between study phase and test subsection could be explained by the significant differences in skills scores at both phases (see Table 21 below for the means and standard deviations for the Pacific and Ontario regions combined). Specifically, the results at the training phase ($F(2.56, 153.66) = 9.46, p < .001, \text{partial } \eta^2 = .14, \text{Obs. Power} = 1.00$)²⁶ indicated that the Frisk scores were significantly lower than all other skills scores (p 's $\leq .001$) except the Low Profile Escort where no differences were observed ($p > .05$).²⁷ In contrast, OC scores were significantly higher than scores for Handcuffing and Leg Irons (p 's $< .001$) but not

²⁶ Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(27) = 233.79, p < .001$, therefore degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity ($\epsilon = .37$).

²⁷ A Bonferroni correction was applied here for the multiple comparisons, $p = .05/28 = .002$.

significantly different from the remaining skills scores (High Profile Escort, $p = .002$; Body Belt, $p = .003$; Low Profile Escort and Radio, p 's $> .05$). Somewhat similar to OC, the scores for Radio were significantly higher than Handcuffing, Leg Irons, Body Belt (p 's $< .001$) and the High Profile Escort ($p = .001$), but not significantly different from the Low Profile Escort ($p = .034$). Scores for Handcuffing, Low Profile Escort, High Profile Escort, and Leg Irons were not significantly different from the remaining skills (i.e., Low Profile Escort, High Profile Escort, Leg Irons, Body Belt with p 's $> .05$).

The results at the transfer phase, ($F(2.66, 130.54) = 17.52, p < .001, \text{partial } \eta^2 = .26, \text{Obs. Power} = 1.00$)²⁸ were similar to that found for the training phase, although a few differences were observed here. Frisk scores were still significantly lower than most other skills (p 's $< .001$) except for the Low and High Profile Escorts where no differences in scores were observed ($p = .045$ and $.117$, respectively). In contrast, scores for the OC were significantly higher than all of the remaining skills scores (p 's $< .001$). Scores for the Radio were still significantly higher than those for the High Profile Escort ($p < .001$), but not significantly different from scores on Handcuffing ($p = .003$), Low Profile Escort ($p = .041$), Leg Irons, and Body Belt (p 's $> .05$). Handcuffing scores were significantly lower than Leg Irons ($p < .001$), but not significantly different from the remaining skills scores (i.e., Low Profile Escort, $p > .05$; High Profile Escort, $p = .003$; and Body Belt, $p > .05$). The Low Profile Escort scores were not significantly different from the remaining skills scores (i.e., High Profile Escort, $p > .05$; Leg Irons, $p = .018$; and Body Belt, $p > .05$). In contrast, the High Profile Escort scores were significantly lower than the Leg

²⁸ Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(27) = 264.16, p < .001$, therefore degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity ($\epsilon = .38$).

Irons ($p < .001$) and Body Belt ($p = .001$). Lastly, as in the training phase, the Leg Irons scores were not significantly different than scores for the Body Belt ($p = .039$).

Table 21

Descriptive Statistics for the Eight Individual Hard Skills Checklist Scores for the Pacific and Ontario Regions Combined (n = 50)

Skills	<i>M (SD)</i>
Training Phase	
Frisk Search	81.03 (14.34)
OC Spray	96.83 (6.80)
Radio	96.42 (5.94)
Handcuffing	88.60 (9.78)
Low Profile Escort	87.00 (31.63)
High Profile Escort	91.38 (14.16)
Application/Removal of Leg Irons	90.47 (10.15)
Application/Removal of Body Belt	92.11 (8.31)
Transfer Phase	
Frisk Search	64.38 (21.65)
OC Spray	98.66 (3.91)
Radio	88.76 (10.46)
Handcuffing	83.12 (10.73)
Low Profile Escort	77.56 (37.88)
High Profile Escort	70.88 (27.98)
Application/Removal of Leg Irons	91.22 (7.71)
Application/Removal of Body Belt	87.43 (11.03)

Attitudes. The descriptive statistics for the attitude measures (i.e., the ATCW and ATI) are presented in Table 22. The results are presented as the percentage of items selected, with higher scores indicating more positive attitudes.

ATCW. A higher score on the ATCW indicates a more positive attitude towards correctional work. In the current study, an internal consistency score of .46 was found for the 12-item ATCW scale pre-training, which indicates a low level of reliability.

However, when the ATCW scores were collected immediately following training and four to six months post-training, the scores increased to .58 and .62, respectively. Despite the higher internal consistency scores found during the training and transfer phase, all of these scores are slightly lower than that found in previous research (Robinson et al., 1993, 1996; Simourd, 1997). Therefore, caution should be used when interpreting these results given the relatively low alpha values (Nunnally, 1978).

Table 22

Descriptive Statistics for the Attitudes Measures

Study Phase	Attitude Measure	Overall			Pacific			Ontario		
		<i>n</i>	<i>M %</i> (<i>SD</i>)	<i>Range</i>	<i>n</i>	<i>M %</i> (<i>SD</i>)	<i>Range</i>	<i>n</i>	<i>M %</i> (<i>SD</i>)	<i>Range</i>
Recruitment	ATCW	102	89.11 (10.92)	58.33-100	37	90.09 (10.91)	58.33-100	51	89.16 (11.21)	58.33-100
Training	ATCW	68	92.03 (8.69)	58.33-100	34	92.40 (9.70)	58.33-100	27	91.36 (7.83)	75-100
Transfer	ATCW	53	84.16 (14.27)	41.67-100	26	85.67 (13.02)	41.67-100	24	81.94 (16.05)	50-100
Recruitment	ATI	102	69.75 (6.94)	51.11-87.22	37	69.65 (5.85)	58.89-82.22	51	70.25 (7.18)	51.11-87.22
Training	ATI	68	67.39 (6.00)	53.33-82.22	34	67.83 (6.49)	86.67-82.22	27	67.16 (5.96)	53.33-77.78
Transfer	ATI	53	60.43 (8.11)	43.89-75.56	26	61.78 (8.04)	43.89-75.56	24	59.28 (8.60)	44.44-74.44

Note. Data for the Quebec region is included in the columns reflecting the sample overall, data for the region itself is provided in Appendix Z.

Generally speaking the results were as expected. The mean ATCW scores were quite high for the sample as a whole with scores ranging from 84-92% found across all phases of the study (see Table 22). Thus, the current sample displayed highly positive attitudes towards correctional work before and after training, as well as four to six months on the job. In order to determine whether or not there were regional differences in ATCW outcomes at each of the study phases, as well as to determine the significance of attitude change across study phases, a 2 x 3 mixed ANOVA was conducted. The between subjects factor was Region (Pacific and Ontario) and the within subjects factor was Study Phase (recruitment, training and transfer). The dependent variable was the scores on the ATCW.

The results show that the main effect of Region was not significant ($F(1, 48) = .55, p = .462, \text{partial } \eta^2 = .01, \text{Obs. Power} = .11$) but there was a significant main effect of Study Phase such that attitude ratings differed significantly across phase, $F(2, 96) = 8.92, p = .000, \text{partial } \eta^2 = .16, \text{Obs. Power} = 1.00$. The Region x Study Phase interaction was not significant ($F(2, 96) = .42, p = .658, \text{partial } \eta^2 = .01, \text{Obs. Power} = .12$), which indicates that attitude ratings were generally consistent across regions.

The main effect of Study Phase was interpreted given the non-significant interaction. Pairwise comparisons revealed that attitude ratings did not differ between the recruitment and training phase ($p = .752$), but ratings at recruitment ($M = 89.55, SD = 11.03$) and training ($M = 91.94, SD = 8.87$) were significantly more positive than ratings obtained during the transfer phase ($M = 83.88, SD = 14.52; p$'s = .024 and .001, respectively) (see Figure 5).

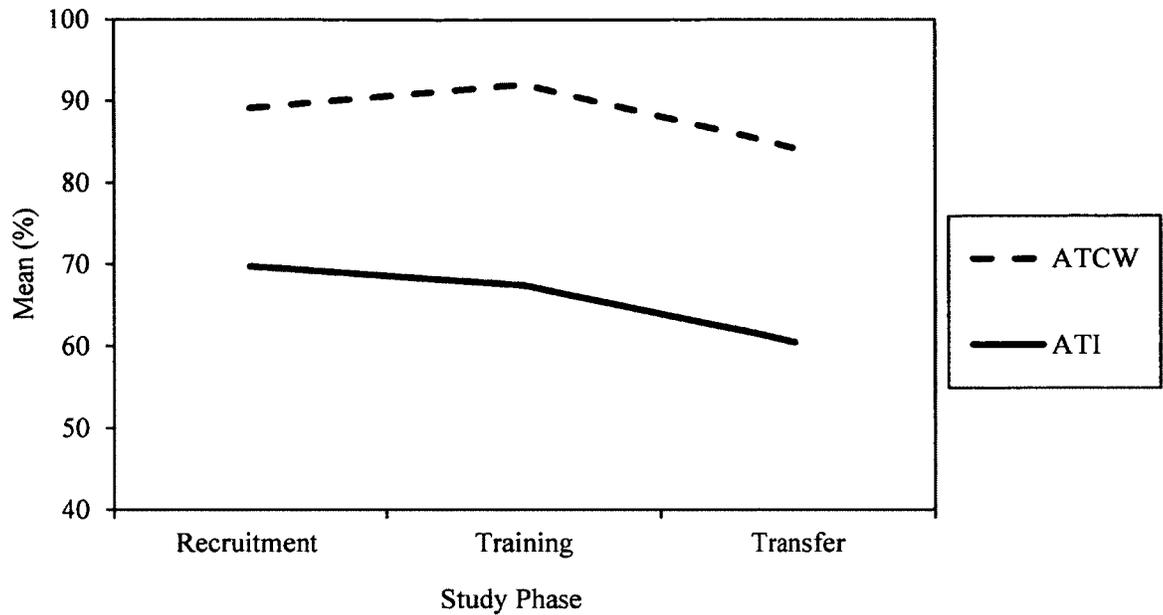


Figure 5. Plot of the attitudes outcomes across study phases.

ATI. Similar to the ATCW measure, a higher score on the ATI indicates a more positive attitude towards inmates. In the current study, an internal consistency score of .87 was found for the 36-item ATI scale during the pre-training phase. Equally high scores (of .85 and .91) were obtained immediately following the training and four to six months post-training, respectively. These values indicate a high level of reliability and are consistent with what has been found in other samples (Melvin et al., 1985; Ortet-Fabregat et al., 1993).

Overall, the mean ATI scores reached moderate levels pre- and post-CTP indicating fairly positive attitudes towards inmates (70% and 67%, respectively; see Table 22). Generally, this is what was expected. Unfortunately, the scores on the ATI dropped to approximately 60% during the transfer phase. In order to determine whether or not there were regional differences in ATI, as well as to determine the degree of attitude change that took place across study phases, a 2 x 3 mixed ANOVA was

conducted. The between subjects factor was Region (Pacific and Ontario) and the within subjects factor was Study Phase (recruitment, training and transfer). The dependent variable was the scores on the ATI.

The results show that the main effect of Region was not significant ($F(1, 48) = 1.10, p = .300, \text{partial } \eta^2 = .02, \text{Obs. Power} = .18$) but that there was a significant main effect of Study Phase, such that attitude ratings differed significantly across phase, $F(2, 96) = 45.90, p = .000, \text{partial } \eta^2 = .49, \text{Obs. Power} = 1.00$. The Region x Study Phase interaction was not significant ($F(2, 96) = 22.09, p = .732, \text{partial } \eta^2 = .00, \text{Obs. Power} = .10$), indicating that attitude ratings were generally consistent across regions.

The main effect of Study Phase was interpreted given the non-significant interaction. Pairwise comparisons revealed that, similar to the ATCW survey, attitude ratings did not differ between the recruitment and training phase ($p = .121$), but ratings at recruitment ($M = 70.00, SD = 6.63$) and training ($M = 67.53, SD = 6.22$) were significantly more positive than ratings obtained during the transfer phase ($M = 60.58, SD = 8.32; p's = .00$) (see Figure 5).

Level 3 Outcomes

Level 3 outcomes included the trainee and supervisor behavioural checklists collected at the transfer phase (frequency and quality scores). The internal reliability scores were relatively high for the two components of the trainee behavioural checklist for the sample as a whole (frequency = .83, quality = .79), as well as for each region separately (Pacific - frequency = .81, quality = .81; Ontario - frequency = .83, quality = .74) (George & Mallery, 2003). As for the supervisor behavioural checklist, the internal reliability scores for the two components were also relatively high for the sample as a

whole (frequency = .85, quality = .79), as well as for each region separately (Pacific - frequency = .90, quality = .84; Ontario - frequency = .59, quality = .74) (George & Mallery, 2003).

Bivariate correlations between trainee and supervisor ratings for each of the individual frequency and quality items on the behavioural checklists were calculated using Spearman's r_s and Pearson's r , respectively (frequency scores are reported as a median while quality items are reported as a mean, see below). In contrast to the hypothesis, the majority of the correlations did not reach significance, meaning that there is little to no relationship between supervisor and trainee self-reports of the frequency and quality of skill use on the job (despite similarities in ratings shown below). The only significant correlation at the corrected alpha level ($p = .05/21 = .002$) was found between trainees and supervisors on the reported quality of use of chemical agents ($r = .71, p < .001$).²⁹

Trainee Behavioural Checklists. Recall that the two components (frequency and quality) of the trainee behavioural checklist were rated on different scales. For the frequency component, trainees were asked to rate how frequently they engaged in various behaviours/skills on average per month using the following five-point scale: 1 (never/no opportunity to use), 2 (1 to 5 times), 3 (6 to 10 times), 4 (11 to 15 times), and 5 (more than 15 times). Higher scores on the frequency component indicate greater use of skills. The descriptive statistics for this scale are provided in Table 23. Due to the way that the frequency of skill use was measured, the median, mode and ranges are provided instead of means and standard deviations.

²⁹ Due to the large number of correlations run here for each of the individual frequency and quality items, for the sake of brevity and to conserve space, only the significant correlations are reported.

Table 23

Descriptive Statistics for the Trainee Behavioural Checklist - Frequency Items

Behaviours/Skills	Overall (n = 53)				Pacific (n = 26)				Ontario (n = 24)			
	NOU (n) ^a	Median	Mode	Range	NOU (n)	Median	Mode	Range	NOU (n)	Median	Mode	Range
Assessing Problematic Situations	0	4	5	2-5	0	5	5	2-5	0	3.50	5	2-5
Identifying Possible Solutions	0	4	5	2-5	0	4	5	2-5	0	3.50	3	2-5
Taking Necessary Measures	1	4	5	1-5	1	4	5	1-5	0	4	5	2-5
Motivating and Influencing Individuals	3	3	2	1-5	1	3.50	2	1-5	1	3	3	1-5
Effective Oral Communication	1	5	5	1-5	1	5	5	1-5	0	5	5	2-5
Effective Communication in Writing	0	3	2	1-5	0	3	3	2-5	0	2.50	2	1-5
Summarizing Information	0	3	2	1-5	0	3.50	2	2-5	0	3	2	1-5
Report Writing:												
OSOR	2	2	2	1-5	1	2.50	2	1-5	1	2	2	1-4
Post Search	8	2	2	1-5	4	2	2	1-5	4	2	2	1-3
Use of Force	29	1	1	1-3	16	1	1	1-2	13	1	1	1-3

Behaviours/Skills	Overall (<i>n</i> = 53)				Pacific (<i>n</i> = 26)				Ontario (<i>n</i> = 24)			
	NOU (<i>n</i>) ^a	Median	Mode	Range	NOU (<i>n</i>)	Median	Mode	Range	NOU (<i>n</i>)	Median	Mode	Range
Offence	14	2	2	1-4	7	2	2	1-4	7	2	2	1-3
Using Communication Equipment (e.g., two-way radio, loudhailer, etc.)	0	5	5	4-5	0	5	5	4-5	0	5	5	4-5
Using Restraint Equipment	1	3	2	1-5	0	3	2	1-5	1	2.50	2	1-5
Engaging in a Search (e.g., area, frisk, strip, etc.)	0	5	5	2-5	0	5	5	3-5	0	5	5	2-5
Escorts (e.g., internal and outside)	1	2	2	1-5	1	2	2	1-5	0	2	2	2-4
Using Chemical and Inflammatory Agents	31	1	1	1-4	13	1	1	1-4	17	1	1	1-2
Utilizing Fire Safety Measures	27	1	1	1-5	11	1	1	1-5	15	1	1	1-5
Using Self Defence Skills	29	1	1	1-5	13	1	1	1-5	15	1	1	1-2
Using Arrest and Control Skills	22	1	1	1-5	10	1	1	1-5	12	1	1	1-5

Behaviours/Skills	Overall (<i>n</i> = 53)				Pacific (<i>n</i> = 26)				Ontario (<i>n</i> = 24)			
	NOU (<i>n</i>) ^a	<i>Median</i>	<i>Mode</i>	<i>Range</i>	NOU (<i>n</i>)	<i>Median</i>	<i>Mode</i>	<i>Range</i>	NOU (<i>n</i>)	<i>Median</i>	<i>Mode</i>	<i>Range</i>
Using Baton Skills	31	1	1	1-2	13	1	1	1-2	18	1	1	1-2

Note. Data for the Quebec region is included in the columns reflecting the sample overall, data for the region itself is provided in Appendix Z.

^aTrainees were given the option to indicate that there was no opportunity to use this skill (abbreviated as NOU to save space). When this response was selected it was obviously not included in the calculation of the descriptive statistics; however, the number of trainees who provided this response is provided.

Based on an examination of Table 23, in contrast to the hypothesis, it can be seen that trainees reported that they utilize the trained soft skills to different degrees depending on the specific skill they are rating. For example, skills that fall under the category of problem solving and communication skills (e.g., assessing problematic situations) are used quite frequently (anywhere from 6-10 times per month, 11-15 times per month or daily). On the other hand, report writing skills are utilized less frequently, ranging from never or no opportunity to use (e.g., writing specific reports, such as the Use of Force Report) to 6-10 times per month (e.g., communicating effectively in writing). Likewise, the frequency of hard skill use varied from those skills that were used daily (e.g., using communication equipment and frisk searches) to those skills that were not used at all (e.g., chemical agents, arrest and control, self defence, etc.).

For the quality component, trainees were asked to rate the quality with which they execute each of the behaviours/skills they rated on the frequency scale. Quality was conceptualized as the consistency between the actual execution of the skills and the way in which the trainee was trained to execute the skills during CTP. Quality was measured on a three-point scale ranging from 1 (below average) to 3 (above average). Higher scores on the quality component indicate a more successful execution of skills according to trainees (i.e., high consistency with how they were taught to execute the skill during CTP).

Overall, the mean checklist scores for quality were moderate ($M = 2.28$, $SD = .35$, range = 1.25-2.94), meaning that trainees felt that they were able to execute the sampled behaviours/skills in a manner that was generally consistent with how they were taught in CTP some of the time (see Table 24 for the individual items). Although higher ratings

were expected, the moderate ratings are not entirely surprising given that there is a certain amount of learning expected to take place on the job and that skill application is likely to change slightly as officers get more comfortable executing the skills in an institutional environment. Similar results were found for each region (Pacific: $M = 2.30$, $SD = .37$, range = 1.25-2.93; Ontario: $M = 2.26$, $SD = .34$, range = 1.53-2.94). An independent samples t -test revealed that, as expected, no significant regional differences existed, $t(48) = -.36$, $p > .05$.

Table 24

Descriptive Statistics for the Trainee Behavioural Checklist - Quality Items

Behaviours/Skills	Overall				Pacific				Ontario			
	<i>n</i>	NOU (<i>n</i>) ^a	Mean (<i>SD</i>)	Range	<i>n</i>	NOU (<i>n</i>)	Mean	Range	<i>n</i>	NOU (<i>n</i>)	Mean	Range
Assessing Problematic Situations	53	0	2.47 (.54)	1-3	26	0	2.42 (.58)	1-3	24	0	2.58 (.50)	2-3
Identifying Possible Solutions	52	1	2.31 (.65)	1-3	25	1	2.32 (.70)	1-3	24	0	2.29 (.62)	1-3
Taking Necessary Measures	52	1	2.27 (.61)	1-3	25	1	2.24 (.74)	1-3	24	0	2.33 (.48)	2-3
Motivating and Influencing Individuals	50	3	2.14 (.73)	1-3	26	0	2.12 (.65)	1-3	21	3	2.19 (.83)	1-3
Communicating Effectively Orally	52	1	2.48 (.61)	1-3	25	1	2.48 (.65)	1-3	24	0	2.50 (.59)	1-3
Communicating Effectively in Writing	53	0	2.36 (.59)	1-3	26	0	2.42 (.58)	1-3	24	0	2.29 (.62)	1-3
Summarizing Information	53	0	2.38 (.56)	1-3	26	0	2.54 (.51)	2-3	24	0	2.17 (.57)	1-3
Report Writing:												
OSOR	52	1	2.46 (.61)	1-3	26	0	2.46 (.58)	1-3	23	1	2.43 (.66)	1-3
Post Search	42	11	2.26 (.88)	1-3	20	6	2.20 (.90)	1-3	19	5	2.37 (.91)	1-3

Behaviours/Skills	Overall				Pacific				Ontario			
	<i>n</i>	NOU (<i>n</i>) ^a	Mean (<i>SD</i>)	Range	<i>n</i>	NOU (<i>n</i>)	Mean	Range	<i>n</i>	NOU (<i>n</i>)	Mean	Range
Use of Force	21	32	2.05 (1.05)	1-3	9	17	1.89 (1.08)	1-3	10	14	2.30 (.96)	1-3
Offence	35	18	2.37 (.92)	1-3	16	10	2.31 (.96)	1-3	16	8	2.44 (.91)	1-3
Using Communication Equipment (e.g., two- way radio, loudhailer, etc.)	53	0	2.40 (.63)	1-3	26	0	2.46 (.65)	1-3	24	0	2.29 (.62)	1-3
Using Restraint Equipment	51	2	2.14 (.74)	1-3	25	1	2.24 (.68)	1-3	23	1	2.04 (.85)	1-3
Engaging in a Search (e.g., area, frisk, strip, etc.)	53	0	2.09 (.66)	1-3	26	0	2.15 (.68)	1-3	24	0	2.00 (.66)	1-3
Escorts (e.g., internal and outside)	51	2	2.24 (.67)	1-3	24	2	2.17 (.74)	1-3	24	0	2.25 (.61)	1-3
Using Chemical and Inflammatory Agents	12	41	2.00 (.89)	1-3	5	21	2.20 (.75)	1-3	6	18	2.00 (.93)	1-3
Utilizing Fire Safety Measures	16	37	2.06 (.95)	1-3	9	17	2.11 (.94)	1-3	6	18	2.17 (.88)	1-3
Using Self Defence Skills	15	38	2.13 (.91)	1-3	7	19	2.29 (.81)	1-3	7	17	2.14 (.93)	1-3
Using Arrest and Control Skills	20	33	2.10 (1.01)	1-3	10	16	2.20 (.93)	1-3	9	15	2.11 (1.04)	1-3

Behaviours/Skills	Overall				Pacific				Ontario			
	<i>n</i>	NOU (<i>n</i>) ^a	<i>Mean</i> (<i>SD</i>)	<i>Range</i>	<i>n</i>	NOU (<i>n</i>)	<i>Mean</i>	<i>Range</i>	<i>n</i>	NOU (<i>n</i>)	<i>Mean</i>	<i>Range</i>
Using Baton Skills	7	46	1.57 (.85)	1-2	4	22	1.75 (.85)	1-2	2	22	1.50 (.72)	1-2

Note. Data for the Quebec region is included in the columns reflecting the sample overall, data for the region itself is provided in Appendix Z.

^aTrainees were given the option to indicate that there was no opportunity to use this skill (abbreviated to NOU to save space). When this response was selected it was obviously not included in the calculation of the mean; however, the number of trainees who provided this response is provided.

Following the frequency and quality ratings, trainees were asked a number of open-ended questions about their ability to apply what they had learned on the job. Specifically, trainees were asked about: (1) if they had ever gone back to review the training materials now that they were on the job, (2) the most frequently used skills, (3) any barriers they may have encountered that may have prevented them from using what they learned, etc. This information was gathered and categorized for formative purposes for CSC. Since no further analyses were planned for the current study, a breakdown of the coding categories as well as the specific comments provided by trainees is provided in Appendix BB. In general, the responses were that about half of trainees indicated that they had gone back to review their training material (either for a general overview or to review a specific content area), whereas the other half reported that such a review was not necessary. In terms of skill use, communication/listening was reported as the behaviour used most commonly on the job as a result of CTP. The second most commonly used behaviour was problem solving. The barrier most often reported as preventing officers from using their skills or knowledge gained in CTP was their coworkers (e.g., existing workers encouraging the newly hired trainee to use skills that weren't compatible with CTP). No opportunity to use skills was the second most common barrier and the third was the CTP training itself (not preparing them adequately for the job or not being realistic).

Supervisor Behavioural Checklists. The descriptive statistics for frequency items of the supervisor behavioural checklist are presented in Table 25. Again, the median, mode and range are presented for the frequency component. Supervisory ratings were only provided for 25 of the 53 trainees who completed the transfer phase ($n = 16$ for

Pacific and $n = 9$ for Ontario). Supervisors reported that these officers utilized the sample of skills differentially depending on which skill they were rating (similar to the ratings provided by trainees). For example, supervisors rated those skills falling under the category of problem solving and communication skills (e.g., assessing problematic situations) as being used quite frequently (anywhere from 6-10 times per month, to 11-15 times per month to daily). Report writing skills were reportedly utilized less frequently, but slightly more often than that reported by trainees, ranging from 1-5 times per month to daily (e.g., writing specific reports).³⁰ Again, in contrast to the hypothesis, the frequency of hard skill use varied, from those skills that were used daily (e.g., using communication equipment and frisk searches) to those skills that were not used at all (e.g., chemical agents, arrest and control, self defence, etc.).

³⁰ Supervisors provided an overall rating for all reports due to a glitch in the document that was emailed by the author. The supervisors were unable to check the boxes related to each specific report separately, instead they could only provide an overall rating for all four reports together. This only occurred for the frequency items. All information was gathered for the quality items.

Table 25

Descriptive Statistics for the Supervisor Behavioural Checklist - Frequency Items

Behaviours/Skills	Overall (n = 25)			Pacific (n = 16)			Ontario (n = 9)		
	<i>Median</i>	<i>Mode</i>	<i>Range</i>	<i>Median</i>	<i>Mode</i>	<i>Range</i>	<i>Median</i>	<i>Mode</i>	<i>Range</i>
Assessing Problematic Situations	3	2	1-5	3	2	1-5	4	5	2-5
Identifying Possible Solutions	3	2	1-5	2.50	2	1-5	4	5	2-5
Taking Necessary Measures	3	2	1-5	2	2	1-5	4	5	2-5
Motivating and Influencing Individuals	2	1	1-5	2	1	1-5	3	4	1-4
Effective Oral Communication	4	5	2-5	4	5	2-5	5	5	2-5
Effective Communication in Writing	3	3	2-5	3	3	2-5	3	3	2-5
Summarizing Information	3	3	2-5	3	2	2-5	3	3	2-4
Report Writing (OSOR, Post Search, Use of Force, Offence, etc.)	3	2	2-5	3	2	2-5	2.33	2	2-4
Using Communication Equipment (e.g., two-way radio, loudhailer, etc.)	5	5	3-5	5	5	3-5	5	5	3-5
Using Restraint Equipment	2	2	1-5	2	2	1-5	3	3	1-5
Engaging in a Search (e.g., area, frisk, strip, etc.)	4	5	1-5	3.50	2	1-5	5	5	2-5

Behaviours/Skills	Overall (<i>n</i> = 25)			Pacific (<i>n</i> = 16)			Ontario (<i>n</i> = 9)		
	<i>Median</i>	<i>Mode</i>	<i>Range</i>	<i>Median</i>	<i>Mode</i>	<i>Range</i>	<i>Median</i>	<i>Mode</i>	<i>Range</i>
Escorts (e.g., internal and outside)	2	2	1-4	2	2	1-4	2	2	1-4
Using Chemical and Inflammatory Agents	1	1	1-3	1	1	1-2	1	1	1-3
Utilizing Fire Safety Measures	1	1	1-5	1	1	1-5	2	2	1-2
Using Self Defence Skills	1	1	1-3	1	1	1-2	1	1	1-3
Using Arrest and Control Skills	1	1	1-3	1	1	1-3	1	1	1-2
Using Baton Skills	1	1	1-1	1	1	1-1	1	1	1-1

Note. Supervisors were given the option to indicate that there was no opportunity to observe this skill; however, none of the supervisors provided this response for any of the frequency items.

For the quality component, supervisors were asked to rate the level of quality with which officers executed the behaviours/skills that were rated on the frequency scale. Again, quality was conceptualized as the consistency with which officers executed the skills in the way they were trained to during CTP. Quality was measured on the same three-point scale ranging from 1 (below average) to 3 (above average); with an option to indicate that they had no opportunity to observe that skill executed by the officer. Higher scores on the quality component indicate a more successful execution of these skills according to supervisors.

Overall, the mean checklist scores for quality were slightly higher than the ratings provided by trainees themselves, although still moderate ($M = 2.41$, $SD = .49$, range = 1.67-3.50) (see Table 26 for the individual items). This means that supervisors felt that officers were able to execute the sampled behaviours/skills in a manner that was generally consistent with how they were taught to during CTP, more in line with expectations about higher quality ratings. Similar results were found for each region (Pacific: $M = 2.85$, $SD = .34$, range = 2.40-3.60; Ontario: $M = 2.91$, $SD = .30$, range = 2.60-3.25). As expected, an independent samples t -test revealed that no significant regional differences existed, $t(23) = .49$, $p > .05$.

Table 26

Descriptive Statistics for the Supervisor Behavioural Checklist - Quality Items

Behaviours/Skills	Overall				Pacific				Ontario			
	<i>n</i>	NOO (<i>n</i>) ^a	<i>Mean</i> (<i>SD</i>)	<i>Range</i>	<i>n</i>	NOO (<i>n</i>)	<i>Mean</i> (<i>SD</i>)	<i>Range</i>	<i>n</i>	NOO (<i>n</i>)	<i>Mean</i> (<i>SD</i>)	<i>Range</i>
Assessing Problematic Situations	23	2	2.35 (.49)	1-3	14	2	2.29 (.47)	2-3	9	0	2.44 (.53)	2-3
Identifying Possible Solutions	23	2	2.26 (.54)	1-3	14	2	2.14 (.54)	1-3	9	0	2.44 (.53)	2-3
Taking Necessary Measures	24	1	2.25 (.53)	1-3	15	1	2.13 (.52)	1-3	9	0	2.44 (.53)	2-3
Motivating and Influencing Individuals	23	2	2.13 (.55)	1-3	15	1	2.07 (.46)	1-3	8	1	2.25 (.71)	1-3
Communicating Effectively Orally	25	0	2.48 (.51)	1-3	16	0	2.38 (.50)	2-3	9	0	2.67 (.50)	2-3
Communicating Effectively in Writing	25	0	2.52 (.51)	1-3	16	0	2.44 (.51)	2-3	9	0	2.67 (.50)	2-3
Summarizing Information	24	1	2.38 (.50)	1-3	15	1	2.33 (.49)	2-3	9	0	2.44 (.53)	2-3
Report Writing: OSOR	25	0	2.44 (.51)	1-3	16	0	2.31 (.48)	2-3	9	0	2.67 (.50)	2-3

Behaviours/Skills	Overall				Pacific				Ontario			
	<i>n</i>	NOO (<i>n</i>)	<i>Mean</i> (<i>SD</i>)	<i>Range</i>	<i>n</i>	NOO (<i>n</i>)	<i>Mean</i> (<i>SD</i>)	<i>Range</i>	<i>n</i>	NOO (<i>n</i>)	<i>Mean</i> (<i>SD</i>)	<i>Range</i>
Post Search	18	7	2.33 (.49)	1-3	10	6	2.30 (.48)	2-3	8	1	2.38 (.52)	2-3
Use of Force	8	17	2.13 (.64)	1-3	4	12	1.75 (.50)	1-2	4	5	2.50 (.58)	2-3
Offence	14	11	2.43 (.51)	1-3	8	8	2.25 (.46)	2-3	6	3	2.67 (.52)	2-3
Using Communication Equipment (e.g., two-way radio, loudhailer, etc.)	25	0	2.54 (.50)	1-3	16	0	2.50 (.52)	2-3	9	0	2.62 (.49)	2-3
Using Restraint Equipment	17	8	2.41 (.51)	1-3	10	6	2.30 (.48)	2-3	7	2	2.57 (.54)	2-3
Engaging in a Search (e.g., area, frisk, strip, etc.)	24	1	2.38 (.50)	1-3	15	1	2.20 (.41)	2-3	9	0	2.67 (.50)	2-3
Escorts (e.g., internal and outside)	20	5	2.30 (.47)	1-3	12	4	2.00 (.00)	2-2	8	1	2.75 (.46)	2-3
Using Chemical and Inflammatory Agents	3	22	2.33 (.58)	1-3	1	15	2.00 (.00)	2-2	2	7	2.50 (.71)	2-3
Utilizing Fire Safety Measures	11	14	2.45 (.52)	1-3	5	11	2.40 (.55)	2-3	6	3	2.50 (.55)	2-3
Using Self Defence Skills	4	21	2.50 (.58)	1-3	2	14	2.50 (.71)	2-3	2	7	2.50 (.71)	2-3

Behaviours/Skills	Overall				Pacific				Ontario			
	<i>n</i>	NOO (<i>n</i>)	<i>Mean</i> (<i>SD</i>)	<i>Range</i>	<i>n</i>	NOO (<i>n</i>)	<i>Mean</i> (<i>SD</i>)	<i>Range</i>	<i>n</i>	NOO (<i>n</i>)	<i>Mean</i> (<i>SD</i>)	<i>Range</i>
Using Arrest and Control Skills	9	16	2.22 (.44)	1-3	7	9	2.14 (.38)	2-3	2	7	2.50 (.71)	2-3
Using Baton Skills	0	25	-	-	0	16	-	-	0	9	-	-

Note. ^aSupervisors were given the option to indicate that there was no opportunity to observe this skill (abbreviated as NOO to save space) and it was not included in the calculation of the mean; however, the number of supervisors who provided this response is provided.

For a variety of reasons it is important to stress that the ratings provided by supervisors be interpreted with a degree of caution. First, only half of the trainees had a checklist completed by their supervisor. Second, it was brought to the author's attention by members of the evaluation team with institutional experience, as well as supervisors themselves, that there is not always ample opportunity to directly observe correctional officers while on duty. In addition, depending on when the supervisor themselves completed CTP; their knowledge of the current program may have been limited. Due to these issues, the Supervisor Checklist scores will not be examined beyond the descriptive statistics and the correlational analyses between evaluation levels (presented below).

Chapter Summary

The majority of the hypotheses with regard to training outcomes were generally supported (a more in-depth discussion of the results is provided below in Chapter 7). For example, it was hypothesized that positive ratings of the training program were expected and trainees did in fact rate CTP very positively with an average rating of 4.59 out of 5 with no regional differences observed. When the items of the Level 1 questionnaire were divided into subsections (i.e., training materials, facilitators, etc.) and question types (i.e., utility, difficulty, etc.), despite high internal reliability, high correlations indicated it was a one-dimensional measure of trainee reactions. This precluded any additional analyses with the subsections.

Similarly, for the Level 2 outcomes, it was hypothesized that significant increases in test scores (regardless of test subsection, skill or region) would be observed across each of the training outcomes between the recruitment and training phases to indicate that learning had occurred and that most of what was learned would be retained at the transfer

phase. Indeed, significant increases in knowledge test scores were observed, as expected, between the recruitment and training phases. While a small but significant loss was observed between the training and transfer phases, knowledge levels remained high. Although no regional differences were found, in contrast to expectations there was a difference in knowledge levels based on test subsection. Specifically, at the recruitment phase (pre-CTP) significantly lower scores were observed for the Stage 3 material compared to Stages 1 and 2. Significant differences between subsections were also found at the other two phases.

Regional differences were also found for the soft skills outcome, in contrast to expectations, with higher scores observed in the Pacific compared to the Ontario region at the training phase. Similarly, when looking at the role play and written components of the soft skills outcome measure, regional as well as test subsection differences were observed. Specifically, in Ontario lower scores were observed in the recruitment phase compared to the latter two phases and scores on the role play component were higher than the written component regardless of study phase. In the Pacific region, scores on the role play component were also significantly greater than the written component across all phases with greater differences observed at the training and transfer phases.

Hard skills were measured only at the training and transfer phases and a significant decrease was observed here. Although it is acknowledged that some loss is expected, the hard skills levels remained high at transfer. In contrast to the hypothesis, differences in individual skills scores were observed at each phase indicating that gains in skills scores differed based on the specific skill. Specifically, at the training phase, scores for the frisk search were significantly lower than the majority of other skills while the

scores for OC spray and radio use were higher and not significantly different than most of the other skills. At the transfer phase, there were more significant differences among skills than that observed in the training phase. Scores for the frisk search as well as the high profile escort were significantly lower than all other skills, while the scores for the OC spray alone remained significantly higher than all other skills.

In contrast to expectations, no significant changes in attitude outcomes were observed between recruitment and training; however, significant losses were observed between the training and transfer phases. Although some loss was expected, the loss was relatively great for the ATI outcome specifically. In fact, lower levels of ATI ratings were observed across all phases of the study. No regional differences were observed here, consistent with expectations.

The hypotheses for the Level 3 outcomes, of high frequency and quality ratings regardless of region, were generally supported, although not fully. For example, frequency ratings of skill use on the job varied depending on the skill. Higher frequencies were observed for the soft skills (e.g., assessing problematic situations), whereas low ratings (including no opportunity to use the skill) were provided for the majority of hard skills (e.g., using restraint equipment). Moderate to high levels of quality ratings were provided indicating that skills were used in a way that was generally consistent with how CTP trains its trainees to use the skills, although higher ratings were expected. No regional differences were observed, as expected. Although supervisors provided ratings that were similar to trainees, in contrast to expectations, correlations only reached significance for one item (use of chemical agents).

CHAPTER 5

Correlations Between Each Level of the Evaluation Framework

In order to examine the relationships between the different levels of Kirkpatrick's expanded evaluation framework (Level 1, 2, and 3), bivariate correlations were calculated (with a modified Bonferroni correction employed in order to control for the rate of Type I errors). The results of the psychometric analyses described above for the reaction data were favourable, therefore the reaction data was examined both in aggregate form as well as per subscale and item type (e.g., training materials overall, general satisfaction, etc.). The various subsections of the Level 2 outcomes examined above were also generally associated with reasonable psychometric results and were included in this analysis. Finally, the individual items of the trainee behavioural checklist (relating to both frequency and quality) as well as the aggregate score for the quality items were included.

Hypotheses

Recall from Chapter 1, that the following hypothesis was tested with respect to relationships between levels of the evaluation framework:

- (1) In general, significant positive correlations between the outcome measures at one level (e.g., Level 1) and the outcome measures at another level (e.g., Level 2) were expected. This is based on the inherent assumption of Kirkpatrick's evaluation framework, that results from one level will predict results from the next. This applies in the current study for relationships between measures at Level 1 and 2, Level 1 and 3, and Level 2 and 3.

Level 1 & 2

Bivariate correlations between Level 1 measures (trainee reaction overall, general satisfaction items, utility items and difficulty items) and Level 2 measures (knowledge test, ATCW, ATI, soft skills assessment, and hard skills) were calculated using Pearson's r . Similar to the analyses presented above, these analyses were carried out separately for the Level 2 measures at each of the study phases (recruitment, training and transfer).

For all three phases, few significant correlations were found between any of the trainee reaction measures and the various learning outcomes (see Table 27-29). Although this finding is in contrast to expectations (based on the assumptions of Kirkpatrick's evaluation framework), it is, however, consistent with the majority of the research literature, which suggests that it is problematic to utilize reaction ratings to imply that learning has taken place (e.g., Alliger & Janak, 1989; Alliger et al., 1997). With very few exceptions, the only significant correlations that were found were within the Level 1 trainee reaction measures themselves (i.e., the subscales of the reaction measure were significantly correlated with each other as well as with the overall score) as well as within the Level 2 outcomes themselves (i.e., the subscales of the knowledge tests were significantly correlated with the overall score, etc.).

Table 27

Correlation Matrix for the Level 1 Outcome (Training Phase) and Level 2 Outcomes (Recruitment Phase)

Outcomes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Trainee Reactions Overall	-																		
2. Training Materials	.93*	-																	
3. Training Program	.90*	.82*	-																
4. Stages of CTP	.86*	.79*	.72*	-															
5. Stage 1	.75*	.70*	.82*	.73*	-														
6. Stage 3	.88*	.83*	.90*	.75*	.81*	-													
7. Facilitators	.81*	.68*	.86*	.61*	.70*	.80*	-												
8. General Satisfaction Items	.98*	.90*	.90*	.91*	.78*	.89*	.81*	-											
9. Utility Items	.81*	.74*	.82*	.72*	.70*	.77*	.70*	.82*	-										
10. Difficulty Items ^a	.90*	.86*	.90*	.77*	.72*	.81*	.79*	.89*	.77*	-									
11. Knowledge Test	-.20	-.23	-.22	-.09	-.14	-.17	-.11	-.19	-.21	-.19	-								
12. Knowledge Test (Stage 1)	-.42*	-.45*	-.34	-.30	-.14	-.35	-.20	-.38	-.32	-.41	.47*	-							
13. Knowledge Test (Stage 2)	-.08	-.08	-.13	-.05	-.7	-.07	-.10	-.08	-.13	-.14	.52*	.40*	-						
14. Knowledge Test (Stage 3)	-.07	.10	-.10	.04	-.11	-.05	-.03	-.05	-.10	-.04	.91*	.11	.27	-					

Outcomes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
15. Soft Skills Assessment	-.17	-.20	-.22	-.08	-.12	-.22	-.22	-.16	-.06	-.23	.10	.13	.03	.07	-				
16. Soft Skills (Role Play)	-.07	-.13	-.10	.02	-.05	-.14	-.15	-.06	.03	-.12	.08	.00	-.03	.10	.89*	-			
17. Soft Skills (Written)	-.24	-.21	-.29	-.18	-.17	-.23	-.22	-.23	-.17	-.29	.09	.27	.10	-.02	.67*	.25	-		
18. ATCW	.24	.23	.36	.21	.37	.27	.24	.26	.10	.22	.04	-.07	.03	.08	-.16	-.10	-.19	-	
19. ATI	.06	.09	.05	.06	.07	.07	.05	.09	.11	-.02	-.01	-.21	-.21	.09	.13	.15	.04	.17	-

Note. Bonferroni correction = $.05/171 = .0003$.

^aRecall that the difficulty items are reverse scored.

*Correlations significant at the corrected alpha level.

Table 28

Correlation Matrix for the Level 1 Outcome (Training Phase) and Level 2 Outcomes (Training Phase)

Outcomes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. Trainee Reactions Overall	-																			
2. Training Materials	.93*	-																		
3. Training Program	.90*	.82*	-																	
4. Stages of CTP	.86*	.79*	.72	-																
5. Stage 1	.75*	.70*	.82*	.73*	-															
6. Stage 3	.88*	.83*	.90*	.75*	.81*	-														
7. Facilitators	.81*	.68*	.86*	.61*	.70*	.80*	-													
8. General Satisfaction Items	.98*	.90*	.90*	.91*	.78*	.89*	.81*	-												
9. Utility Items	.81*	.74*	.82*	.72*	.70*	.77*	.70*	.82*	-											
10. Difficulty Items ^a	.90*	.86*	.90*	.77*	.72*	.81*	.79*	.89*	.77*	-										
11. Knowledge Test	-.14	-.11	-.13	-.15	-.28	-.22	-.18	-.16	-.23	-.09	-									
12. Knowledge Test (Stage 1)	-.06	-.04	-.07	-.14	-.09	-.20	-.05	-.09	-.02	-.06	.49*	-								
13. Knowledge Test (Stage 2)	.10	.10	.08	.04	-.01	.11	-.06	.08	-.03	.06	.35*	.01	-							
14. Knowledge Test (Stage 3)	-.17	-.14	-.14	-.10	.27	-.18	-.15	-.16	-.25	-.08	.80*	.05	.01	-						
15. Soft Skills Assessment	-.22	-.24	-.10	-.10	.15	-.11	-.06	-.17	.03	-.20	-.24	-.14	-.27	-.13	-					
16. Soft Skills Assessment (Role Play)	-.17	-.22	-.08	-.09	.12	-.12	-.04	-.14	.07	-.18	-.20	-.04	-.32	-.12	.92*	-				

Outcomes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
17. Soft Skills Assessment (Written)	-.20	-.18	-.08	-.07	.15	-.04	-.07	-.14	-.04	-.15	-.20	-.25	-.09	-.10	.76*	.43*	-			
18. Hard Skills	-.05	-.05	-.05	.05	-.08	-.03	.02	.01	.05	-.09	-.08	-.18	.01	-.01	.20	.14	.22	-		
19. ATCW	.13	.06	.26	.06	.19	.17	.32	.17	.17	.21	.12	.16	-.22	.14	.01	.01	.02	-.02	-	
20. ATI	.13	.12	.09	.07	.11	.11	.14	.13	.19	.06	-.16	.01	-.17	-.11	-.03	-.05	.00	-.02	.15	-

Note. Bonferroni correction = $.05/190 = .0003$.

^aRecall that the difficulty items are reverse scored.

*Correlations significant at the corrected alpha level.

Table 29

Correlation Matrix for the Level 1 Outcome (Training Phase) and Level 2 Outcomes (Transfer Phase)

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. Trainee Reactions Overall	-																			
2. Training Materials	.93*	-																		
3. Training Program	.90*	.82*	-																	
4. Stages of CTP	.86*	.79*	.72*	-																
5. Stage 1	.75*	.70*	.82*	.73*	-															
6. Stage 3	.88*	.83*	.90*	.75*	.81*	-														
7. Facilitators	.81*	.68*	.86*	.61*	.70*	.80*	-													
8. General Satisfaction Items	.98*	.90*	.90*	.91*	.78*	.89*	.81*	-												
9. Utility Items	.82*	.74*	.82*	.72*	.70*	.77*	.70*	.82*	-											
10. Difficulty Items ^a	.90*	.86*	.90*	.77*	.72*	.80*	.79*	.89*	.77*	-										
11. Knowledge Test	-.20	-.14	-.24	-.07	-.25	-.22	-.19	-.19	-.21	-.15	-									
12. Knowledge Test (Stage 1)	-.07	-.00	-.17	.08	-.07	-.05	-.07	-.06	-.10	-.15	.57*	-								
13. Knowledge Test (Stage 2)	.10	.12	.06	.05	.04	.04	.10	.08	-.08	.11	.47*	.21	-							

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
14. Knowledge Test (Stage 3)	-.19	-.13	-.21	-.06	-.25	-.21	-.19	-.17	-.15	-.09	.90*	.32	.33	-						
15. Soft Skills Assessment	-.19	-.17	-.30	-.02	-.16	-.20	-.35	-.16	-.23	-.28	.00	.04	-.03	.00	-					
16. Soft Skills Assessment (Role Play)	-.18	-.15	-.28	.04	-.12	-.16	-.28	-.13	-.24	-.23	-.07	.00	-.09	-.08	.89*	-				
17. Soft Skills Assessment (Written)	-.12	-.11	-.20	-.10	-.14	-.18	-.29	-.13	-.11	-.21	.11	.09	.07	.13	.72*	.31	-			
18. Hard Skills	-.18	-.10	-.22	-.09	-.27	-.28	-.24	-.18	-.10	-.10	.22	-.03	.08	.23	-.00	-.02	.03	-		
19. ATCW	.32	.19	.22	.16	.10	.16	.32	.29	.06	-.17	-.03	-.04	.20	-.06	.18	.14	.16	-.17	-	
20. ATI	-.10	-.05	-.12	-.18	-.13	-.08	-.07	-.10	-.19	-.16	-.02	.03	.13	.00	.02	-.06	.14	-.07	.29	-

Note. Bonferroni correction = $.05/190 = .0003$.

^aRecall that the difficulty items are reverse scored.

*Correlations significant at the corrected alpha level.

Level 1 & 3

Bivariate correlations (Spearman's r_s and Pearson's r) were also calculated between Level 1 measures (trainee reaction overall, general satisfaction items, utility items and difficulty items) and Level 3 measures (individual frequency and quality items as well as quality overall for the trainee behavioural checklist). Note that the analyses were carried out for the Level 3 outcome measures for the trainees only (as supervisor ratings on the job would not be expected to correlate with trainee reactions during training). In contrast to expectations, no significant correlations were found after the Bonferroni correction was applied to the results other than those within the Level 1 measures themselves (see Table 30; correlations for the individual frequency and quality items of the trainee behavioural checklist were not provided in order to save space and for the sake of brevity).

One would generally expect reactions to the training (as items from this questionnaire tap into the perceived practicality and utility of training) to be somewhat related to the transfer measures (i.e., Level 3 outcomes). Perhaps, however, reactions to training are not related to how frequently one engages in certain skills on the job, or how well the skills are executed because of the impact of the transfer climate itself. Rather, the frequency and quality of skill use may be determined much more so by the realities of the job than reactions to training (i.e., the number of incidents with inmates, the level of security of the institution, the post within the institution that is worked by the officer, etc.). For example, a trainee who rated CTP very positively would not be able to express their arrest and control skills (and express these skills with high quality) if there have not been opportunities to do so within their assigned institution. Furthermore, research

indicates that relationships between Level 1 and 3 measures are rarely examined and when they are, despite the assumptions of Kirkpatrick's framework, the correlations are rarely significant (Alliger et al., 1997).

Table 30

Correlation Matrix for the Level 1 and 3 Outcome Measures

Variables	1	2	3	4	5	6	7	8	9	10	11
1. Trainee Reactions Overall	-										
2. Training Materials	.93*	-									
3. Training Program	.90*	.82*	-								
4. Stages of CTP	.86*	.79*	.72*	-							
5. Stage 1	.75*	.70*	.82*	.73*	-						
6. Stage 3	.88*	.83*	.90*	.75*	.81*	-					
7. Facilitators	.81*	.68*	.86*	.61*	.70*	.80*	-				
8. General Satisfaction Items	.98*	.90*	.90*	.91*	.78*	.89*	.81*	-			
9. Utility Items	.82*	.74*	.82*	.72*	.70*	.77*	.70*	.82*	-		
10. Difficulty Items ^a	.90*	.86*	.90*	.77*	.72*	.80*	.79*	.89*	.77*	-	
11. TBC Quality Overall	.26	.22	.25	.18	.17	.10	.23	.26	.08	.23	-

Note. Bonferroni correction = $.05/210 = .0002$. The Bonferroni calculation includes the non-significant individual items of the Level 3 measures not shown here.

^aRecall that the difficulty items are reverse scored.

*Correlations significant at the corrected alpha level..

Level 2 & 3. Finally, bivariate correlations between Level 2 measures (knowledge, skills and attitudes) and Level 3 measures (individual frequency and quality items as well as quality overall for the trainee and supervisor behavioural checklists) were calculated using Spearman's r_s and Pearson's r (for analyses related to the frequency data and the qualitative data, respectively) (see Table 31).

In general, the hypothesis was not supported. Although the subsection scores for the Level 2 outcomes were included in the analyses, none were significantly correlated with the Level 3 outcomes. The only significant correlations with respect to these subsections were within the Level 2 outcomes themselves (e.g., subsections with other subsections, as shown above in the correlation matrices for the training outcome results). Additionally, none of the individual frequency and quality items from trainee and supervisor behavioral checklists were significantly correlated with any of the other measures. Therefore, in an effort to limit redundancy and save space, the results of the Level 2 subsections and individual items were not included in the correlation matrix below.

For all study phases there was only one significant correlation after the Bonferroni correction was applied. A significant correlation was found between the ATCW ratings (collected at the recruitment phase) and the quality ratings overall on the trainee behavioural checklist ($r = .49, p = .000$), such that a positive relationship was found between attitudes towards correctional work before training and the self-reported quality of skills executed on the job. Some other correlations approached significance. For example, although not significant, the ATCW and ATI measures (collected during the transfer phase) were correlated with the quality scores overall, provided by the trainees in

the transfer phase ($r = .40, p = .003$; $r = .36, p = .008$, respectively). This suggests that there is a potential positive relationship between attitudes towards correctional work and inmates after training and the self-reported quality of skills executed on the job.

Table 31

Correlation Matrix for the Level 2 and 3 Outcome Measures

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Knowledge Test (Recruitment)	-															
2. Knowledge Test (Training)	.26	-														
3. Knowledge Test (Transfer)	.36	.54*	-													
4. Soft Skills Assessment (Recruitment)	.10	.10		-												
5. Soft Skills Assessment (Training)	.17	-.24	.05	.17	-											
6. Soft Skills Assessment (Transfer)	.22	-.12	-.00	.28	.31	-										
7. Hard Skills (Training)	.13	-.08	.08	-.02	.20	-.09	-									
8. Hard Skills (Transfer)	.06	.22	.22	.01	-.04	-.00	.34	-								
9. ATCW (Recruitment)	-.06	.04	-.10	-.21	-.11	-.19	-.14	-.25	-							
10. ATI (Recruitment)	-.10	-.06	-.04	.13	-.14	.15	-.00	-.08	.15	-						
11. ATCW (Training)	.10	.12	.15	.05	.01	-.19	-.02	-.12	.39	-.04	-					
12. ATI (Training)	-.14	-.16	-.17	.07	-.03	-.06	-.02	.06	.02	.13	.15	-				
13. ATCW (Transfer)	-.12	-.14	-.03	.03	-.11	.18	-.10	-.17	.35	.08	.41	.07	-			
14. ATI (Transfer)	-.16	-.18	.02	-.01	-.13	.02	-.06	-.07	.08	.52	.07	.51*	.29	-		

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
15. TBC Quality Overall	-.27	.00	-.19	-.04	-.10	-.17	-.24	-.09	.49*	.14	.35	.25	.40	.36	-	
16. SBC Quality Overall	.04	-.12	-.04	-.08	.34	.01	-.01	-.16	-.29	.03	-.20	-.04	.07	.24	-.25	-

Note. Bonferroni correction = $.05/820 = .00006$. The Bonferroni calculation includes the non-significant subsections of the Level 2 measures not shown here as well as the individual frequency and quality items of the Level 3 measures.

*Correlations significant at the corrected alpha.

If significant correlations had been found between the levels, hierarchical multiple regression (HMR) analyses were planned in order to determine whether the expanded form of Kirkpatrick's (1959a) training evaluation framework provides added value compared to the original version. For example, within the current context, HMR would have enabled the author to examine the power of certain evaluation criteria (e.g., attitude levels) that are not typically included in Kirkpatrick's model for predicting other training outcomes (e.g., ratings of skill quality while on the job) while controlling for the predictive power of evaluation criteria (e.g., reaction scores) that are included in Kirkpatrick's model (Axtell et al., 1997; Tan et al., 2003). In the present study, HMR was not carried out because so few significant relationships between the evaluation levels were found.

Chapter Summary

In general, the hypothesis that outcomes at lower levels of the framework will predict outcomes at higher levels of the framework was not supported, with only a few significant relationships found between measures at the differing levels (a more in-depth discussion of these results is provided in Chapter 7). While no significant correlations were found between Level 1 and 3 measures, a few significant correlations were found between Level 1 and 2 measures as well as Level 2 and 3 measures. In contrast to expectations, a negative correlation was observed between knowledge test scores on the Stage 1 subsection prior to CTP (Level 2) and trainee reactions during the training phase (Level 1) ($r = -.42$). Similarly, a significant negative correlation was found between knowledge test scores on the Stage 1 subsection prior to CTP (Level 2) and trainee reactions on the training materials subsection of the questionnaire during the training

phase (Level 1) ($r = -.45$). These results indicate that a negative relationship exists between Stage 1 test scores before training and trainee reactions during training (both overall and specifically related to training materials). The only other significant correlation was found between the ATCW ratings at the recruitment phase (Level 2) and the quality ratings of skill use overall at the transfer phase (Level 3) ($r = .49$). This finding suggests that there is indeed a positive relationship between attitudes towards correctional work before training and the self-reported quality of skill use (overall) on the job.

CHAPTER 6

The Influence of Training Effectiveness Factors

The examination of training effectiveness factors consisted of several different sets of analyses. The first set of analyses examined the reliability of the subscales of the LTSI. This is an important component of the current study given that the LTSI has never been used within the correctional context (and thus might not generalize to it). This analysis adds to the somewhat limited pool of psychometric data that is available for the LTSI. Although an exploratory factor analysis would have also been desirable, this was not possible in the current study. This is because Holton and his colleagues analyzed all of the raw data that was collected as part of the agreement for utilizing their instrument. Specifically, Holton's colleague (Reid Bates) calculated all of the subscale scores and reliability statistics. At no time was information shared about the content of each subscale (i.e., items making up each subscale). This information is also not available in the published research on the LTSI.

Informed by the reliability of the LTSI subscales, the next set of analyses determined how the training effectiveness factors related to the training outcomes obtained in Levels 1 through 3. This was accomplished by calculating correlations between the training effectiveness measures and the training outcomes. These analyses provide insight, for example, into whether those who rate themselves as possessing higher levels of self-efficacy also know, learn, and transfer more than those who rate themselves as possessing lower levels of self-efficacy.

Lastly, in order to determine how the LTSI subscales moderate the relationships between the different levels of the expanded Kirkpatrick framework, moderated multiple

regression (MMR) analyses were conducted. The use of MMR allows us to answer questions of the general form: Does training effectiveness variable A moderate the relationship that is found between evaluation criterion X and evaluation criterion Y? Or, as a more specific example: Is the lack of a significant correlation between learning outcomes after CTP (Level 2 learning data) and trainee ratings of quality of on the job behaviour (Level 3 transfer data) explained at all by the lack of supervisor support for training that is experienced by trainees (a training effectiveness variable)?

Sample size was a concern for this type of analysis. In most cases, the guidelines for MMR with respect to sample size suggest that sample size to variable ratios of 15:1 or 10:1 are most desirable (Cohen & Cohen, 1983; Field, 2009). Clearly, this requirement prevented the analysis of all 17 LTSI subscales. However, not all of the scales had high reliability scores in this study (details are provided below), which is also a requirement of MMR (Cohen & Cohen, 1983; Frazier, Tix, & Barron, 2004). Thus, by analyzing a smaller set of highly reliable subscales, an attempt was made to overcome this issue. In addition, it is important to note that not all subscales were relevant for all MMR analyses and thus, some subscales could be eliminated from analyses. For example, it makes little sense to examine how perceptions of supervisor support moderate the relationship between reaction-level data and learning-level data because supervisor support will only matter once the trainee is working within an institution.

Hypotheses

Recall from Chapter 1, the following hypotheses were tested with respect to the training effectiveness factors examined in the current study:

- (1) Based on the existing literature that has examined the psychometric properties of the LTSI, it is expected that high reliability levels will be found for the majority of the 17 subscales.
- (2) In addition, based on relationships found in the existing literature, positive correlations are expected between the training outcomes and the majority of the LTSI subscales (with the exception of supervisor opposition which is measured in a negative direction).
- (3) In addition, it is expected that the majority of the subscales will be rated at moderate to high levels by trainees.

Despite expectations with regard to subscale reliability, only those with reliability scores of .80 or greater will be used for the moderator analysis. At this time it is not known which scales will meet that criterion. If highly reliable scales are identified, moderator analyses will be carried out and depending on the specific moderators identified; the predictor and outcome variables of interest will be outlined (including training outcome variables from the various study phases).

- (4) In general, it is hypothesized that each moderator will strengthen the relationship between the predictor and outcome variables.

LTSI

The individual subscale reliabilities in the current study range from .41 to .92 (see Table 32). These values seem relatively consistent with our hypothesis as well as previous research (Holton et al., 2000), with the exception of the following scales that displayed low reliabilities (<.60): personal capacity for transfer, opportunity to use, and transfer effort-performance expectations. Although the factor structure could not be

verified via factor analysis in the current sample, the subscale reliabilities provide valuable information to inform us of which factors are suitable for further analyses. The training effectiveness variables included in the subsequent analysis were the LTSI scales with the highest reliability scores ($\geq .80$), specifically: supervisor support, perceived content validity, resistance-openness to change, transfer design and performance self-efficacy.

The descriptive statistics for the various LTSI subscales are also presented in Table 32 for the sample as a whole as well as for the Pacific and Ontario regions separately. Higher scores on these subscales (out of a maximum of 5) indicate a more favorable rating by trainees, with the exception of the supervisor opposition subscale where lower scores indicate less supervisor opposition. Overall, the means for each of the subscales ranged from 2.45 to 3.95 indicating that trainees felt anywhere from neutral (e.g., peer support) to positive (e.g., motivation to transfer) for the majority of the subscales, which is consistent with expectations. Correlations between the various subscales of the LTSI are provided in Appendix CC.

Bivariate correlations between the retained training effectiveness factors and the various training outcomes were calculated using Pearson's r . These correlations are provided in Table 33. In contrast to expectations, only one significant positive correlation was found between an LTSI subscale and a training outcome at the corrected alpha level. Specifically, attitudes towards correctional work at the transfer phase (i.e., on the job) were significantly correlated with transfer design ratings. The only other significant correlations were among the training outcomes or LTSI factors themselves. With regard to the LTSI subscales, ratings on openness to change and supervisor ratings were

significantly positively correlated ($r = .61$). In addition ratings on both self-efficacy and transfer design were positively correlated with content validity ($r = .49$ and $.58$, respectively).

Table 32

Descriptive Statistics and Internal Consistency Scores for the Subscales of the LTSI (n = 53)

LTSI Subscale	Overall (n = 53)			Pacific (n = 26)		Ontario (n = 24)	
	<i>M (SD)</i>	<i>Range</i>	<i>Alpha</i>	<i>M (SD)</i>	<i>Range</i>	<i>M (SD)</i>	<i>Range</i>
Supervisor Support	2.66 (.77)	1.00-4.17	.92	2.78 (.83)	1.00-4.17	2.54 (.74)	1.00-3.67
Perceived Content Validity	3.07 (.90)	1.00-4.40	.88	3.09 (.96)	1.00-4.40	2.97 (.83)	1.40-4.40
Resistance-Openness to Change	2.85 (.69)	1.00-4.17	.85	2.86 (.83)	1.00-4.17	2.88 (.55)	1.33-3.83
Transfer Design	3.81 (.55)	2.00-5.00	.82	3.98 (.41)	3.25-5.00	3.62 (.62)	2.00-4.50
Performance Self-Efficacy	3.62 (.56)	2.00-4.75	.80	3.47 (.63)	2.00-4.75	3.75 (.45)	2.50-4.75
Negative Personal Outcomes	2.91 (.67)	1.00-4.00	.77	3.00 (.63)	1.75-4.00	2.83 (.75)	1.00-4.00
Peer Support	3.12 (.69)	1.75-5.00	.76	3.13 (.66)	1.75-4.00	3.10 (.78)	2.00-5.00
Intent to Transfer	3.91 (.39)	3.00-4.75	.75	3.87 (.41)	3.00-4.75	3.93 (.38)	3.00-4.75
Positive Personal Outcomes	2.45 (.66)	1.00-3.67	.70	2.50 (.63)	1.00-3.67	2.35 (.70)	1.00-3.33
Performance-Outcomes Expectations	3.38 (.55)	2.20-4.60	.70	3.39 (.58)	2.20-4.40	3.33 (.54)	2.20-4.60
Supervisor Opposition ^a	2.50 (.55)	1.00-4.33	.68	2.44 (.52)	1.33-3.33	2.57 (.62)	1.00-4.33
Learner Readiness	3.14 (.65)	1.50-4.00	.68	3.16 (.68)	1.75-4.00	3.03 (.63)	1.50-4.00
Performance Coaching	3.43 (.54)	1.50-4.25	.63	3.43 (.63)	1.50-4.00	3.42 (.49)	2.50-4.25
Motivation to Transfer	3.95 (.49)	3.00-5.00	.62	3.95 (.50)	3.00-4.50	3.87 (.42)	3.25-4.75
Opportunity to Use	3.45 (.53)	2.00-4.75	.58	3.38 (.61)	2.00-4.75	3.58 (.41)	2.50-4.25

LTSI Subscale	Overall (<i>n</i> = 53)			Pacific (<i>n</i> = 26)		Ontario (<i>n</i> = 24)	
	<i>M</i> (<i>SD</i>)	<i>Range</i>	<i>Alpha</i>	<i>M</i> (<i>SD</i>)	<i>Range</i>	<i>M</i> (<i>SD</i>)	<i>Range</i>
Transfer Effort-Performance Expectations	3.87 (.39)	3.00-5.00	.58	3.88 (.43)	3.00-5.00	3.86 (.35)	3.25-4.50
Personal Capacity for Transfer	3.39 (.48)	2.00-4.25	.41	3.35 (.53)	2.00-4.25	3.42 (.41)	2.50-4.25

Note. The subscales are listed in order of their alpha levels (highest to lowest). Alphas are only provided for the sample overall as this data was analyzed by Reid Bates and he was not given instructions to provide this information per region at the time of analysis.

^aRecall the Supervisor Opposition subscale is a negative scale so higher average scores reflect more opposition.

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	
15. Trainee Reactions Overall	-.20	-.14	-.20	-.17	-.22	-.19	-.05	-.18	.24	.06	.13	.13	.32	-.10	-								
16. TBC Quality Overall	-.28	.00	-.19	-.03	-.10	-.17	-.24	-.09	.43	.14	.35	.25	.40	.36	.26	-							
17. SBC Quality Overall	.04	-.12	-.04	-.08	.34	.01	-.01	-.16	-.29	.03	-.20	-.04	.07	.24	-.05	-.25	-						
18. Supervisor Support	-.27	-.28	-.30	-.16	-.30	-.20	-.02	-.22	.30	.08	.16	.16	.39	.31	.43	.34	-.13	-					
19. Content Validity	.06	-.09	.06	.10	.12	-.12	.18	-.01	.16	.01	.31	-.03	.35	.29	.29	.25	.13	.44	-				
20. Openness to Change	-.20	-.04	-.14	-.11	-.28	-.30	.07	-.14	.36	.07	.28	.29	.23	.13	.43	.32	-.17	.61*	.30	-			
21. Personal Self-Efficacy	-.06	-.22	-.08	.11	-.10	-.05	.18	.09	.05	.03	.10	.04	.35	.21	.31	.14	-.11	.41	.49*	.44	-		
22. Transfer Design	.01	-.06	.05	.19	-.17	-.04	-.12	-.05	.05	-.04	.26	-.08	.54*	.15	.50	.24	-.07	.38	.58*	.10	.41	-	

Note. Bonferroni correction = $.05/253 = .0002$.

*Correlations significant at the corrected alpha level.

In order to determine if there were regional differences in the ratings for the 17 scales of the LTSI a MANOVA was carried out. The independent variable was Region (Pacific versus Ontario) and the dependent variables were the scores on the 17 subscales of the LTSI. Recall that during the data screening process it was determined that the results of the main analyses did not differ when using transformed and non-transformed data; therefore, the results of the MANOVA with non-transformed data were retained for ease of interpretation. Assumptions of homogeneity of variance and covariance were met. Using Pillai's trace, in contrast to expectations, there was a significant effect of region on the subscale scores, $V = .55$, $F(17, 32) = 2.26$, $p = .023$, partial $\eta^2 = .55$, Obs. Power = .92. Separate univariate ANOVAs on the outcome variables revealed that this effect could be attributed to a significant regional difference in the mean ratings of the transfer design subscale, $F(1, 48) = 6.02$, $p = .018$, partial $\eta^2 = .11$, Obs. Power = .67, such that it was rated significantly higher in the Pacific region compared to Ontario.

Moderation Effects

In order to determine if the relationships between the various training outcomes were moderated by the trainee effectiveness variables, hierarchical multiple regression was used. Several assumptions must be met for this type of analysis to be appropriate. As mentioned above, one of the main assumptions of multiple regression analysis is that the potential moderators and predictor variables must demonstrate high reliability scores (Frazier et al., 2004). Given this, the training effectiveness variables with reliability scores $> .80$ were targeted for these analyses. The five training effectiveness variables included in the analysis are described in Table 34.

Table 34

Training Effectiveness Variables Included in the MMR Analyses

LTSI Subscale	Description
Supervisor Support	The extent to which supervisors/managers support and reinforce the use of training on the job.
Perceived Content Validity	The extent to which trainee's judge training content to reflect job requirements accurately.
Resistance-Openness to Change	The extent to which prevailing group norms are perceived by individuals to resist or discourage the use of skills and knowledge acquired in training.
Performance Self-Efficacy	An individual's general belief that he is able to change his performance when he wants to.
Transfer Design	The degree to which (1) training has been designed and delivered to give trainees the ability to transfer learning to the job, and (2) training instructions match job requirements.

The other assumptions that must be met for MMR to be appropriate include the standard statistical assumptions of the absence of outliers, multicollinearity and singularity as well as the normality, linearity and homoscedasticity of residuals. Standardized residuals were examined to ensure that regression assumptions were not violated (Tabachnick & Fidell, 2007). Univariate outliers (i.e., standard scores greater than ± 3.29) were identified for two variables (ATCW at the training phase and the Transfer Design Subscale), but no multivariate outliers were present in the data. Residual scatterplots (standardized predicted values by standardized residuals) were examined for deviations from normality, linearity, and homoscedasticity. Plots did not display any curvilinear patterns, and no skewed or uneven error distributions; however, significance

tests revealed that several standardized variables were significantly skewed and kurtotic. The ATCW at the training phase was severely negatively skewed ($-5.34, p < .001$) and significantly leptokurtic ($5.41, p < .001$). The hard skills score at the training phase was also severely negatively skewed ($-3.77, p < .001$). Lastly, the transfer design subscale was moderately negatively skewed ($-2.76, p < .001$) and leptokurtic ($3.35, p < .01$). As indicated above in the data screening section of Chapter 3, no differences in results were found when using untransformed or transformed data, thus the data was analyzed untransformed. Neither multicollinearity nor singularity was an issue with the variables employed in the current hierarchical regression analyses (i.e., all r 's between variables were $< .80$; Tabachnick & Fidell, 2007). Finally, scatterplots of cases by residuals and cases by each outcome variable were examined, indicating no visible patterns across errors or outcomes for any cases. Similarly, there is no reason to suspect that there are significant relationships between the trainees (i.e., independence of cases) or that errors across trainees are correlated (i.e., independence of errors).

The predictor and moderator variables in the analyses described below were all considered to be continuous variables. The predictor variables included the knowledge and skills tests scores as well as attitude ratings from various study phases. The moderator variables included the five subscales of interest from the LTSI. Although the scales are technically considered interval measures, they are treated as continuous in the current study. The outcome variables, similar to the predictor variables, included the knowledge and skills tests scores as well as attitude ratings from the latter two study phases. More specific details are provided below for each of the analyses of the moderator effects.

It is recommended that all of the continuous predictor and moderator variables be centered or standardized in order to reduce any chances of multicollinearity in the conditional or moderator effects, as well as to ease the interpretation of results (Fairchild & MacKinnon, 2009; Frazier et al., 2004). In the current study, these variables were standardized by converting each into a z score. Following this standardization procedure, the standardized scores of both the predictor and moderator variables were multiplied together to create the interaction terms.

The results presented below are categorized by the moderator variables (i.e., the LTSI subscales of interest). For each potential moderator, the expected findings are outlined along with the regression results; however, as indicated above, the general hypothesis for each analysis is that higher levels of each moderator will strengthen the relationship between the predictor variables and the outcome variables.³¹

Supervisor support. This section presents the results of the regression analyses performed to test whether self-reported ratings of supervisor support moderate various relationships between Level 2 training outcomes (between training and transfer), as well as relationships between Level 2 and 3 outcomes (between training and transfer). The Level 2 outcomes of interest included soft skill scores, hard skill scores, and attitude ratings (towards corrections and inmates). The Level 3 outcomes of interest included the quality component of the trainee behavioural checklist. It is expected that the relationships between the predictor and outcome measures will be significantly stronger when trainees perceive that their supervisor is supportive (i.e., when the supervisor reinforces the use of training on the job). For example, the relationship between soft skills

³¹ According to Frazier and colleagues (2004), this would be considered an enhancing interaction, which is defined by the authors as a situation where “both the predictor and moderator affect the outcome variable in the same direction and together have a stronger than additive effect” (p.117).

scores between training and transfer will be strengthened under conditions where trainees perceive high levels of supervisor support. See Table 35 below for the variables included in the proposed hierarchical regression analyses.

Table 35

Proposed Hierarchical Regression Analyses to Examine the Potential Moderating Effect of Supervisor Support

Relationship	Predictor Variable	Moderator	Outcome Variable
A	Soft Skill Assessment (Training Phase)	Supervisor Support	Soft Skill Assessment (Transfer Phase)
B	Soft Skill Assessment (Training Phase)	Supervisor Support	Quality of Skill Use (TBC)
C	Hard Skills Test Score (Training Phase)	Supervisor Support	Hard Skills Test Score (Transfer Phase)
D	Hard Skills Test Score (Training Phase)	Supervisor Support	Quality of Skill Use (TBC)
E	ATCW (Training Phase)	Supervisor Support	ATCW (Transfer Phase)
F	ATI (Training Phase)	Supervisor Support	ATI (Transfer Phase)

In order to determine whether the interaction term accounts for a significant amount of variance in the outcome variable over and above the effects of the predictor and the moderator variable themselves, the predictor and moderator variables are entered into Block 1, followed by the interaction term (predictor x moderator) in Block 2. In the case of relationship A, in order to determine whether the interaction term accounted for a significant amount of variance in soft skill retention above and beyond that accounted for

by the individual variables, soft skill scores at the training phase and supervisor support were entered into the model first (Block 1), followed by the interaction term (Block 2).

In contrast to expectations, the results showed that the moderator did not significantly impact relationship A, and thus did not account for significantly more variance in soft skill retention than each of the conditional effects alone (see Table 36 below for the R^2 change). As also indicated in the t -tests results of Block 2 (see Table 37 below for the beta coefficients), the interaction regression coefficient was not significant. Given that the moderated effect was not significant, the first-order effects were interpreted without the interaction term included in the model (i.e., as represented in Block 1 in Table 37) (Jaccard & Turrisi, 2003), and neither one of the first order effects was significant.

The same procedure as that outlined for relationship A was used to examine the rest of the relationships (B-F) (refer to Table 37 to see the variables entered into Block 1 and 2). In contrast to expectations, the results showed that the potential moderator effect of supervisor support was not significant for any of the relationships tested (see Table 36 for the R^2 change values and Table 37 for the B coefficients). Given that the moderator effect was not significant, the first-order effects in Table 37 were interpreted without the interaction term included in the model (i.e., as represented in Block 1 for each relationship in Table 37). These results are presented next for relationships B through F.

Table 36

Summary of Hierarchical Multiple Regression Analyses for the Moderator of Supervisor Support

Relationship	Block/Model	R	R^2	$R^2\Delta$	$F\Delta$	p
A	1	.331	.109	.109	3.07	.055
	2	.337	.114	.004	.23	.631
B	1	.336	.113	.113	3.19	.050
	2	.353	.125	.012	.67	.418
C	1	.400	.160	.160	4.75	.013
	2	.403	.163	.003	.17	.678
D	1	.407	.166	.166	4.98	.011
	2	.407	.166	.000	.00	.998
E	1	.525	.276	.276	9.53	.000
	2	.528	.279	.003	.18	.673
F	1	.564	.318	.318	11.66	.000
	2	.564	.319	.001	.04	.839

Note. Δ = change. The degrees of freedom for the F statistics were the same for all relationships; Block 1 = 2, 50 and Block 2 = 1, 49.

Table 37

Regression Coefficients for Each Effect and Their Associated Confidence Intervals and T-Tests – Supervisor Support

Relationship	Block and Predictor	B(SE)	95% CI	β	t	p	Power O.E.	Power M.E.
A	Block 1							
	Soft Skills (Training)	.867(.443)	-.022 to 1.757	.274	1.96	.056		
	Supervisor Support	-.378(.439)	-1.260 to .503	-.121	-.86	.393		
	Block 2							
	Soft Skills (Training) × Supervisor Support	.212(.439)	-.671 to 1.095	.065	.48	.631	.68	.06
B	Block 1							
	Soft Skills (Training)	.001(.049)	-.097 to .100	.004	.03	.976		
	Supervisor Support	.117(.048)	.020 to .214	.337	2.42	.019		
	Block 2							
	Soft Skills (Training) × Supervisor Support	.039(.048)	-.058 to .136	.110	.817	.418	.72	.12
C	Block 1							
	Hard Skills (Training)	2.693(1.038)	.607 to 4.778	.336	2.59	.012		
	Supervisor Support	-1.741(1.068)	-3.885 to .404	-.211	-1.63	.109		
	Block 2							
	Hard Skills (Training) × Supervisor Support	-.442(1.059)	-2.569 to 1.686	-.056	-.42	.678	.83	.06
D	Block 1							
	Hard Skills (Training)	-.078(.044)	-.165 to .010	-.230	-1.78	.081		
	Supervisor Support	.115(.045)	.025 to .205	.333	2.58	.013		

Relationship	Block and Predictor	<i>B</i> (SE)	95% <i>CI</i>	β	<i>t</i>	<i>p</i>	Power O.E.	Power M.E.
E	Block 2							
	Hard Skills (Training) × Supervisor Support	.001(.044)	-.089 to .090	.002	.02	.988	.84	.05
	Block 1							
	ATCW(Training)	.559(.192)	.173 to .944	.355	2.91	.005		
	Supervisor Support	.574(.209)	.155 to .993	.335	2.74	.008		
F	Block 2							
	ATCW(Training) × Supervisor Support	-.067(.159)	-.387 to .252	-.054	-.43	.673	.97	.06
	Block 1							
	ATI(Training)	7.145(1.779)	3.573 to 10.717	.475	4.02	.000		
	Supervisor Support	3.477(1.727)	.009 to 6.946	.238	2.01	.049		
	Block 2							
	ATI(Training) × Supervisor Support	.327(1.603)	-2.895 to 3.548	.024	.20	.839	.98	.05

Note. Power O.E. stands for the power of the overall effect and Power M.E stands for the power for the moderating effect.

For relationship B, an examination of the *B* coefficients indicates that soft skill scores at the training phase did not significantly contribute to an increase in quality of skill use on the job while holding supervisor support ratings constant. In contrast, holding soft skills scores at the training phase constant, quality of skill use on the job is predicted to increase by .117 units as ratings of supervisor support increase by one point.

For relationship C, while holding perceived supervisor support constant, hard skill scores at the transfer phase are predicted to increase by 2.693 units as hard skill scores at the training phase increase by one point. In contrast, holding hard skills scores at the training phase constant, supervisor support ratings did not contribute to a significant increase in hard skill scores at the transfer phase.

For relationship D, hard skill scores at the training phase did not contribute to a significant increase in quality of skill use on the job while holding supervisor support ratings constant. However, quality of skill use on the job is predicted to increase by .115 units as perceived supervisor support ratings increase by one point, while holding hard skills scores at the training phase constant.

For relationship E, while holding supervisor support constant, ATCW ratings at the transfer phase are predicted to increase by .559 units as ATCW ratings at the training phase increase by one point. Similarly, while holding ATCW ratings at the training phase constant, ATCW ratings at the transfer phase is predicted to increase by .574 units as perceived supervisor support ratings increase by one point. The standardized slopes can be compared to assess the relative importance of each factor in terms of explaining the variation in ATCW ratings at the transfer phase. As shown in Table 37 (Block 1, relationship E), the standardized slopes for ATCW ratings at the training phase and

perceived supervisor support ratings are comparable, indicating that both variables are approximately equally effective in predicting ATCW ratings on the job.

Finally, for relationship F, while holding supervisor support constant, ATI ratings at the transfer phase are predicted to increase by 7.145 units as ATI ratings at the training phase increase by one point. Similarly, while holding ATI ratings at the training phase constant, ATI ratings at the transfer phase are predicted to increase by 3.477 units as perceived supervisor support ratings increase by one point. The standardized slopes were compared to assess the relative importance of each factor in terms of explaining the variation in ATI ratings at the transfer phase. As shown in Table 37 (Block 1, relationship F), the standardized slope for ATI ratings at the training phase was higher than that for the perceived supervisor support ratings, indicating that the ATI ratings at the training phase are approximately two times more effective than supervisor support in predicting ATI ratings on the job.

Resistance-Openness to Change. This section presents the results of the regression analyses performed to test whether self-reported ratings of openness to change moderates various relationships between Level 2 training outcomes at the training and transfer phases, as well as relationships between Level 2 and 3 outcomes at the training and transfer phases. Higher scores on this scale indicate stronger feelings that the organizational culture encourages the use of skills and knowledge acquired in training. The Level 2 training outcomes of interest included soft skills scores and hard skills scores. The Level 3 outcome of interest included the quality component of the trainee behavioural checklist.

It is expected that the relationships between the predictor and outcome measures will be significantly stronger when perceived group norms encourage the use of skills and knowledge acquired during training. For example, the relationship between soft skills scores between training and transfer is expected to be stronger when trainees perceive higher levels of openness to change. See Table 38 for the variables included in the proposed hierarchical regression analyses.

Table 38

Proposed Hierarchical Regression Analyses for the Moderation Effect of Openness to Change

Relationship	Predictor Variable	Moderator	Outcome Variable
A	Soft Skill Assessment (Training Phase)	Openness to Change	Soft Skill Assessment (Transfer Phase)
B	Soft Skill Assessment (Training Phase)	Openness to Change	Quality of Skill Use (TBC)
C	Hard Skills Test Score (Training Phase)	Openness to Change	Hard Skills Test Score (Transfer Phase)
D	Hard Skills Test Score (Training Phase)	Openness to Change	Quality of Skill Use (TBC)

The same procedure as outlined above was used for the relationships tested here (A-D) for the potential moderator of openness to change (refer to Table 40 to see the variables entered into Block 1 and 2). In contrast to expectations, the results showed that the potential moderator effect of openness to change was not significant for any of the relationships tested (see Table 39 for the R^2 change values and Table 40 for the B coefficients). Given that the moderator effect was not significant, the first-order effects in

Table 40 were interpreted without the interaction term included in the model (i.e., as represented in Block 1 for each relationship in Table 40). These results are presented next for relationships A through D.

Table 39

Summary of Hierarchical Multiple Regression Analyses for the Moderator of Resistance-Openness to Change

Relationship	Block/Model	R	R^2	$R^2\Delta$	$F\Delta$	p
A	1	.384	.148	.148	4.33	.018
	2	.423	.179	.032	1.88	.176
B	1	.321	.103	.103	2.88	.066
	2	.325	.106	.003	.14	.706
C	1	.375	.140	.140	4.08	.023
	2	.405	.164	.023	1.38	.247
D	1	.412	.170	.170	5.12	.009
	2	.448	.201	.031	1.90	.174

Note. Δ = change. The degrees of freedom for the F statistics were the same for all relationships; Block 1 = 2, 50 and Block 2 = 1, 49.

Table 40

Regression Coefficients for Each Effect and Their Associated Confidence Intervals and T-Tests – Resistance-Openness to Change

Relationship	Block and Predictor	B(SE)	95% CI	β	<i>t</i>	<i>p</i>	Power O.E.	Power M.E.
A	Block 1							
	Soft Skills (Training)	.777(.430)	-.087 to 1.640	.245	1.81	.077		
	Openness to Change	-.740(.426)	-1.595 to .115	-.236	-1.74	.088		
	Block 2							
	Soft Skills (Training) × Openness to Change	.569(.415)	-.264 to 1.403	.178	1.37	.176	.86	.25
B	Block 1							
	Soft Skills (Training)	-.003(.049)	-.101 to .095	-.009	-.07	.946		
	Openness to Change	.111(.048)	.013 to .208	.319	2.29	.026		
	Block 2							
	Soft Skills (Training) × Open to Change	.018(.048)	-.078 to .114	.051	.38	.706	.65	.06
C	Block 1							
	Hard Skills (Training)	2.807(1.053)	.693 to 4.922	.350	2.67	.010		
	Openness to Change	-1.313(1.083)	-3.487 to .862	-.159	-1.21	.231		
	Block 2							
	Hard Skills (Training) × Openness to Change	-1.031(.879)	-2.797 to .735	-.175	-1.17	.247	.83	.19
D	Block 1							
	Hard Skills (Training)	-.087(.044)	-.175 to .000	-.259	-2.01	.050		

Relationship	Block and Predictor	<i>B</i> (SE)	95% <i>CI</i>	β	<i>t</i>	<i>p</i>	Power O.E.	Power M.E.
	Openness to Change	.118(.045)	.028 to .208	.339	2.62	.011		
	Block 2							
	Hard Skills (Training) \times Openness to Change	.050(.036)	-.023 to .123	.201	1.38	.174	.90	.24

Note. Power O.E. stands for the power of the overall effect and Power M.E stands for the power for the moderating effect.

For relationship A, while holding perceived openness to change constant, soft skill scores at the transfer phase are predicted to increase by 2.807 units as soft skill scores at the training phase increase by one point. In contrast, while holding perceived soft skills scores at the training phase constant, openness to change ratings did not significantly contribute to an increase in soft skills scores at the transfer phase.

For relationship B, while holding ratings of openness to change constant, quality of skill use on the job is predicted to decrease by .087 units as soft skills scores increase by one point. Similarly, while holding soft skills scores at the training phase constant, quality of skill use on the job is predicted to increase by .118 units as openness to change is rated one point higher. The standardized slopes were compared to assess the relative importance of each factor in terms of explaining the variation in quality of skill use at the transfer phase. As shown in Table 40 (Block 1, relationship E), the standardized slope for soft skill scores at the training phase was slightly lower and in a negative direction compared to that for the perceived openness to change ratings ($\beta = -.259$ and $\beta = .339$, respectively), indicating that the soft skill scores at the training phase may be less effective than the ratings of openness to change in predicting quality of skill use on the job.

For relationship C, neither one of the first order effects was significant. For relationship D, while holding openness to change constant, hard skill scores at the training phase did not significantly contribute to increases in quality of skill use on the job. In contrast, while holding hard skills scores at the training phase constant, quality of skill use on the job is predicted to increase by .111 units as perceived openness to change ratings increase by one point.

Content Validity. This section presents the results of the regression analyses performed to test the ability of self-reported ratings of content validity to moderate various relationships between Level 2 training outcomes at the training and transfer phases, as well as relationships between Level 2 and 3 outcomes at the training and transfer phases. The Level 2 training outcomes of interest included the knowledge, soft skills and hard skills test scores. The Level 3 outcome of interest included the quality component of the trainee behavioural checklist. It is expected that the relationships between the predictor and outcome measures will be significantly stronger when trainees perceive greater levels of content validity (i.e., greater perceived connection between training content and what is required on the job). For example, the relationship between knowledge test scores between training and transfer are expected to be stronger when perceived content validity is high. See Table 41 for the variables included in the proposed hierarchical regression analyses.

Table 41

Proposed Hierarchical Regression Analyses for the Moderation Effect of Content Validity

Relationship	Predictor Variable	Moderator	Outcome Variable
A	Knowledge Test (Recruitment Phase)	Content Validity	Knowledge Test (Training Phase)
B	Knowledge Test (Training Phase)	Content Validity	Knowledge Test (Transfer Phase)
C	Soft Skill Assessment (Recruitment Phase)	Content Validity	Soft Skill Assessment (Training Phase)
D	Soft Skill Assessment (Training Phase)	Content Validity	Soft Skill Assessment (Transfer Phase)
E	Soft Skill Assessment (Training Phase)	Content Validity	Quality of Skill Use (TBC)
F	Hard Skills Test Score (Training Phase)	Content Validity	Hard Skills Test Score (Transfer Phase)
G	Hard Skills Test Score (Training Phase)	Content Validity	Quality of Skill Use (TBC)

The same procedure as outlined above was used for the relationships tested here (A-G) (refer to Table 43 to see the variables entered into Block 1 and 2). In contrast to expectations, the results showed that the potential moderator effect of content validity was not significant for the relationships that were tested, with the exception of relationship B (see Table 42 for the R^2 change values and Table 43 for the B coefficients). Given that the moderator effect was not significant for most relationships, the first-order effects were interpreted without the interaction term included in the model (i.e., as represented in Block 1 for each relationship in Table 43). In the case of relationship B,

however, the moderator effects are presented and interpreted. These results are presented below.

Table 42

Summary of Hierarchical Multiple Regression Analyses for the Moderator of Content

Validity

Relationship	Block/Model	R	R^2	$R^2\Delta$	$F\Delta$	p
A	1	.281	.079	.079	2.15	.127
	2	.282	.080	.000	.02	.896
B	1	.549	.301	.301	10.76	.000
	2	.651	.424	.123	10.44	.002
C	1	.197	.039	.039	1.01	.373
	2	.262	.068	.030	1.56	.217
D	1	.350	.122	.122	3.49	.038
	2	.355	.126	.004	.20	.656
E	1	.285	.081	.081	2.22	.120
	2	.311	.097	.016	.84	.364
F	1	.346	.120	.120	3.41	.041
	2	.348	.121	.001	.06	.815
G	1	.383	.147	.147	4.30	.019
	2	.383	.147	.000	.01	.927

Note. Δ = change. The degrees of freedom for the F statistics were the same for all relationships; Block 1 = 2, 50 and Block 2 = 1, 49.

Table 43

Regression Coefficients for Each Effect and Their Associated Confidence Intervals and T-Tests – Content Validity

Relationship	Block and Predictor	B(SE)	95% CI	β	<i>t</i>	<i>p</i>	Power O.E.	Power M.E.
A	Block 1							
	Knowledge(Recruitment)	1.118(.568)	-.022 to 2.258	.268	1.97	.054		
	Content Validity	-.418(.547)	-1.517 to .681	-.104	-.76	.448		
	Block 2							
	Knowledge (Recruitment) × Content Validity	-.116(.883)	-1.891 to 1.659	-.021	-.13	.896	.53	.05
B	Block 1							
	Knowledge(Training)	3.059(.663)	1.727 to 4.392	.547	4.61	.000		
	Content Validity	.654(.718)	-.788 to 2.096	.108	.91	.367		
	Block 2							
	Knowledge (Training) × Content Validity	2.282(.706)	.863 to 3.701	.370	3.23	.002	.97	.72
C	Block 1							
	Soft Skills (Recruitment)	.622(.561)	-.504 to 1.749	.155	1.11	.273		
	Content Validity	.425(.550)	-.679 to 1.529	.108	.77	.443		
	Block 2							
	Soft Skills (Training) × Content Validity	.669(.535)	-.407 to 1.744	.174	1.25	.217	.47	.24
D	Block 1							
	Soft Skills (Training)	1.045(.423)	.197 to 1.894	.330	2.47	.017		

Relationship	Block and Predictor	<i>B</i> (<i>SE</i>)	95% <i>CI</i>	β	<i>t</i>	<i>p</i>	Power O.E.	Power M.E.
	Content Validity	-.511(.419)	-1.352 to .330	-.163	-1.22	.228		
E	Block 2 Soft Skills (Training) × Content Validity	.177(.396)	-.619 to .974	.062	.45	.656	.73	.06
	Block 1 Soft Skills (Training)	-.046(.048)	-.142 to .050	-.130	-.95	.345		
	Content Validity	.094(.047)	-.001 to .189	.270	1.98	.053		
	Block 2 Soft Skills (Training) × Content Validity	-.041(.045)	-.130 to .049	-.130	-.92	.364	.61	.14
F	Block 1 Hard Skills (Training)	2.821(1.081)	.651 to 4.992	.352	2.61	.012		
	Content Validity	-.585(1.111)	-2.818 to 1.647	-.071	-.53	.601		
	Block 2 Hard Skills (Training) × Content Validity	.249(1.057)	-1.875 to 2.373	.032	.24	.815	.71	.05
G	Block 1 Hard Skills (Training)	-.098(.045)	-.188 to -.008	-.291	-2.19	.033		
	Content Validity	.107(.046)	.014 to .199	.308	2.32	.025		
	Block 2 Hard Skills (Training) × Content Validity	-.004(.044)	-.092 to .084	-.012	-.09	.927	.79	.05

Note. Power O.E. stands for the power of the overall effect and Power M.E stands for the power for the moderating effect.

For relationship A, none of the conditional effects were significant. However, in line with our expectations, the results showed that the moderated effect was significant for relationship B (i.e., the interaction between knowledge test scores at the training phase and content validity accounted for significantly more variance in knowledge retention than each of the conditional effects alone; see Table 43, relationship B). As also indicated in the *t*-tests results of Block 2 (see Table 42, relationship B), the interaction regression coefficient was significantly different from zero. Since the interaction term/moderated effect was significant, a simple slopes analysis was conducted to further investigate the expected moderated effect. The simple slopes web utility provided by Preacher, Curran and Bauer (2012) was used for this purpose. The simple slopes analysis is plotted in Figure 6.

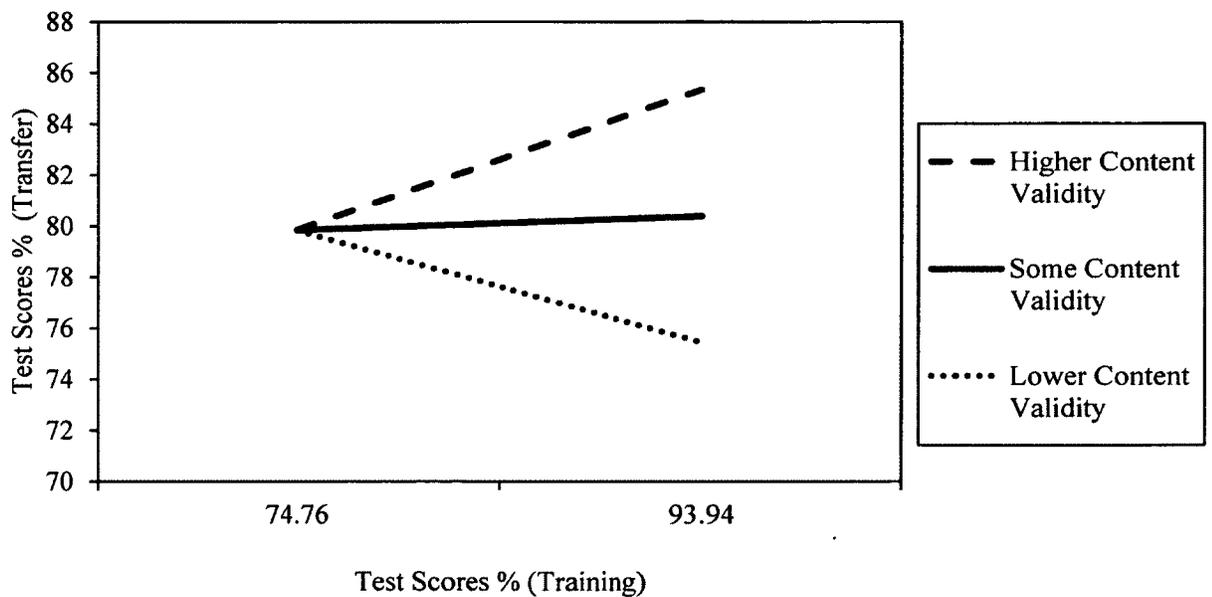


Figure 6. Results of the simple slopes analysis, displaying the relationship between knowledge test scores at the training and transfer phases as a function of content validity ratings.

As shown in Table 44, the simple slope of the regression line for knowledge test scores at the training phase on knowledge tests scores at the transfer phase was significant for higher ratings of content validity. Neither one of the other simple slopes were significantly different from zero (i.e., for the mean or lower ratings of content validity). Therefore, the significant moderation effect was found only for higher levels of content validity.

Table 44

Summary of the Results of a Simple Slopes Analysis Showing the Relationship Between Knowledge Test Scores at the Training and Transfer Phases as a Function of Content Validity Ratings

Moderator Level	<i>B</i> (SE)	<i>t</i>	<i>p</i>
Higher Content Validity Ratings	2.536(1.096)	2.31	.025
Some Content Validity Ratings	.254(.819)	.31	.758
Lower Content Validity Ratings	-2.028(1.245)	-1.63	.110

For relationship C, none of the conditional effects were significant. However, for relationship D, while holding content validity ratings constant, soft skill scores at the transfer phase are predicted to increase by 1.045 units as soft skills scores at the training phase increase by one point. In contrast, while holding soft skills scores at the training phase constant, ratings of content validity did not contribute significantly to an increase in soft skills scores at the transfer phase.

For relationship E, neither of the conditional effects were significant ($p > .05$). For relationship F, while holding content validity ratings constant, hard skill scores at the

transfer phase are predicted to increase by 2.821 units as hard skills scores at the training phase increase by one point. In contrast, while holding hard skills scores constant at the training phase, content validity ratings did not contribute to a significant increase in hard skill scores in the transfer phase.

For relationship G, while holding ratings of content validity constant, quality of skill use on the job is actually predicted to decrease by .098 units as hard skills scores at the training phase increase by one point. On the other hand, while holding hard skills scores at the training phase constant, quality of skill use on the job is predicted to increase by .107 units as content validity is rated one point higher. The standardized slopes were compared to assess the relative importance of each factor in terms of explaining the variation in quality of skill use at the transfer phase. As shown in Table 43 (Block 1, relationship G), the standardized slopes for hard skill scores at the training phase was slightly lower and in a negative direction compared to that for the content validity ratings ($\beta = -.291$ and $\beta = .308$, respectively), indicating that both the hard skill scores at the training phase and the ratings of content validity appear to be equally effective in predicting quality of skill use on the job, although in opposite directions.

Self-efficacy. This section presents the results of the regression analyses performed to test the ability of self-reported ratings of self-efficacy to moderate various relationships between Level 2 training outcomes at the training and transfer phases, as well as relationships between Level 2 and 3 outcomes at the training and transfer phases. The Level 2 training outcomes of interest included the knowledge, soft skills and hard skills test scores. The Level 3 outcome of interest included the quality component of the trainee behavioural checklist. It is expected that the relationships between the predictor

and outcome measures will be significantly stronger when trainees perceive greater levels of self-efficacy (i.e., greater belief that an individual is able to change their behaviour when they want to). For example, the relationship between knowledge test scores between training and transfer is expected to be strengthened under conditions when trainees perceive higher levels of self-efficacy. See Table 45 below for the variables included in the proposed hierarchical regression analyses.

Table 45

Proposed Hierarchical Regression Analyses for the Moderation Effect of Self-Efficacy

Relationship	Predictor Variable	Moderator	Dependent Variable
A	Knowledge Test (Recruitment Phase)	Self-efficacy	Knowledge Test (Training Phase)
B	Knowledge Test (Training Phase)	Self-efficacy	Knowledge Test (Transfer Phase)
C	Soft Skill Assessment (Recruitment Phase)	Self-efficacy	Soft Skill Assessment (Training Phase)
D	Soft Skill Assessment (Training Phase)	Self-efficacy	Soft Skill Assessment (Transfer Phase)
E	Soft Skill Assessment (Training Phase)	Self-efficacy	Quality of Skill Use (TBC)
F	Hard Skills Test Score (Training Phase)	Self-efficacy	Hard Skills Test Score (Transfer Phase)
G	Hard Skills Test Score (Training Phase)	Self-efficacy	Quality of Skill Use (TBC)

The same procedure outlined above was used for the relationships tested here (A-G) (refer to Table 47 to see the variables entered into Block 1 and 2). In contrast to

expectations, the results showed that the potential moderator effect of self-efficacy was not significant for the relationships tested here, with the exception of relationship B (see Table 46 for the R^2 change values and Table 47 for the B coefficients). Given that the moderator effect was not significant for most relationships, the first-order effects were interpreted without the interaction term included in the model (i.e., as represented in Block 1 for each relationship in Table 47). However, for relationship B, the moderator effects are presented and interpreted. These results are presented below.

Table 46

Summary of Hierarchical Multiple Regression Analyses for the Moderator of Self-Efficacy

Relationship	Block/Model	R	R^2	$R^2\Delta$	$F\Delta$	p
A	1	.334	.112	.112	3.14	.052
	2	.334	.112	.000	.02	.899
B	1	.540	.291	.291	10.26	.000
	2	.594	.353	.062	4.67	.036
C	1	.204	.042	.042	1.08	.346
	2	.210	.044	.002	.13	.722
D	1	.311	.096	.096	2.67	.079
	2	.344	.118	.022	1.22	.275
E	1	.165	.027	.027	.70	.503
	2	.188	.035	.008	.41	.525
F	1	.341	.116	.116	3.28	.046
	2	.355	.126	.010	.57	.452
G	1	.301	.091	.091	2.50	.092
	2	.323	.104	.013	.73	.399

Note. Δ = change. The degrees of freedom for the F statistics were the same for all relationships; Block 1 = 2, 50 and Block 2 = 1, 49.

Table 47

Regression Coefficients for Each Effect and Their Associated Confidence Intervals and T-Tests – Self-Efficacy

Relationship	Block and Predictor	B(SE)	95% CI	β	<i>t</i>	<i>p</i>	Power O.E.	Power M.E.
A	Block 1							
	Knowledge (Recruitment)	1.044(.557)	-.075 to 2.164	.250	1.87	.067		
	Self-Efficacy	-.837(.537)	-1.917 to .242	-.208	-1.56	.126		
	Block 2							
	Knowledge (Recruitment) × Self-Efficacy	-.104(.815)	-1.741 to 1.534	-.021	-.13	.899	.68	.05
B	Block 1							
	Knowledge (Training)	3.059(.682)	1.688 to 4.430	.547	4.48	.000		
	Self-Efficacy	.257(.739)	-1.227 to 1.740	.042	.35	.729		
	Block 2							
	Knowledge (Training) × Self-Efficacy	1.577(.729)	.111 to 3.043	.262	2.16	.036	.99	.44
C	Block 1							
	Soft Skills (Recruitment)	.716(.561)	-.410 to 1.842	.178	1.28	.208		
	Self-Efficacy	-.475(.550)	-1.579 to .629	-.120	-.86	.392		
	Block 2							
	Soft Skills (Training) × Self-Efficacy	-.226(.633)	-1.497 to 1.045	-.051	-.36	.722	.33	.05
D	Block 1							
	Soft Skills (Training)	.977(.428)	.118 to 1.836	.309	2.29	.027		

Relationship	Block and Predictor	<i>B</i> (SE)	95% <i>CI</i>	β	<i>t</i>	<i>p</i>	Power O.E.	Power M.E.
	Self-Efficacy	-.044(.424)	-.896 to .807	-.014	-.11	.917		
	Block 2							
	Soft Skills (Training) \times Self-Efficacy	-.479(.434)	-1.350 to .393	-.150	-1.10	.275	.70	.18
E	Block 1							
	Soft Skills (Training)	-.029(.049)	-.128 to .069	-.084	-.60	.554		
	Self-Efficacy	.046(.049)	-.051 to .144	.134	.95	.345		
	Block 2							
	Soft Skills (Training) \times Self-Efficacy	-.032(.050)	-.133 to .069	-.091	-.64	.525	.27	.09
F	Block 1							
	Hard Skills (Training)	2.671(1.083)	.496 to 4.847	.334	2.47	.017		
	Self-Efficacy	.259(1.114)	-1.979 to 2.496	.031	.23	.817		
	Block 2							
	Hard Skills (Training) \times Self-Efficacy	-.665(.877)	-2.428 to 1.098	-.110	-.76	.452	.73	.10
G	Block 1							
	Hard Skills (Training)	-.091(.046)	-.184 to .002	-.270	-1.97	.054		
	Self-Efficacy	.066(.048)	-.029 to .162	.192	1.40	.169		
	Block 2							
	Hard Skills (Training) \times Self-Efficacy	-.032(.037)	-.107 to .043	-.126	-.85	.399	.64	.12

Note. Power O.E. stands for the power of the overall effect and Power M.E stands for the power for the moderating effect.

For relationship A, none of the conditional effects were significant. However, in line with expectations, the results showed that the moderated effect was significant for relationship B (i.e., the interaction between knowledge test scores at the training phase and self-efficacy accounted for significantly more variance in knowledge retention than each of the conditional effects alone; see Table 47, relationship B). As also indicated in the *t*-tests results of Block 2 (see Table 47 relationship B), the interaction regression coefficient was significantly different from zero. Since the interaction term/moderated effect was significant, a simple slopes analysis was conducted to further investigate the expected moderated effect. Again, the simple slopes web utility provided by Preacher and colleagues (2012) was used for this purpose. The simple slopes analysis is plotted in Figure 7.

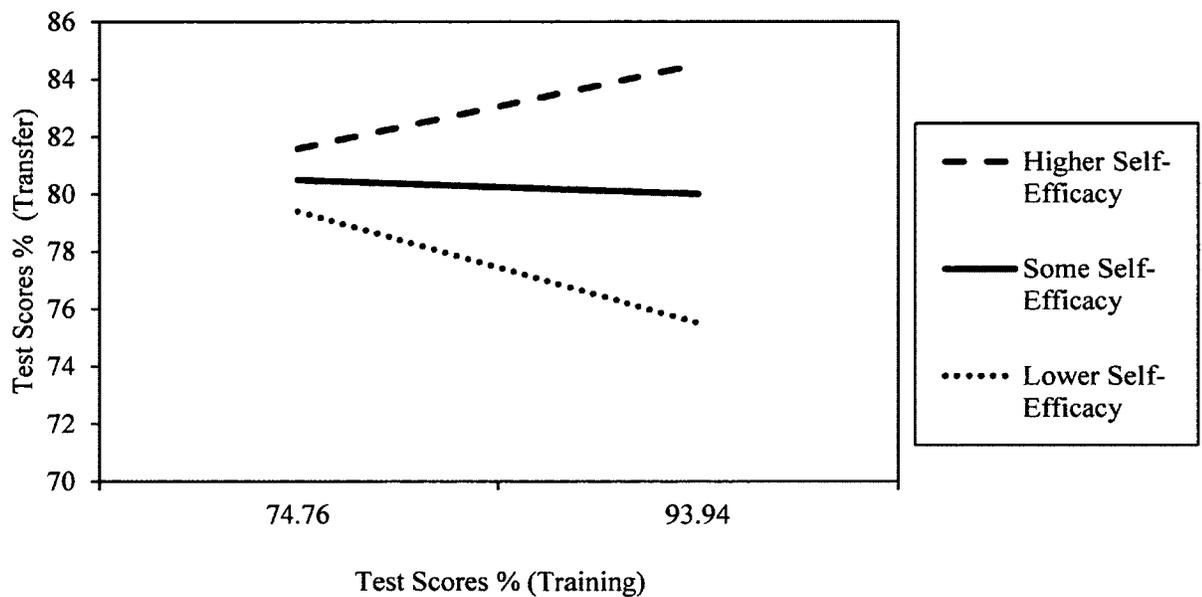


Figure 7. Results of the simple slopes analysis, displaying the relationship between knowledge test scores at the training and transfer phases as a function of self-efficacy ratings.

As shown in Table 48, none of the simple slopes were significantly different from zero. Thus, while the expected pattern of results did emerge, the slope of the regression line for knowledge tests scores at the transfer phase on knowledge test scores at the training phase did not significantly vary as a function of higher or lower ratings of perceived self-efficacy.

Table 48

Summary of the Results of a Simple Slopes Analysis Showing the Relationship Between Knowledge Test Scores at the Training and Transfer Phases as a Function of Self-Efficacy Ratings

Moderator Level	<i>B</i> (SE)	<i>t</i>	<i>p</i>
Higher Self-Efficacy Ratings	1.355(1.073)	1.26	.212
Some Self-Efficacy Ratings	-.222(.864)	-0.26	.798
Lower Self-Efficacy Ratings	-1.799(1.341)	-1.341	.186

For relationship C, none of the conditional effects were significant. However, for relationship D, while holding self-efficacy ratings constant, soft skill scores at the transfer phase are predicted to increase by .997 units as soft skills scores at the training phase increase by one point. In contrast, while holding soft skills scores at the training phase constant, self-efficacy ratings did not contribute to a significant increase in soft skills scores at the transfer phase.

For relationship E, none of the conditional effects were significant. For relationship F, while holding self-efficacy ratings constant, hard skill scores at the transfer phase are predicted to increase by 2.671 units as hard skills scores at the training phase increase by one point. In contrast, while holding hard skills scores constant at the

training phase, self-efficacy ratings did not contribute to a significant increase in hard skills test scores at the transfer phase. For relationship G, as shown in Table 47, neither of the conditional effects were significant ($p > .05$).

Transfer Design. This section presents the results of the regression analyses performed to test the ability of self-reported ratings of transfer design to moderate various relationships between Level 2 training outcomes at the training and transfer phases, as well as relationships between Level 2 and 3 outcomes at the training and transfer phases. The Level 2 training outcomes of interest included soft skills scores and hard skills scores. The Level 3 outcome of interest included the quality component of the trainee behavioural checklist. It is expected that the relationships between the predictor and outcome measures will be significantly stronger when trainees perceive greater levels of transfer design (i.e., greater perception that training has been designed with transfer in mind). For example, the relationship between soft skill scores between training and transfer are expected to be strengthened under condition when trainees perceive higher levels of transfer design. See Table 49 below for the variables included in the proposed hierarchical regression analyses.

Table 49

Proposed Hierarchical Regression Analyses for the Moderation Effect of Transfer Design

Relationship	Predictor Variable	Moderator	Dependent Variable
A	Soft Skill Assessment (Training Phase)	Transfer Design	Soft Skill Assessment (Transfer Phase)
B	Soft Skill Assessment (Training Phase)	Transfer Design	Quality of Skill Use (TBC)
C	Hard Skills Test Score (Training Phase)	Transfer Design	Hard Skills Test Score (Transfer Phase)
D	Hard Skills Test Score (Training Phase)	Transfer Design	Quality of Skill Use (TBC)

The same procedure as outlined above was used for the relationships tested here (A-D) (refer to Table 51 to see the variables entered into Block 1 and 2). In contrast to expectations, the results showed that the potential moderator effect of transfer design was not significant for any of the relationships tested (see Table 50 for the R^2 change values and Table 51 for the B coefficients). Given that the moderator effect was not significant the first-order effects were interpreted without the interaction term included in the model (i.e., as represented in Block 1 for each relationship in Table 51).

For relationship A, while holding transfer design constant, soft skill scores at the transfer phase are predicted to increase by .992 units as soft skill scores at the training phase increase by one point. In contrast, while holding soft skill scores at the training phase constant, transfer design ratings did not contribute to a significant increase in soft skill scores at the transfer phase.

For relationship B, neither one of the first order effects was significant. For relationship C, while holding transfer design constant, hard skill scores at the transfer phase are predicted to increase by 2.71 units as hard skill scores at the training phase increase by one point. In contrast, while holding hard skill scores at the training phase constant, transfer design ratings did not contribute to a significant increase in hard skill test scores at the transfer phase. For relationship D, neither one of the first order effects was significant.

Table 50

Summary of Hierarchical Multiple Regression Analyses for the Moderator of Transfer Design

Relationship	Block/Model	R	R^2	$R^2\Delta$	$F\Delta$	p
A	1	.311	.097	.097	2.67	.079
	2	.323	.105	.008	.44	.512
B	1	.242	.059	.059	1.56	.221
	2	.285	.081	.022	1.20	.279
C	1	.339	.115	.115	3.25	.047
	2	.370	.137	.022	1.23	.247
D	1	.315	.099	.099	2.75	.074
	2	.316	.100	.001	.05	.818

Note. Δ = change. The degrees of freedom for the F statistics were the same for all relationships; Block 1 = 2, 50 and Block 2 = 1, 49.

Table 51

Regression Coefficients for Each Effect and Their Associated Confidence Intervals and T-Tests – Transfer Design

Relationship	Block and Predictor	B(SE)	95% CI	β	<i>t</i>	<i>p</i>	Power O.E.	Power M.E.
A	Block 1							
	Soft Skills (Training)	.992(.432)	.125 to 1.860	.314	2.30	.026		
	Transfer Design	.061(.428)	-.799 to .920	.019	.14	.888		
	Block 2							
	Soft Skills (Training) × Transfer Design	-.333(.505)	-1.349 to .682	-.092	-.66	.512	.65	.09
B	Block 1							
	Soft Skills (Training)	-.020(.049)	-.118 to .077	-.058	-.42	.677		
	Transfer Design	.078(.048)	-.019 to .175	.225	1.62	.112		
	Block 2							
	Soft Skills (Training) × Transfer Design	-.062(.057)	-.176 to .052	-.155	-1.09	.279	.54	.18
C	Block 1							
	Hard Skills (Training)	2.712(1.073)	.558 to 4.867	.339	2.53	.015		
	Transfer Design	-.043(1.103)	-2.259 to 2.173	-.005	-.04	.969		
	Block 2							
	Hard Skills (Training) × Transfer Design	1.147(1.036)	-.935 to 3.229	.158	1.11	.274	.76	.18
D	Block 1							

Relationship	Block and Predictor	<i>B</i> (SE)	95% <i>CI</i>	β	<i>t</i>	<i>p</i>	Power O.E.	Power M.E.
	Hard Skills (Training)	-.071(.046)	-.163 to .021	-.211	-1.56	.125		
	Transfer Design	.073(.047)	-.021 to .167	.210	1.56	.126		
	Block 2							
	Hard Skills (Training) × Transfer Design	.010(.045)	-.079 to .100	.034	.23	.818	.63	.05

Note. Power O.E. stands for the power of the overall effect and Power M.E stands for the power for the moderating effect.

Chapter Summary

Some but not all of the hypotheses relating to the training effectiveness factors were supported in the current study. It was hypothesized that high reliability levels would be found for the majority of the 17 subscales of the LTSI. In general the LTSI subscales were moderately to highly reliable (with five subscales meeting the requirement of scores at .80 or greater for inclusion in the moderator analyses – supervisor support, openness to change, content validity, self-efficacy, and transfer design). It was also hypothesized that positive correlations would be found between the training outcomes and the majority of the LTSI subscales. In contrast to expectations only one significant correlation was found between the training outcomes and the subscales. Specifically, a significant positive relationship between attitudes towards correctional work at the transfer phase and ratings of transfer design was found ($r = .54$). However, several significant correlations were found among the LTSI subscales. Specifically, supervisor support and openness to change ($r = .61$), self-efficacy and content validity ($r = .49$), and transfer design and content validity ($r = .58$). It was hypothesized that the majority of the subscales would be rated at moderate to high levels and in general the subscales were indeed rated at moderate to high levels. However, in contrast to expectations, regional differences were found for the ratings on the transfer design subscale, such that higher ratings were observed in Pacific compared to Ontario.

Lastly, it was hypothesized that each moderator would strengthen the relationship between the predictor and outcome variables, In contrast to expectations, of the five subscales included in the moderator analyses only two were found to significantly moderate a single relationship each. Specifically, content validity and self efficacy both

significantly moderated the relationship between knowledge test scores at training and transfer. In the absence of moderator effects, first-order effects were interpreted and several significant relationships were found.

For supervisor support, hard skills scores at the training phase contributed to a significant increase in hard skills scores on the job, whereas perceived supervisor support did not. In contrast, hard skills scores at training did not significantly contribute to an increase in quality of skill use on the job, whereas perceived supervisor support did. Both ATCW ratings at the training phase and perceived supervisor support significantly contributed (in an equal fashion) to increases in ATCW ratings on the job. Lastly, both ATI ratings at the training phase and perceived supervisor support significantly contributed to increases in ATCW ratings on the job, although the ATI scores were two times more effective at predicting the increase.

For openness to change, soft skills scores at the training phase contributed significantly to an increase in soft skills scores on the job, whereas openness to change ratings did not. Both soft skills scores at the training phase and ratings of openness to change ratings significantly contributed to an increase in quality of skill use on the job, with openness to change being a better predictor. Similar to supervisor support, hard skills scores at training did not significantly contribute to an increase in quality of skill use on the job, whereas ratings of openness to change did.

For content validity, similar to that found above, soft skills scores at the training phase contributed significantly to an increase in soft skills scores on the job, whereas perceived content validity did not. The same results were observed with hard skills

scores. However, both hard skills scores and perceived content validity significantly contributed (in an equal fashion) to an increase in quality of skill use on the job.

For both self-efficacy and transfer design, similar to that found above, both the soft and hard skills scores at the training phase significantly contributed to an increase in the corresponding soft and hard skills scores on the job, whereas ratings of self-efficacy and transfer design did not.

CHAPTER 7

Discussion

The purpose of this dissertation was three-fold. First, an expanded version of the evaluation framework originally proposed by Kirkpatrick (1959a) was used to determine the effectiveness of CTP. Second, the relationships between the levels of the expanded version of Kirkpatrick's framework (Levels 1-3) were examined in an attempt to provide insight into the often conflicting and limited amount of empirical research in this area. Lastly, factors that may influence the effectiveness of training (measured via the LTSI) were examined to determine their impact on the relationships explored in the second part of the dissertation. Overall, it was thought that by conducting a Level 3 evaluation of the Correctional Training Program delivered by the Correctional Service of Canada would not only provide evidence of its potential effectiveness, but also provide insights as to the suitability of an expanded version of Kirkpatrick's evaluation framework for use in training evaluations more generally.

The following section outlines the ways in which Kirkpatrick's evaluation framework was expanded while also exploring the findings associated with each of the training outcomes of CTP measured at Levels 1 through 3. Following a discussion of the training outcomes at each level, the relationships, or lack thereof, between each of the levels are discussed. In addition, an examination of the findings related to the impact of training effectiveness factors is undertaken. Lastly, a summary of all the findings as they relate to Kirkpatrick's four-level framework is provided along with a review of the limitations of the study and some directions for future research in this area.

Training Outcomes

Level 1. Previous research has suggested that including measures of perceived training utility and difficulty in reaction surveys, above and beyond the general satisfaction items that are typically included, may allow reaction surveys to be more predictive of higher-order training outcomes (Alliger et al., 1997; Axtell et al., 1997, Baumgartel et al., 1984; Ruona et al., 2002; etc.). For example, compared to the very low correlations that are found between general satisfaction with training and other training outcomes, small but significant relationships have occasionally been found between utility reactions and both learning and transfer (Alliger et al., 1997). Thus, in the current study, items tapping into utility and difficulty were included in the reaction survey (in addition to general satisfaction items that are typically included in a Kirkpatrick-style reaction measure) to test if reaction surveys are more useful if treated as multidimensional measures.

In the current study, we did in fact find that utility items tended to be rated slightly more positively compared to satisfaction and difficulty items, although this difference was not significant. However, as evidenced by the extremely high correlations between item types (i.e., satisfaction, utility, and difficulty items) that were included in the reaction survey, data from the current study suggests that examining the overall reaction score was perhaps the most fruitful way forward. The treatment of reactions to training as a one-dimensional construct is consistent with some previous research (e.g., Mathieu, Martineau, & Tannenbaum, 1993), but inconsistent with other research (e.g., Morgan & Casper, 2000).

Of course this does not mean that it would not be useful in other training evaluation studies to separate item types for the purpose of analysis. Indeed, in contrast to the current study, items related solely to utility (and even difficulty) may be much more beneficial when evaluating other types of training programs. One must remember that CTP is an entry-level training program designed to provide new correctional officers with the knowledge and skills required to work in a correctional institution. CTP trainees do not come into the program with high levels of relevant corrections-based knowledge or skills, nor with a complete understanding of their future job duties or organizational climate. Therefore, it is possible that trainees in programs such as CTP do not have the necessary frame of reference to accurately determine whether the training has utility (and thus, utility items do not stand out as being more useful than other types of reaction items). Current employees attending training programs as part of refresher training would have a much better repertoire they could draw upon to respond to utility (and even difficulty) items on a reaction questionnaire.

With all that said, regardless of whether the overall reaction measure was examined or the specific subscales on the reaction questionnaire, the trainee reaction ratings were very positive. Indeed, all of the average responses on the reaction survey were greater than 4.50 out of 5.00, except for the ratings related to Stage 1 (the new on-line portion of CTP), which had a mean of 4.37. While these ratings reflect very positively on CTP and suggest that it was well received, it should also be noted that there are often inherent biases in self-report measures. In fact, some research would suggest that trainees often have a desire to respond favourably on reaction questionnaires (Fisher

& Katz, 2000; Zerbe & Paulhus, 1987). Despite assurances of confidentiality and anonymity, the responses from trainees could still (in part at least) reflect these biases.

Another potential limitation with the reaction measure, beyond potentially biased responding, has to do with how the reaction data was collected. As indicated previously, although it is often suggested that reaction measures be collected at various points throughout a training program, because of the length of CTP and its multifaceted nature it was deemed most appropriate to collect reactions to training at the end of the in-class portion (i.e., after Stage 3). If reaction measures had been collected after each facet of the training program was completed, trainees would have filled out a very large number of surveys (>20), which would undoubtedly have led to concerns about the reliability of their responses (e.g., due to response fatigue). On the other hand, it is possible that more precise results may have been obtained if the reaction measures had been distributed after each stage of CTP. Despite these limitations it is important to note that for the current evaluation of CTP it was indeed possible to ensure that trainees do react well to the training.

Level 2. While a Level 2 evaluation based on Kirkpatrick's framework often includes a focus on knowledge outcomes solely (e.g., as might be measured by a multiple choice test), the expanded framework that was tested in this dissertation expanded the Level 2 evaluation to include both skills and attitudes in addition to knowledge. These additions are particularly important in the current context given the mandate of CTP. Recall that CTP is aimed at providing potential correctional officers with the knowledge, skills, *and* attitudes necessary to effectively manage offenders within correctional institutions once CSC hires them. Thus, in order to determine the extent to which learning

occurred as a result of CTP, it was important that knowledge, skills and attitudinal outcomes were examined before and after training, as well as four to six months on the job.

Analyses of these learning outcomes indicated that significant learning gains were observed for most outcomes immediately following CTP, whereas slight, but significant losses were observed at the transfer phase. For example, compared to pre-test scores, an impressive and significant amount of general knowledge gain (24%) was observed immediately following the completion of CTP. However, when knowledge levels were assessed again four to six months after training, a significant decrease in knowledge (approximately 5%) was observed. What is important to note, however, is that knowledge retention levels remained at 80%, which is still very high considering the vast amount of content taught during CTP over the long (six month) training period. The loss in knowledge does, however, indicate the need to ensure that appropriate job aids or refresher training are provided in a timely manner, as one would assume that as more time passes the result will potentially be an even greater deficit of knowledge.

Interestingly, the same pattern of gains and losses were observed for each of the subsections in the knowledge tests relating to the three stages of CTP. Specifically, significant gains were found for all stages of CTP immediately following CTP (gains of 12% were observed for Stages 1 and 2 while impressive gains of 35% were observed for Stage 3). Significantly lower scores were observed for the Stage 3 subsection at the recruitment phase compared to scores for Stages 1 and 2 (in addition, higher scores were observed for Stage 1 material compared to Stage 2 material). Although significant differences in subsection scores were observed at the training and transfer phases, the

differences were much less drastic compared to pre-CTP levels. Looking at the differences between the training and transfer phases, significant losses were observed for each of the test subsections four to six months on the job (losses of 4% were observed for Stages 1 and 2, while losses of 5% were observed for Stage 3). There are a number of possible explanations for the lower scores observed for Stage 3 material prior to training as well as for the higher training gains for this subsection, but perhaps the most likely explanation has to do with the content.

Unlike Stage 1 and 2 of CTP that focus on general information about corrections, which is presumably something that trainees may already know a bit about before applying for a job in corrections, Stage 3 focuses on very specific content such as Law and Policy, Security, Self Defence Skills, Arrest, Control and Baton Skills, Use of Force, Firearms, and Fire Safety. It is reasonable to assume that trainees do not come into CTP with a lot of prior knowledge or experience in these areas and thus will have the most to learn in this stage. The training content in Stage 3 may also be perceived as more relevant and applicable to the job compared to the more general information presented in Stage 1 and 2, which could further explain the significant knowledge gains for this stage. What is important to note, however, is that by the end of CTP, trainees had acquired relatively comparable levels of knowledge in the material covered in all three stages of the training program (although significant differences indicated that higher levels of knowledge were observed for Stage 1 followed by Stage 2 and then Stage 3).

It should also be noted that the assessment of knowledge retention at the transfer phase might provide a conservative estimate of retention because necessarily, a significant amount of learning takes place on the job after officers have completed CTP.

This knowledge may not have been fully incorporated by the time the evaluation took place four to six months on the job. Despite the fact that the majority of the training content of CTP is applicable in all institutions, little opportunity to cover institution-specific content beyond that related to differing security levels is provided (i.e., information relating to minimum, medium, and maximum security level institutions is covered, but that for specific institutions is not). Thus, the assessment of knowledge retention at a time period in addition to six months after training, such as one year, may be beneficial to ensure that knowledge retention levels are still maintained past six months while also allowing for sufficient time for on the job learning to take place. As indicated above, the deficits in knowledge retention observed at the transfer phase may indeed increase if opportunities for on the job learning are limited or not provided at all. Multiple follow up evaluations may highlight more specifically when and where deficits are occurring.

In addition to knowledge, both soft and hard skills were assessed as part of the Level 2 evaluation as well. In terms of soft skills, trainees came into CTP with high levels of soft skills overall as assessed by the Skills and Abilities Assessment (which looks at problem solving, de-escalation, and communication skills). Interestingly, however, there were significant regional differences found in these skills at the training phase of the study. Recall that small deficits were observed in the Pacific region (2%), while small gains were observed in the Ontario region (2%), whereas no differences were observed at the recruitment or transfer phases. Since this is a national program with a standardized curriculum, regional differences were not expected. This finding is likely due to slight differences in the delivery of the training, or the focus of training, in each of the regions.

Inevitably, instructors will have differing strengths when it comes to facilitation, as well as differences in background knowledge in the subject matter being taught.

Despite the significant regional differences at the training phase, large variations in the scores across all study phases were not observed, indicating that very little change in soft skill levels were found as a result of CTP, at least as it was measured in the current study. This lack of change, or restricted range of scores, could simply represent a ceiling effect given the high performance level of soft skills at the recruitment stage. This seems a likely possibility, especially when one considers that the Skills and Abilities Assessment tool, which was used to assess soft skills throughout this study, is in fact a recruitment tool. This means that only those trainees with high levels of soft skill abilities were selected to participate in CTP in the first place, thus limiting the room for observed skill level increases. Other possible explanations also exist, however.

Since we know that communication and problem solving skills are specifically targeted in CTP, another potential explanation for the stable level of soft skill performance across phases may be that the specific soft skills taught during CTP are not adequately captured by the Skills and Abilities Assessment tool. As indicated above, this assessment measure is the current recruitment tool used for the selection of new officers and the scenario used for the assessment of soft skills is currently a customer service situation, whereby the trainee must de-escalate an upset customer using problem solving and communication skills. In many ways, this situation is different from the scenarios that correctional officers will face in institutions. For example, in the customer service scenario trainees would arguably be more likely to give the customer what they want, utilizing a “customer is always right” mentality. In contrast, in an institution, officers are

taught to deal with situations without giving in to an offender's every demand. Although, it was assumed that the Skills and Abilities Assessment tool was appropriate as an assessment of soft skill ability, perhaps utilizing a corrections-based scenario would yield different (and more valid) results (e.g., allowing us to find improvements in soft skills across phases). Future research should be carried out to determine if the program actually does not result in improvements to soft skills or if a different assessment method is needed.

When looking at the subsections of the soft skills measure, scores on the role play component were consistently higher than the written component at each phase of the study. What is concerning is that the scores on the written component decreased (although not significantly) after training and never returned to pre-CTP levels. Report writing is a major focus in CTP and they can become important legal documents. Thus, increases in these skills are desirable.

Similar to the findings observed for the overall soft skill scores, regional differences were also observed on the subsections of the soft skill assessment. Again, it is not entirely clear why regional differences were observed here, other than the fact that perhaps the Ontario region covered the written components of the report writing sessions more effectively than the Pacific region, whereas the scenario-based components relating to communication and problem solving skills were covered more effectively in the Pacific region. Despite these regional differences, high scores were obtained in each of the subsections of the soft skills assessment.

Due to ethical and safety reasons, hard skills were not assessed prior to the start of CTP, but they were assessed immediately following the program and again after four to

six months on the job. Similar to the losses in knowledge and skill levels, which were reported above, significant losses (8%) in hard skills scores were observed on the job (compared to scores observed immediately after training). However, it is important to note that, similar to knowledge retention levels, hard skills scores still remained high (83%) at the transfer phase despite the overall loss of skill. In general, losses were observed for each of the eight individual skills with the greatest losses observed for the Frisk Search, as well as Low and High Profile Escorts. The only gain that was observed was a 2% increase found for OC Spray. It was expected that trainees would achieve comparable levels on each of the individual hard skills scores; however, that was not the case in the current study. Trainees were able to execute some skills better than others, and this differed based on training phase (e.g., scores for the radio and OC spray were not significantly different at the training phase, while at the transfer phase, scores for the OC spray were significantly higher than those for radio).

It could of course be possible that other hard skills focused on in CTP might have shown significant improvements (as indicated above, over 60 skills are taught to recruits during CTP). However, there were good reasons to evaluate the specific skills that were chosen in this study. Indeed, the major reason why these eight skills were focused on was that the subject matter experts that were brought together for this study deemed them to be the most important.

While there are perhaps some exceptions, the main contributor to the loss of hard skill ability observed in the current study likely stems from the reported lack of use of many of the skills on the job (a more in-depth discussion of this issue is provided below in the section on transfer). Given this, the concern is that these skills may continue to

decrease if they are not used by the trainees on a regular basis. A number of other factors may also compound this problem, including the lack of comprehensive refresher training and a resistant organizational culture that may discourage the use of newly trained skills (e.g., it was determined during the course of this study that existing officers were sometimes telling new officers to execute the frisk searches differently from how they were taught on CTP). Given these concerns with hard skills retention, it would be important to assess these skills at multiple time points, including a year after the completion of CTP, in addition to six months, to determine if they continue to decline, and to perhaps assess a broader range of hard skills.

Finally, while trainees' attitudes towards corrections were quite positive across all phases of the study, attitudes towards inmates were not as positive. Despite these differences between the attitude measures, both sets of attitudes (towards corrections and inmates) showed similar patterns of results in that there were no significant changes in attitude ratings as a result of CTP, while significant decreases were observed four to six months on the job (decreases of 1% and 12% were observed for attitudes towards correction and inmates, respectively). Thus, at the training phase, it appears as though more content relating to the realities of working with an inmate population may be needed, and at the transfer phase, it appears as though the organizational or transfer climate is once again having a negative impact, especially on attitudes toward inmates.

Somewhat surprising were the low correlations found between the different measures of Level 2 outcomes at each of the study phases (e.g., knowledge tests and soft skill assessments at recruitment, training, and transfer, etc.). It would appear as though the measures of learning that tap into different domains (i.e., knowledge, skills and

attitudes) are unrelated as evidenced by the non-significant correlations found between them. This is generally consistent with findings reported in previous research, which has suggested that content of the Level 2 measures used in this study (knowledge, soft skills, attitudes towards corrections, etc.) are distinct from one another, despite all being based on valid course content (e.g., there is no overlapping content or redundancies) (Alliger et al., 1997). For example, when correlations between different Level 2 measures are reported in meta-analyses, they tend to be quite low (e.g., $r = .18$, Alliger et al., 1997; $r = .09$, Tracey et al., 1995).

However, for the Level 2 measures with subsections (e.g., knowledge and soft skills), significant positive correlations were generally found between the different subsections and the overall score of the same measure at each study phase (e.g., recruitment, training, and transfer). For example, for both knowledge and soft skills, the subsections of the tests and the overall test score were significantly correlated at all study phases. Similarly, significant correlations were observed between the Stage 1 and 2 subsections of the knowledge test (for the recruitment phase only) and the role play and written subsections of the soft skills assessment were significantly correlated at the training phase. The results provide an indication that for knowledge and soft skills, the subsections appear to be distinct but related measures to the overall score because the significant correlations did not exceed levels of .80.

When examining the correlations between the same Level 2 measures across training phases (i.e., investigating the relationships between knowledge test scores at the training phase and knowledge tests scores at the transfer phase), consistent with previous research, some significant relationships were found (Alliger & Janak, 1989; Alliger et al.,

1997). For example, a significant positive relationship was found between the acquisition of knowledge (during training) and the retention of that knowledge (during transfer) ($r = .54$). Similarly, a significant positive relationship was found between ratings of attitudes towards inmates after CTP and four to six months on the job ($r = .51$). However, many of the other correlations between the Level 2 measures (soft skills, hard skills and attitudes towards corrections) across study phases did not reach significance. It is not clear why relationships were not found here beyond that for knowledge and attitudes towards inmates. However, it is possible that the lack of variation in soft skills scores and the significant losses of hard skills scores on the job could provide an explanation.

Level 3. Now that it has been determined that trainees reacted very positively to the training program and that learning gains were observed in most areas, and largely retained, it is interesting to consider the results of the Level 3 analysis wherein officers (and their supervisors) were asked to rate the frequency and quality of skill use on the job. Based on the results reported above, it appears that the CTP skills that were utilized most frequently on the job included a combination of soft and hard skills. The soft skills that were used most frequently included: assessing problematic situations, identifying possible solutions, taking necessary measures and communicating effectively orally. The most common hard skills included using communication equipment and engaging in searches. It is not surprising that the soft and hard skills mentioned above are utilized frequently as these skills are consistent with what a new correctional officer would be expected to use on a daily basis. What is surprising is the apparent lack of engagement in the vast majority of hard skills taught during CTP, such as report writing, escorting

inmates, the use of chemical agents, implementing fire safety measures, self defence, and arrest and control skills.

Due to the large amount of training time dedicated to these skills, it was expected that higher frequency scores would be observed. This finding is certainly cause for concern because it is well known in the training literature that when individuals are given no opportunity to use or practice the skills learned during training they are likely to degrade (Baldwin & Ford, 1988; Ford et al., 1992; Holton et al., 1997; Tannenbaum & Yukl, 1992; Tracey et al., 1995). The current training climate with CSC does little to ease such concerns. While officers are required to complete numerous security-related mandatory refresher courses yearly, not all knowledge and skill areas are covered in these courses and in some cases the refresher courses are really just requalification courses versus training that involves practice, etc. These refresher courses provide the only opportunity for officers to practice their skills beyond using them on the job. Thus, if the skills taught during CTP are not utilized or required on the job this represents a serious problem (i.e., the skills may degrade to such a point that the officers cannot carry out their duties effectively).

Of course, this finding may also have implications for CTP. For example, if skills taught in CTP are actually not necessary in the day to day duties of correctional officers, the number of skills focused on in CTP (and the refresher training) could be reduced or streamlined. A review of the CTP curriculum to determine the most necessary and frequently used skills across institutions could not only reduce the load on the learner and increase the chances of higher levels of retention (and thus use of skills on the job), but this may also decrease the time needed for training. Alternatively, if it is determined that

all CTP-trained skills are necessary, time for remedial training may have to be built into the duties of correctional officers and shifts in the transfer climate may be necessary to ensure that opportunities for practice are provided.

Thus, a potential explanation for the large decreases in retention levels for some of the soft and hard skills (e.g., the low and high profile escorts) is the frequency scores provided by officers. Specifically, some skills (e.g., the low and high profile escorts) were rated as never being used on the job and therefore it is not surprising that these skills were not retained. With that being said, however, frequency of use is clearly not the only factor that may be impacting retention levels. For example, conducting a search was rated as one of the most frequently used skills on the job, yet this skill is one of the eight skills where we observed the greatest loss in skill levels.

As mentioned above, another factor that may explain problems with skill retention while on the job is the organizational climate. For example, throughout the course of this study it came to our attention that new officers were often ridiculed, not only by other officers but also by inmates, about their searching procedures when they first started their job (e.g., in terms of how thorough they were and the length of time it took them to conduct a search). Indeed, it appears that on occasion officers are told to forget what they learned in CTP with regard to searching and to do it more quickly. This of course is problematic for several reasons. First, officers are risking their personal safety as well as the safety of other officers and inmates by not following correct procedures. Second, if the organizational culture does not support the training initiatives of the organization, a lot of time and money are wasted only to have officers re-learn a different way of doing things while on the job.

Officers also reported that the quality of their skills was moderate to high, with an average of 2.28 out of 3. In other words, officers indicated that they generally felt they were able to execute the skills that were sampled in the current study in a way that was reasonably consistent with what they learned in CTP. Although higher ratings of quality would of course be desirable, it is possible that these levels were not observed because a certain degree of learning on the job is necessary. Indeed, performing very well in a classroom environment is one thing, but performing the same skills in a real institution is quite another (thus, it is quite impressive that officers felt that they made the transition from classroom to institution as well as they did).

It should be noted, however, that there is a discrepancy between some of the objective measures of skill use (as evidenced by the hard skill test scores) and the respective quality ratings provided by the officers, such that officers perceived high levels of quality for some skills that they actually had difficulty executing (e.g., frisk search). This finding highlights an inherent limitation in the transfer measure (for both the frequency and quality ratings), specifically that these ratings are measured via self-report and officers may have difficulty providing accurate and precise estimates of skill use on the job. With that said, the self-report measures were arguably as good as they could be, and were constructed in accordance with current recommendations for the assessment of skills (Phillips & Phillips, 2007). For example, trainees were informed at the outset of the study and reminded immediately after CTP that they would be filling out a questionnaire that would be assessing the frequency and quality of the skills they learned in CTP. The questionnaire was designed to be as simple and clear as possible to ensure that responding was not difficult. Detailed instructions were provided on how to

rate the frequency of skill use on the job along with a clear frame of reference for each rating scale (i.e., on average per month) and explicit intervals (e.g., 2 corresponds to 1-5 times).

A potential method for overcoming some of the limitations of self-report measures of skill use is to obtain supervisor ratings of transfer in addition to those provided by the officers themselves. This was done in the current study. Unfortunately, we did not receive supervisor ratings of skill use and quality for all the officers in the study and thus, we were not able to analyze this data in a very sophisticated fashion (in addition, the sample of supervisors was approximately half that of trainees). However, based on the data we did collect it appears that supervisors provided similar (although slightly lower) frequency ratings to those provided by the officers (particularly with regard to problem solving and communication skills). The quality of skill use was also rated similarly by the supervisors, although these scores were slightly higher than the ratings provided by the officers. Surprisingly, in light of other research (Axtell et al., 1997; Rouiller & Goldstein, 1993; Xiao, 1996), the correlations between trainee and supervisor frequency and quality ratings did not reach significance in the current study (with the exception of quality of use of chemical agents, $r = .71$).

Certainly the ratings provided by supervisors should be viewed with an appropriate degree of caution. Indeed, there are several limitations associated with this measure. First is the already stated limitation of sample size. However a second, equally important limitation has to do with how meaningful the data is that was obtained with this measure. After the current study had begun it was brought to the author's attention that supervisors do not always have the opportunity to directly supervise their officers and

therefore their ratings may not always have been based on actual observations of skill use. Other research, where supervisors did not spend very much time with employees on the job after training, has found discrepancies between the ratings provided by trainees and their supervisors (Van der Klink, Gielen, & Nauta, 2001). Although that was not entirely the case here, of concern is that the lack of direct supervision in the current study may have led supervisors to consult with officers before completing the measurement tool, providing responses that are similar but not based on true observation.

Correlations Between the Levels

Correlations between the different levels of Kirkpatrick's framework were examined in order to determine whether or not the evaluation framework is truly hierarchical in nature as implied by Kirkpatrick (1994) (i.e., that results from lower levels will predict results from higher levels). In the current study, the results are very consistent with previous research indicating that the outcomes at one level of the framework do not generally predict outcomes at other levels (Alliger & Janak, 1989; Alliger et al., 1997; Brown, 2005; Colquitt, LePine, & Noe, 2000; Mathieu et al., 1992; Noe & Schmitt, 1986; Warr & Bunce, 1995, etc.). In other words, there were very few instances where reaction, learning, and transfer outcomes were significantly correlated. Thus, the conclusion drawn in many studies that a relationship between trainee reactions, learning and/or transfer outcomes does not exist is supported in the current study.

Turning our attention first to the hypothesized relationship between reactions and other training outcomes, the lack of significant results is very interesting. Indeed, while our results are consistent with the vast majority of evaluation research that does not support a link between reactions and other outcomes, there is quite a large body of

literature, which suggests that a link should exist. For example, studies of affect more generally (in terms of mood and emotion) suggest that positive or negative states experienced during training should influence levels of motivation (Erez & Isen, 2002) and engagement during learning (Ainley, Hidi, & Berndorff, 2002; Brown, 2005). Similarly, research suggests that higher levels of affect can enhance the ability to process information in creative and integrative ways (Estrada, Isen, & Young, 1997; Isen, 1993). In addition, both cognitive and skills-based learning outcomes have been predicted by engagement, persistence, and effort (Ainley, et al., 2002; Fisher & Ford, 1998). So, given these findings, why is there no significant relationship found between reactions and learning and/or transfer in this and other studies?

Several researchers have suggested that the lack of a relationship between reactions and learning or transfer outcomes has to do with the way reaction measures are conceptualized. For example, some researchers have argued that reaction measures might act as moderators between other training effectiveness factors and training outcomes, rather than having a direct impact on those outcomes (Holton, 1996). For example, Mathieu and colleagues (1992) found that reactions to training significantly moderated the relationship between motivation to learn (referred to as training motivation) and learning. They also found that reactions to training significantly mediated the relationship between motivation to learn coupled with self-assignment to training (i.e., individuals who volunteered to attend training versus those assigned to attend) and post-test performance scores. The authors concluded that reactions to training can have complex indirect impacts in training contexts (Mathieu et al., 1992). It would have been interesting to formally test these possibilities in the current study; however, the low

reliability scores observed for the motivation to learn subscale of the LTSI prevented its examination (more about this is provided below). However, based on the majority of evidence pointing to the inability of reaction-level data to predict learning and/or transfer outcomes, this research conceptualizing trainee reactions as a secondary, rather than a direct outcome, provides a plausible explanation.

In addition, the timing of the collection of reaction measures should be considered. Specifically, Alliger et al. (1997) suggest that reactions relating to the utility of training may be more useful if gathered after training, once trainees have had some experience on the job. In addition, as indicated above in Chapter 4, it is often beneficial and desirable to collect reaction measures at multiple time points during training (especially for longer training programs) to ensure that participants can accurately reflect on the elements of the program being assessed (Phillips & Phillips, 2007). Although this was not done in the current study, in order to prevent fatigue in responding and the potential result of unreliable data, perhaps different results would be obtained by collecting this data at multiple time points. However, as indicated by Phillips and Phillips (2007), it is still important to gather data at the end of a training program as this allows participants to reflect on the training experience as a whole.

Because we expanded Kirkpatrick's original framework at each level to include additional outcome measures, we can also speak to a range of other questions that have been asked in the training evaluation literature. For example, recall that some researchers have suggested that different types of trainee reactions (e.g., utility and difficulty) may be more likely to predict learning and transfer outcomes than the general satisfaction items that are typical of Kirkpatrick's Level 1 measures (Warr & Bunce, 1995). Although, high

ratings of utility were observed in the current study (as discussed above), these ratings were not significantly correlated with training outcomes at any other phase of the study. The same can be said for the difficulty items.

It is unclear why difficulty ratings did not correlate significantly with other training criteria, but there is a likely explanation for the lack of significant correlations associated with the utility ratings. Indeed, the lack of correlations in this case are likely to be due to the fact that CTP targets new employees, most of which have a limited frame of reference with respect to the utility of what is covered during training. Despite the fact that some studies have reported significant correlations between utility ratings and other training criteria (e.g., Alliger et al., 1997; Ruona et al., 2002; Tan et al., 2003; Tannenbaum & Yukl, 1992) this could have been because the trainees in these cases had knowledge of the work environment to which they would return and apply their learning. Indeed, the vast majority of evaluation studies have examined training programs like this; programs that have targeted current employees who are on the job (e.g., Ruona et al., 2002; Tan et al., 2003, etc.). Thus, the assertion that different types of trainee reactions may be more likely to predict learning and transfer outcomes may depend, to some extent at least, on the type of training being offered or the target group of the training program. Future studies are required that examine different reaction ratings, and their relationships to higher-order training outcomes, while exploring different types of training programs (e.g., development, refresher and core training programs) and types of employees (e.g., new or current).

Turning now to the results of learning and transfer outcomes, the current study produces similar results to those found with the reaction measures (i.e., non-significant

correlations for the most part). These findings are consistent with some previous research (e.g., Alliger & Janak, 1989), but not other research that has found small but significant relationships between what was learned during training and ratings of transfer (e.g., Axtell et al., 1997; Alliger et al., 1997; Chen, Thomas, & Wallace, 2005; Warr & Bunce, 1995).³² It is possible that the lack of correlations between Level 2 and 3 measures in the current study is due to the Level 3 measures focused on (i.e., trainee ratings of frequency and quality of skill use). In fact, most of the studies reporting significant correlations between training and transfer outcomes utilize ratings of skill application on the job by supervisors or peers. For example, Richman-Hirsch (2001) found a small but significant correlation between knowledge test scores and frequency ratings of skill use on the job ($r = .14$) provided by supervisors and coworkers after a customer service skills training program. Similarly, Warr and Bunce (1995), found a significant correlation after a management training program between learning scores (determined by a tutor, not an objective test completed by trainees) and transfer ratings provided by immediate supervisors four months after training ($r = .24$). However, this relationship was not maintained when measured again seven months after training. Although supervisor ratings were also gathered in the current study, as part of the expansion of the evaluation framework, further analyses were not possible due to low response rates coupled with data quality concerns. It is possible that different results, similar to that found in other studies may have been found if this data could have been analyzed further.

However, it should be noted that relying solely on transfer ratings from supervisors or coworkers (as was the case in the research described above) may be

³² Although it must be said that some of this research (e.g., the meta-analysis by Alliger et al. (1997)) is not particularly convincing (e.g., because it was based on only two studies).

problematic as this may not be the most reliable indicator of training transfer and therefore conclusions about the predictive relationship between training outcomes and transfer may be misleading. For the same reasons as that observed in the current study, the reliability of supervisor data may be questionable given that these individuals that may not have a chance for observation, they may be unaware of the objectives of the training program, they may also have a vested interest in providing favourable ratings, etc.

Regardless of measurement differences, it still seems counterintuitive that correlations between these measures do not exist. As indicated above, this finding indicates that perhaps there were issues with the measurement tools (either at training or transfer). For example, with the use of more objective transfer measures of actual job performance (e.g., an assessment of written reports completed by trainees on the job to determine whether they are completed correctly, etc.) it may be possible to find correlations between learning outcomes and transfer outcomes. Alternatively, other factors relating to the transfer of training may be having an impact on the results, such as the training effectiveness factors that will be discussed next. This last possibility is particularly important, because if training effectiveness factors are indeed having an impact here, and one simply relies on the lack of significant correlations between training outcomes and application on the job to draw conclusions about the training, then it is likely that inappropriate conclusions about the effectiveness of training will be made (Bates, 2004; Holton, 2005).

Training Effectiveness Factors

One of the major criticisms of Kirkpatrick's four-level framework is that it does not explicitly include a measure of training effectiveness factors that may impact the outcomes measured at each of the levels, and the relationships between these outcomes (Guerra-Lopez, 2008; Holton, 1996; 2005; Holton et al., 2000; Phillips, 1996; Phillips & Phillips, 2007). For example, one of the most-cited models proposed by Baldwin and Ford (1988) posits that many factors will either negatively impact or positively enhance the effectiveness of training. Baldwin and Ford's model relates broadly to three areas: trainee characteristics (e.g., self-efficacy), training design (e.g., content validity, transfer design) and work environment (e.g., openness to change, supervisor support). They suggest that evaluation frameworks that do not include measures of these factors are not likely to provide a full picture of the reasons why a program is deemed effective or ineffective.

Other, more recent models of training evaluation also include such factors. For example, Phillips and Phillips (2007) specifically address these types of training effectiveness factors in their ROI Methodology, referring to them as barriers and enablers. Phillips and Phillips call for their measurement at various levels and times during an evaluation. For example, some factors can be measured as part of a Level 1 questionnaire (e.g., the degree to which learners are motivated to learn the material). Other factors that could be measured in a Level 2 evaluation include one's confidence and motivation to apply what was learned in training. At Level 3 (learning transfer in the ROI Methodology) measures such as supervisor support, transfer design and opportunities to use what was learned could be incorporated.

Finally, over the last several years, there has been an even greater emphasis placed on the transfer climate as a potential training effectiveness factor (Holton, 2005). Specifically, research has now made it clear that training can be effective, but still not transfer to the job if the organizational climate is problematic (Holton et al., 2000). Indeed, some researchers have argued that if a training program itself is found to be effective, yet transfer or organizational results are not favourable, the measurement of transfer climate could identify that an intervention is needed at the organizational level instead of at the training level (Holton, 2005; Phillips & Phillips, 2007).

To help accomplish this, Holton (1996) developed and validated the LTSI, which was designed to provide a systematic measure of 16 (now 17) training effectiveness factors. Holton (1996) designed the LTSI to provide a means to identify potential training factors of relevance for the specific organization or program being evaluated and to alleviate issues surrounding the use of unreliable and invalidated measures of training effectiveness.

While there is no debating that appropriate steps must be undertaken to ensure the reliability and validity of instruments used to assess training effectiveness factors in any study, one can see from the results of the current study that utilizing scales, which are found to be reliable and valid for one training program or organization may not be reliable and valid in another. For example, despite the large amount of effort that has been put into validating the LTSI, not all of its scales were found to be reliable in the current study. Specifically, only five subscales on the LTSI had reliability scores over .80: supervisor support, perceived content validity, resistance-openness to change, transfer design and performance self-efficacy.

Despite some concerns with the reliability of the LTSI in the correctional context, it was used in the current study in the hopes that it would provide interesting insights into many training effectiveness factors. It is believed that this occurred. For example, the majority of factors on the LTSI were only rated moderately (i.e., around 3 out of 5), indicating that there are many potential weaknesses identified in the transfer system at CSC with regard to CTP. This finding is consistent with the findings of Holton, Chen, and Naquin (2003) where a variety of organizations were studied. Factors that were particularly problematic according to CTP trainees included: positive personal outcomes, supervisor opposition, supervisor support, resistance-openness to change, negative personal outcomes, perceived content validity, peer support and learner readiness, as evidenced by mean ratings under three.

Interestingly, the degree to which specific training effectiveness factors are perceived as problematic by trainees in the broader literature does appear to vary, both across organizational settings and types of training. For example, in a study by Holton and colleagues (2003), significant differences were found in the degree to which training effectiveness factors were deemed problematic across private, public, and nonprofit organizations for all of the LTSI subscales except for learner readiness and perceived self-efficacy. With respect to training type, significant differences were found in Holton and colleagues (2003) study across different types of training (e.g., supervisory vs. sales training) for all of the LTSI subscales except for perceived content validity and performance coaching. This type of research highlights the fact that each organization and each training program has different objectives, applications, and impeding factors and thus, evaluation of training must be customized to each organization and each

training program (Mathieu et al., 1992). This research also suggests that the particular problem areas identified by CTP trainees are likely to be somewhat specific to CTP or to the correctional context more generally.

Using the five factors of the LTSI that were found to be most reliable in the current study (supervisor support, openness to change, content validity, self-efficacy and transfer design) correlations between the training effectiveness factors and the training outcomes (e.g., knowledge scores, behavioral checklists, etc.) were examined. Surprisingly, these correlations were all non-significant, with the exception of the significant correlation between ratings of attitudes towards corrections at the transfer phase and ratings of transfer design ($r = .54$). This correlation is likely due to the fact that attitudes towards corrections, as rated by trainees on the job, are closely linked to their experiences transitioning into the institutional environment, and thus the degree to which they perceive that training has set them up for success on the job.

In order to further examine the relationship between the five training effectiveness factors and the training outcomes, we examined whether any of the factors moderated specific relationships between training outcome measures. The use of hierarchical multiple regression analysis allowed us to answer questions of the type: Does the lack of supervisor support impact the relationship between one's scores on various skill tests during training and the transfer of those skills to the job?

The regression results indicated that overall, very few of the training effectiveness factors were significant moderators of the relationships between the various training outcomes. Specifically, three out of the five training effectiveness factors were not found to be significant moderators for any of the relationships examined (e.g., supervisor

support, resistance-openness to change, and transfer design). However, content validity was found to significantly moderate the relationship between knowledge scores at the training phase and knowledge scores at the transfer phase. In other words, for those officers who perceived the training to have high levels of content validity, their levels of knowledge acquisition after CTP related significantly to their level of knowledge after four to six months on the job. Similarly, performance self-efficacy significantly moderated the relationship between knowledge scores at the training phase and knowledge scores at the transfer phase. In other words, for those officers who perceived themselves to have average levels of self-efficacy, their levels of knowledge acquisition after CTP related significantly to their level of knowledge after four to six months on the job.

While the findings related to content validity are somewhat promising, it was expected that content validity would moderate more than one relationship between outcomes in the current study. Although there is not a vast amount of research examining the role of content validity as a moderator in the training evaluation literature, previous research has shown that the more the training content is related to the work environment (i.e., the greater the relevance), the more likely the content will be transferred to the job (e.g., Axtell et al., 1997, Bhatti & Kaur, 2010; Gagne, Baker, & Foster, 1950; Holton et al., 2000; Thorndike & Woodworth, 1901; Underwood, 1951; Velada et al., 2007). The significant results found in the current study suggest that connections between knowledge-based training content and job requirements are made, whereas similar connections for skills-based training content are lacking.

Only moderate ratings of content validity were observed, thus it is possible that trainees feel that not enough connections are made regarding the use of skills learned on the job, or that the connections made are too general given the broad (i.e., non-institutional) perspective of CTP. Indeed, trainees indicated the desire for more institution-specific training content in their qualitative feedback. The LTSI was completed at the transfer phase and therefore ratings of content validity about CTP were provided four to six months after training, which could have contributed to the lower ratings observed here. For example, once on the job, as indicated by the frequency ratings, trainees were not able to execute a large portion of the skills taught during CTP and this likely influenced their perceptions regarding the content validity of the training.

It was also expected that self-efficacy would moderate more than one relationship between outcomes in the current study. Although more attention is given to self-efficacy as a moderator in the literature, the moderating relationship often involves different predictors than those focused on in the current study. Some researchers have found that, similar to trainee reactions, self-efficacy might act as a moderator between other training effectiveness factors (specifically motivation to learn) and learning or transfer outcomes (Axtell et al., 1997; Kanfer & Ackerman, 1989; Mathieu et al., 1993). Thus, it is possible that if motivation to learn was included as the predictor variable (instead of the learning outcomes) in the current regression analyses, significant moderating effects may have been observed. Unfortunately, it was not possible to test this relationship due to the low reliability of the motivation to learn subscale of the LTSI.

Another potential explanation for the lack of significant findings with respect to self-efficacy, especially for the relationships involving training outcomes that go beyond

knowledge, is the fact that there is little opportunity to execute a large portion of the skills learned during CTP when the trainees finally get a job (as indicated by the frequency ratings of skill use on the job). Despite officers feeling that they are capable of executing the skills taught during CTP (as evidenced by the high ratings of self-efficacy) they may be discouraged to execute them in the manner they were taught, thereby preventing the moderator effects that were expected.

Beyond content validity and self-efficacy, it was also expected that moderating effects would be found for the three remaining training effectiveness factors – supervisor support, resistance-openness to change, and transfer design. However, similar to content validity, previous research has shown that relationships exist between these factors and training outcomes but there is a lack of research examining them as moderators. For example, previous research has shown that increased perceptions of supervisor support often results in greater transfer (Foxon, 1997) and the more opportunities there are to use newly acquired knowledge and/or skills, the more likely it is that transfer will occur (Clarke, 2002). In the current study, these two factors were not rated very highly. In other words, it appears as though, at best, there are moderate levels of perceived supervisor support as well as openness to change by the larger institutional environment. Therefore, without their presence, the absence of moderating effects becomes less surprising.

On the other hand, it is not entirely clear why transfer design was not a significant moderator as trainees rated this subscale quite highly (although the high ratings were not uniform across the entire sample, significant regional differences indicated that higher ratings were observed in the Pacific region). But again, previous research indicates that only significant relationships via correlations have been found versus moderator effects.

For example, the greater the degree to which training has been designed and delivered to give trainees the ability to transfer learning to the job, the greater the chance that what is learned during training will transfer to the job (Holton, 2005).

The most likely reasons for the lack of significant moderating effects could simply be that no relationship exists in the current study. In addition, conditions such as the low sample size ($n = 53$ at the transfer phase) and the power concerns that accompany this issue can contribute greatly to the lack of significant findings. Lastly, overall, the range of responses was limited for the majority of LTSI subscales and this can also have an impact of the lack of findings. It is also possible that for each of the moderator relationships tested, alternate predictor and outcome variables may have been appropriate. As indicated above it is possible other LTSI subscales could themselves be predictors; however, due to low reliability scores these relationships could not be tested. For the most part, in the current study, the predictor and outcome variables are considered appropriate and inclusive.

Of course, beyond the potential explanations already discussed above, which may explain the lack of significant moderator effects associated with LTSI factors, there are many other limitations associated with the LTSI (beyond reliability issues) that may have limited its ability to be useful in the current study. For example, despite being designed for research purposes, the LTSI remains very difficult to administer and to score. As is typical in training evaluation studies, the 93-item LTSI was administered after CTP trainees completed their training program. Given the length of the LTSI, and the fact that it was only one of a number of evaluation tools completed at one time, it is likely that this led to some response fatigue while completing it. Perhaps if the survey could be broken

up according to subscales, researchers and practitioners could choose to administer the subscales of interest at various time points throughout the training program, as appropriate. For example, subscales that are relevant to pre-training issues (e.g., motivation to learn) could be administered before the training commences and this might lead to more meaningful responses (as well as limit fatigue). Due to the length of CTP it was likely that trainees in our study experienced difficulty reflecting back to their perceived state of mind months before the training program began (e.g., to assess their motivation to learn).

In addition to these administration issues, it was also impossible to score the LTSI due to the fact that Holton does not share the coding scheme for scoring or analysis. Instead, all LTSI data must be sent to one of Holton's colleagues for this purpose. This appears to be done in order to ensure that data is shared with the developers of the LTSI. However, this procedure is obviously far from ideal and it does create serious problems. In the current study, for example, timing was a serious issue. It took over a year to receive the subscale scores for the LTSI. Practitioners, and researchers working within applied settings, will have a difficult time justifying the use of the LTSI to their organizations when such delays occur. Even more problematic than the time issue though is the fact that the current scoring and analysis procedure prevents researchers from carrying out important analyses, such as confirming the factor structure of the LTSI items. This is a serious limitation that must be overcome in order for the LTSI to reach its full potential.

Finally, like other survey instruments, there may be a response bias associated with the LTSI, which makes it difficult to interpret the results emerging from it. For

example, the majority of the subscales received moderate scores (around 3 out of 5). This may of course represent the actual perceptions of trainees, but it could also represent a response bias, especially if survey fatigue was an issue (as mentioned above). Holton (1996) suggests that after potential barriers are identified through the use of the LTSI follow up analyses could be conducted via focus groups, interviews, or additional questionnaires in order to flesh out the findings. The problem with doing this in the current case of course is that the majority of the LTSI factors were rated as moderately problematic. Conducting follow up analyses on all of these factors would be a very lengthy, costly, and time consuming process.

Summary of Questions Addressed in the Current Study

As outlined at the outset of the dissertation and reiterated at the beginning of the discussion, the purpose of this dissertation was three-fold. In the following section, each question asked in this thesis will be answered and implications for the suitability of Kirkpatrick's expanded framework will be discussed. In addition, recommendations for modifications to CTP as well as the expansion of the evaluation framework will be provided stemming from the answers to the three questions.

The first question asked whether or not CTP was an effective training program, as determined using an expanded version of Kirkpatrick's evaluation framework (1959a). It is fair to say that using this framework we were able to determine that CTP is indeed an effective training program. This is evidenced by the favourable trainee reactions, knowledge gains, high level of hard skill acquisition, and maintenance of soft skills and attitude levels. Important information was also gleaned regarding the retention levels of knowledge, skills and attitudes using transfer tests and self-reported frequency and

quality ratings of skill use, all of which were conducted after trainees had been on the job for four to six months. Despite significant losses on all training outcomes at the transfer phase (with the exception of soft skills), the levels of knowledge, skills, and positive attitudes remained high.

The value of expanding Kirkpatrick's evaluation framework (1959a) varied depending on the level of the framework. The expansions to Level 2 were especially valuable for CTP staff (and CSC more generally). By expanding Level 2 measures to go beyond the typical declarative knowledge outcomes (to include both soft and hard skills, and attitudinal measures), it was possible to speak to very specific training gains and losses that are important for CTP trainers and curriculum designers. However, due to the highly correlated nature of the Level 1 measures, including items on the reaction survey that tapped into utility or difficulty made little difference given that their unique effects could not be independently tested. Similarly, for the Level 3 measures, although it was informative to expand the original framework to include feedback from supervisors, issues of reliability and low sample size prevented any sophisticated analyses from taking place.

The second question asked if significant relationships did in fact exist between the levels of the expanded Kirkpatrick (1959a) framework in the current study. This issue was examined in an attempt to provide insight into the often conflicting and limited amount of empirical research in this area. Unfortunately, the results of the current study tend to support other research that has failed to find relationships between the levels of Kirkpatrick's framework (Alliger & Janak, 1989; Alliger et al., 1997; Tracey et al., 1995, etc.). Indeed, nothing changed simply because the levels of Kirkpatrick's original

framework were expanded – there is still no evidence that significant relationships exist across the levels of Kirkpatrick’s framework.

As indicated above with regard to trainee reactions, the lack of support for relationships between Level 1 measures with Level 2 and 3 measures warrants further investigation and thus future studies should be undertaken to examine not only how reaction measures are conceptualized but when they are measured. For example, further research is needed to determine if reaction measures play more of an indirect or secondary role as moderators between training effectiveness factors (such as motivation to learn) and training outcomes instead of being conceptualized as primary outcomes. In addition, further research is needed to determine the most valuable time to collect reaction measures (although it is acknowledged that this will be informed in large part by the training program being evaluated) and to determine if different types of reaction measures (e.g., utility or difficulty) are inappropriate for some types of training programs (e.g., entry-level versus developmental training, etc.). Perhaps, reaction measures collected after training, when trainees are back on the job is more fruitful than those collected during training (especially for entry-level programs).

Lastly, the lack of correlations between the Level 2 and 3 measures also warrants further investigation, specifically with regard to the various types of transfer measures utilized. Unfortunately in the current study, inconsistencies in this area could not be cleared up due to the inability to analyze the supervisor data collected as part of the expanded evaluation framework. For example, future research should be conducted whereby trainee, supervisor and coworker transfer ratings are collected and compared in various training contexts.

The last question asked was whether or not there are factors that may influence the effectiveness of training (as measured via the LTSI) as well as the relationships within and between levels. One of the criticisms of Kirkpatrick's framework (1994) is that it does not explicitly include a measurement of factors that may impact the extent to which trainees learn the material taught during training and apply it on the job. Thus by adding this component in the current study, we were again able to provide CSC with valuable information they can use to understand their trainees, CTP, as well as the broader organizational culture. However, the low reliability of the LTSI subscales along with low sample sizes and low power prevented further statistical testing of all factors.

Consistent with previous research, overall the ratings for the training effectiveness factors were moderate to high (Holton et al., 2003). Those factors with the lowest levels highlight the weakest areas in the transfer climate at CSC with regard to CTP. In this study, these included: positive personal outcomes, supervisor support, resistance-openness to change, negative personal outcomes, perceived content validity, peer support and learner readiness. Those rated more positively include: self-efficacy, motivation and intent to transfer what was learned, and transfer design. Thus, trainees appear to be coming out of CTP feeling equipped and ready for their job, however, several organizational barriers, such as supervisor support, openness to change, peer support and content validity, may be getting in the way. The only significant moderator effects identified were that higher levels of content validity as well as self-efficacy coupled with higher knowledge scores at the end of CTP predicted greater knowledge retention levels.

Beyond the practical recommendations that come out of this study (provided below), the current study is also beneficial to training evaluation literature as a whole due

to the fact that studies including measures at Levels beyond 1 and 2 are rare (Alliger & Janak, 1989). In addition, few studies have relied on a pre-post design such as the design used in the current study. Overall, the results have indicated that despite the recommended modifications to Kirkpatrick's four level evaluation framework, we were still unable to provide evidence that this model can overcome the limitations associated with it. Indeed, our results support the criticisms of this framework made by other researchers (e.g., Alliger & Janak, 1989; Bates, 2004; Guerra-Lopez, 2008, etc.). For example, despite the use of an expanded framework, lower order training criteria in the model were still not correlated with the higher-order training criteria.

Despite the fact that CTP effectively provided new officers with the knowledge, skills, and attitudes necessary to work in an institutional environment, several areas for improvement (with regard to CTP) were identified in this evaluation. In addition, several recommendations for conducting training evaluations in the future are provided. The following are the recommendations stemming from the results of the evaluation and the evidence (or lack thereof) for the expanded version of Kirkpatrick's evaluation framework (1994). Those recommendations relating specifically to CTP are presented first, followed by the general recommendations for future training evaluations.

CTP-Specific Recommendations

1. Information gathered via the trainee reaction questionnaires should be used for formative purposes only and can be used to provide context to the Level 2 results (i.e., training outcomes).
2. The focus of most of the current refresher training is on skills, therefore to ensure knowledge deficits do not increase further, it is recommended that appropriate

knowledge-based content topics be incorporated into refresher training where applicable and that job aids are created that can serve as a quick reference when needed by officers on the job (although their use should be encouraged by supervisors and management).

3. Review the coverage of soft skills content in CTP as levels around 70% were achieved (which is the cutoff for passing all formal tests in CTP) and because they are the skills utilized most frequently on the job, higher levels should be attained as result of CTP (although it is acknowledged that it is possible different outcomes may be observed with a different measurement method than that used in the current study).
4. Ensure that the high levels of hard skills scores achieved following CTP and at transfer are maintained. Enhance opportunities for practice and remedial training on the job (e.g., via the on the job training portion of CTP), and ensure yearly refresher training is as comprehensive as possible. Review curriculum to determine if the number of hard skills can be reduced as we focused on only a small sample of the vast number of skills learned during CTP.
5. Ensure that appropriate content is added to CTP to prepare trainees for their jobs within the correctional environment, with time spent educating new officers on the realities of working with an inmate population.
6. Conduct follow up testing at a later period (such as one year after training) to determine if knowledge, skills, and attitude levels were maintained.
7. Results from the lower levels of the evaluation framework (e.g., high levels of trainee satisfaction) should not be used to predict results from the higher levels of

the framework (e.g., knowledge gains or high frequency of soft skill use on the job). In other words, high ratings of satisfaction did not lead to higher training outcomes in CTP.

8. Further investigation of the impact of the training effectiveness factors by CSC is recommended, to more fully determine their impact on CTP as well as the transfer of what is learned on the job. The low subscales on the LTSI in the current study can serve as a starting point.

General Recommendations for Training Evaluations

1. Trainee reaction questionnaires should be administered during training to gather important formative information about the training program being delivered. The items included in trainee reaction questionnaires should tap into any areas of interest for the training program being evaluated (taking into consideration whether or not it is an entry-level program or developmental for current employees). Items pertaining to general satisfaction, utility and difficulty (relating to the training content) are recommended. Reliability and correlational analyses are recommended with every sample to determine whether or not the measure is one- or multi-dimensional.
2. Continue to ensure that if multiple domains (i.e., knowledge, skills, attitudes, etc.) are targeted during training that the Level 2 measures utilized for an evaluation should be expanded to match these domains in order to provide meaningful information about learning gains and deficits. Similarly, depending on the training context, multiple measures of transfer (Level 3) may be beneficial.

3. Ensure that the Level 1 measure (i.e., trainee reactions) is not relied on to determine whether or not learning has occurred as a result of training or that transfer of training outcomes has occurred. Similarly, ensure that Level 2 measures (i.e., learning) are not relied on to ensure that transfer has occurred. The current study has provided evidence that predictive relationships between levels of the evaluation framework were not found.
4. Continue to examine training effectiveness factors relating to the trainee, the training itself and the transfer climate that may impact the results of training; however, the use of the LTSI for these purposes is limited due to the vast amount of factors included for measurement and the difficulties with administering and scoring.

Additional Study Limitations and Directions for Future Research

A number of limitations have been highlighted throughout the discussion, such as collecting trainee reactions at the end of CTP instead of throughout the program, the potential problems with administering and scoring the LTSI and the use of self-report measures at the transfer phase. However, there are several, more general limitations associated with the current study that are also worthy of mention.

One of the main limitations of this study is the relatively small sample size and therefore the statistical power associated with some of the analyses that were conducted. Although studies with multiple phases, administered over a long period of time, in an applied setting, necessarily impose challenges in this regard, one cannot minimize the potential problems associated with small samples. Data for only 53 trainees was available across all three phases of the study and for some statistical tests this meant relatively low

post-hoc power. Thus, a degree of caution is certainly required when drawing conclusions from the tests. Obviously a larger sample would have been desirable and continued attempts should be made to expand upon this evaluation of CTP to collect data from additional trainees. With this all being said, however, it is important to stress that this study does speak to the potential value of CTP and it does contribute to the larger field of evaluation research by contributing to evidence associated with the value (or not) of Kirkpatrick's four-level evaluation framework.

Another limitation, which has been touched upon already in the discussion, relates to some of the measures that were used in the current study. Indeed, it is entirely possible that the lack of correlations found in the current study between the levels of the evaluation framework was due (to some extent at least) to the way that training outcomes were measured. For example, recall that the soft skills measure consisted of a tool already utilized by CSC for the purpose of trainee recruitment. In hindsight, this may not have been ideal. Not only may it not tap into the soft skills related to CTP, it may also bias the results because, as a recruitment tool, it almost ensures high scores at the outset of the program. In addition, for the knowledge and hard skills tests, it was not possible to test all the content covered in CTP due to the vast amount of time and resources that would have been required to do this. Instead, a random selection approach was taken to sample a small amount of content and skill areas (e.g., only 8 of 60 hard skills were actually assessed). It is possible that the wrong content and skills areas were assessed and certainly future research should carefully consider this issue.

In terms of the Level 3 measure, it is also important to highlight that the behavioural checklist covered dimensions that, although pre-defined as central to an

officer's role and stemming from the objectives of CTP, only included skill related measures and did not directly tap into trainee knowledge. Due to the nature of transfer measures, which are aimed at determining performance on the job, they tend to focus on observable behaviours and skills. This naturally precludes knowledge related measures, which are generally thought to be less observable. In hindsight though, some knowledge related measures (e.g., knowledge of policies) could have been included in the transfer tools to enable a more well-rounded measure of transfer to the job. In addition, the behavioural checklist used to categorize the frequency of behaviours on the job was not suitable for mean-level analysis and thus precluded the frequency of skill use to be included in some of the analyses (e.g., regression). It may be useful in the future to include other types of measures to assess job transfer that are able to capture knowledge and skill use more accurately. Although quite intensive, these might include observational studies, or experience sampling using technology that captures behaviour on a daily, or even more fine grained basis.

It also cannot be ignored that some measures, such as the trainee reactions, attitude ratings, behavioural checklists, and the LTSI were self-report measures. Thus, they come with the inherent risk of bias in terms of providing responses that trainees think are expected, or the risk that people do not have access to the information being requested of them (i.e., problems with introspection). As already highlighted, research has clearly indicated the potential biases that might be common when using self-report measures (Fisher & Katz, 2000; Paulhaus, 1984; Zerbe & Paulhaus, 1987) and some researchers have argued that people are not good at reporting on their learning experiences (Alliger & Janak, 1989; Hofstadter & Dennett, 1981; Thorndike, 1935).

Thus, it could be argued that the results obtained with these rating tools are a methodological artifact rather than an accurate reflection of reality. While it is very difficult to overcome these issues, researchers in the future should consider whether there are other ways of measuring training outcomes that may result in more reliable data. With that said, trainee perceptions are important and they should not be ignored (Axtell et al., 1997; Phillips & Phillips, 2007; Wexley & Baldwin, 1986).

Finally, a control group (such as a group of trainees taking part in older versions of CTP) was not used in the current study. Obviously this would have allowed us to speak more directly to the value of CTP as it is currently being delivered. However, a control group would have been extremely difficult to set up and very costly. More importantly, the development of a control group based on the previous version of CTP was simply not possible given that the updated CTP was already being implemented nationally during the data collection phase, and CSC indicated that there was no need or desire to examine the previous version of CTP as it was outdated. Without a full control group design, the author acknowledges that any differences between pre- and post-test measures of learning cannot conclusively be attributed to CTP. However, the current pre-post design certainly provides strong evidence of the effectiveness of this training program and provides many advantages over post-test only designs, which has certainly been the norm in the field of evaluation research to date (Guerra-Lopez, 2008; Parker, 1976; Phillips & Stawarski, 2008).

References

- Ainley, M., Hidi, S., & Berndorff, D. (2002). Interest, learning and the psychological processes that mediate their relationship. *Journal of Educational Psychology, 94*, 545-561.
- Al-Athari, A., & Zairi, M. (2002). Training evaluation: An empirical study in Kuwait. *Journal of European Industrial Training, 26*, 241-251.
- Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's level of training criteria: Thirty years later. *Personnel Psychology, 42*, 331-342.
- Alliger, G. M., Tannebaum, S. I., Bennett, W., Traver, H., Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology, 50*, 341-358.
- Allison, P. D. (2001). Missing data. *Quantitative Applications in the Social Sciences* (Vol. 136). Philadelphia, PA : Sage Publications, Inc.
- Alvarez, K., Salas, E., & Garofano, C. M. (2004). An integrated model of training evaluation and effectiveness. *Human Resource Development Review, 3*, 385-416.
- American Society for Training and Development (ASTD). (1996). *The 1996 American Society for Training and Development report on trends that affect corporate learning and performance* (2nd ed.). Alexandria, VA: American Society for Training and Development.
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369-406.

- Antheil, J. H., & Casper, I. G. (1986). Comprehensive evaluation model: A tool for the evaluation of non-traditional training programs. *Innovative Higher Education, 11*, 55-64.
- Axtell, C. M., Maitlis, S., & Yearta, S. K. (1997). Predicting immediate and longer-term transfer of training. *Personnel Review, 26*, 201-213.
- Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology, 41*, 63-105.
- Bandura, A. (1977). *Social Learning Theory*. New York: General Learning Press.
- Bates, R. (2004). A critical analysis of evaluation practice: The Kirkpatrick model and the principle of beneficence. *Evaluation and Program Planning, 27*, 341-347.
- Bates, R., Holton, E. F., Seyler, D. L., & Carvalho, M. A. (2000). The role of interpersonal factors in the application of computer-based training in an industrial setting. *Human Resource Development International, 3*, 19-42.
- Baumgartel, H. J., Reynolds, M., & Pathan, R. (1984). How personality and organizational climate variables moderate the effectiveness of management development programs: A review and some recent research findings. *Management and Labour Studies, 9*, 1-16.
- Bensimon, P. (2005). *Correctional officer recruits during the college training period: An examination*. Ottawa: Correctional Service of Canada. Research Report R-165.
- Bhatti, M. A., & Kaur, S. (2010). The role of individual and training design factors on training transfer. *Journal of European Industrial Training, 34*, 656-672.

- Blanchard, P. N., Thacker, J. W., & Way, S. A. (2000). Training evaluation: Perspectives and evidence from Canada. *International Journal of Training and Development*, 4, 295-304.
- Bookter, A. (1999). *Convergent and divergent validity study of the learning transfer questionnaire*. Unpublished Doctoral Dissertation, Louisiana State University, Baton Rouge, LA.
- Brief, A. P., & Weiss, H. M. (2002). Organizational behavior: Affect in the workplace. *Annual Review of Psychology*, 53, 279-307.
- Broad, M. L., & Newstrom, J. W. (1992). *Transfer of training*. Reading, MA: Addison-Wesley.
- Brown, K. G. (2005). An examination of the structure and nomological network of trainee reactions: A closer look at "smile sheets". *Journal of Applied Psychology*, 90, 991-1001.
- Burke, L. A., & Baldwin, T. T. (1999). Workforce training transfer: A study of the effect of relapse prevention training and transfer climate. *Human Resource Management*, 38, 227-242.
- Burke, L. A., & Hutchins, H. M. (2007). Training transfer: An integrative literature review. *Human Resource Development Review*, 6, 263-296.
- Burnard, P. (1996). Teaching the analysis of textual data: An experiential approach. *Nurse Education Today*, 16, 287-281.
- Chandler, T. A. (1978). The questionable status of student evaluations of teaching. *Teaching of Psychology*, 5, 150-152.

- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293-332.
- Chen, H. (2003). *Cross-cultural construct validation of the learning transfer system inventory in Taiwan*. Unpublished Doctoral Dissertation, Louisiana State University, Louisiana.
- Chen, H. C., Holton, E. F., & Bates, R. (2005). Development and validation of the learning transfer system inventory in Taiwan. *Human Resource Development Quarterly, 16*, 55-84.
- Chen, G., Thomas, B., & Wallace, J. C. (2005). A multilevel examination of the relationships among training outcomes, mediating regulatory processes, and adaptive performance. *Journal of Applied Psychology, 90*, 827-841.
- Cheng, E. W. L., & Hampson, I. (2008). Transfer of training: A review and new insights. *International Journal of Management Reviews, 10*, 327-341.
- Clark, R. C., Nguyen, F., & Sweller, J. (2006). *Efficiency in learning: Evidence based guidelines to manage cognitive load*. San Francisco, CA: Pfeiffer.
- Clarke, N. (2002). Job/work environment factors influencing training transfer within a human service agency: Some indicative support for Baldwin and Ford's transfer climate construct. *International Journal of Training and Development, 6*, 146-162.
- Clement, R. W. (1982). Testing the hierarchy theory of training evaluation: An expanded role for trainee reactions. *Public Personnel Management Journal, 11*, 176-184.
- Coates, D. E. (2006). *People skills training: Are you getting a return on your investment?* Retrieved from <http://www.2020insight.net/PeopleSkills.htm>

- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Colquitt, J. A., LePine, J. A., & Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology, 85*, 678-707.
- Dixon, N. M. (1987). Meet training's goals without reaction forms. *Personnel Journal, 66*, 108-115.
- Elias, P. K., Elias, M. F., Robbins, M. A., & Gage, P. (1987). Acquisition of word-processing skills by younger, middle-age, and older adults. *Psychology of Aging, 2*, 340-348.
- Erez, A., & Isen, A. M. (2002). The influence of positive affect on the components of expectancy motivation. *Journal of Applied Psychology, 89*, 1055–1067.
- Estrada, C. A., Isen, A. M., & Young, M. J. (1997). Positive affect facilitates integration of information and decreases anchoring in reasoning among physicians. *Organizational Behavior and Human Decision Processes, 72*, 117–135.
- Fairchild, A. J., & Mackinnon, D. P. (2009). A general model for testing mediation and moderation effects. *Prevention Science, 10*, 87-99.
- Field, A. P. (2009). *Discovering Statistics Using SPSS (3rd Ed.)*. London: Sage.
- Facteau, J. D., Dobbins, G. H., Russell, J. E. A., Ladd, R. T., & Kudish, J. D. (1995). The influence of general perceptions of the training environment on pretraining motivation and perceived training transfer. *Journal of Management, 21*, 1-25.
- Fisher, S. L., & Ford, J. K. (1998). Differential effects of learner effort and goal orientation on two learning outcomes. *Personnel Psychology, 51*, 397-420.

- Fisher, R. J., & Katz, J. E. (2000). Social-desirability bias and the validity of self-reported values. *Psychology & Marketing, 17*, 105-120.
- Ford, J. K., Quinones, M. A., Segó, D. J., & Sorra, J. S. (1992). Factors affecting the opportunity to perform trained tasks on the job. *Personnel Psychology, 45*, 511-27.
- Foxon, M. (1997). The influence of motivation to transfer, action planning, and manager support on the transfer process. *Performance Improvement Quarterly, 10*, 42-63.
- Frazier, P. A., Tix, A. P., & Barron, K. E. (2004). Testing moderator and mediator effects in counseling psychology research. *Journal of Counseling Psychology, 51*, 115-134.
- Gagne, R. M., Baker, K. E., & Foster, H. (1950). Transfer of discrimination training to a mobile task. *Journal of Experimental Psychology, 40*, 314-328.
- Gaudine, A. P., & Saks, A. M. (2004). A longitudinal quasi-experiment on the effects of post-training transfer interventions. *Human Resource Development Quarterly, 15*, 57-76.
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.
- Graneheim, U. H., & Lundman, B. (2004). Qualitative content analysis in nursing research: Concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today, 24*, 105-112.
- Guerra-Lopez, I. J. (2008). *Performance evaluation: Proven approaches for improving program and organizational performance*. San Francisco, CA: Jossey-Bass.

- Hand, H. H., Richards, M. D., & Slocum, J. W. (1973). Organizational climate and the effectiveness of an human relations training programme. *Academy of Management Journal*, 16, 185-195.
- Hofstadter, D. R., & Dennett, D. C. (1981). *The mind's I: Fantasies and reflections on self and soul*. New York: Bantam Books.
- Holton, E. F. (1996). The flawed four-level evaluation model. *Human Resources Quarterly*, 7, 5-15.
- Holton, E. F. (2005). Holton's evaluation model: New evidence and construct elaborations. *Advances in Developing Human Resources*, 7, 37-54.
- Holton, E. F., Bates, R. A., & Ruona, W. E. A. (2000). Development of a generalized learning transfer system inventory. *Human Resource Development Quarterly*, 11, 333-360.
- Holton, E. F., Bates, R. A., Seyler, D. L., & Carvalho, M. B. (1997). Toward construct validation of a transfer climate instrument. *Human Resource Development Quarterly*, 8, 95-113.
- Holton, E. F., Chen, H., & Naquin, S. S. (2003). An examination of learning transfer system characteristics across organizational settings. *Human Resource Development Quarterly*, 14, 459-482.
- House, R. (1968). Leadership training: Some dysfunctional consequences. *Administrative Science Quarterly*, 12, 556-571.
- Holzbach, R. L. (1978). Rater bias in performance ratings: Superior, self, and peer ratings. *Journal of Applied Psychology*, 63, 579-588.

- Hsieh, H., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*, 1277-1288.
- Isen, A. M. (1993). Positive affect and decision making. In M. Lewis & J. Haviland-Jones (Eds.), *Handbook of Emotions* (pp. 261–277). NY: Guilford.
- Jaccard, J., & Turrisi, R. (2003). Interaction effects in multiple regression (2nd ed.). *Quantitative Applications in the Social Sciences* (Vol. 72). Philadelphia, PA: Sage Publications, Inc.
- Jackson, K. M., & Trochim, W. M. K. (2002). Concept mapping as an alternative approach for the analysis of open-ended survey responses. *Organizational Research Methods, 5*, 307-336.
- Jones, J. W., Steffy, B. D., & Bray, D. W. (1991). *Applying psychology in business*. New York, NY: Lexington Books.
- Juurink, D. N., & Detsky, A. S. (2005). Kappa statistic. *Canadian Medical Association Journal, 173*, 16.
- Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology, 74*, 657-690.
- Kaplan, R. M., & Pascoe, G. C. (1977). Humorous lectures and humorous examples: Some effects upon comprehension and retention. *Journal of Educational Psychology, 69*, 61-65.
- Khasawneh, S. A. (2004). *Construct validation of an Arabic version of the learning transfer system inventory for use in Jordan*. Unpublished Doctoral Dissertation, Louisiana State University, Louisiana.

- Kirkpatrick, D. L. (1959a). Techniques for evaluating training programs. *Journal of the ATSD, 13*, 3-9.
- Kirkpatrick, D. L. (1959b). Techniques for evaluating training programs: Part 2 – Learning. *Journal of the ATSD, 13*, 21-26.
- Kirkpatrick, D. L. (1960a). Techniques for evaluating training programs: Part 3 – Behavior. *Journal of the ATSD, 14*, 13-18.
- Kirkpatrick, D. L. (1960b). Techniques for evaluating training programs: Part 4 – Results. *Journal of the ATSD, 14*, 28-32.
- Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), *Training and development handbook: A guide to human resource development* (pp. 18-1-18-27). New York, NY: McGraw Hill.
- Kirkpatrick, D. L. (1994). *Evaluating training programs: The four levels*. San Francisco, CA: Berrett-Koehler.
- Kirkpatrick, D. L., & Kirkpatrick, J. D. (2005). *Transferring learning to behavior: Using the four levels to improve performance*. San Francisco, CA: Berrett-Koehler.
- Kontoghiorghes, C. (2004). Reconceptualizing the learning transfer conceptual framework: Empirical validation of a new systemic model. *International Journal of Training and Development, 8*, 210-221.
- Kozlowski, S. W. J., & Salas, E. (1997). An organizational systems approach for the implementation and transfer of training. In J. K. Ford & Associates (Eds.), *Improving training effectiveness in work organizations* (pp. 247-290). Hillsdale, NJ: LEA.

- Kraiger, K. (2002). Decision-based evaluation. In K. Kraiger (Ed.), *Creating, implementing, and managing effective training and development: State-of-the-art lessons for practice* (pp. 331-375). San Francisco, CA: Jossey-Bass.
- Kraiger, K., Ford, K. J., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78*, 311-328.
- Machles, D. (2003). Evaluating the effectiveness of safety training. *Occupational Health & Safety*. Retrieved on November 26, 2008, from <http://ohsonline.com/Articles/2003/06/Evaluating-the-Effectiveness-of-Safety-Training.aspx>.
- MacLennan, R. N. (1993). Interrater reliability with SPSS for Windows 5.0. *The American Statistician, 47*, 292-296.
- Mann, S. (1997). Implications of the response-shift bias for management. *Journal of Management Development, 16*, 328-336.
- Mathieu, J. E., Tannenbaum, S. I., & Salas, E. (1992). Influences of individual and situational characteristics on measures of training effectiveness. *Academy of Management Journal, 4*, 828-847.
- Mathieu, J. E., Martineau, J. W., & Tannenbaum, S. I. (1993). Individual and situational influences on the development of self-efficacy: Implications for training effectiveness. *Personnel Psychology, 46*, 125-147.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*, 30-46.

- Melvin, K. B., Gramling, L. K., & Gardner, W. M. (1985). A scale to measure attitudes toward prisoners. *Criminal Justice and Behavior*, *12*, 241-253.
- Morgan, R. B., & Casper, W. J. (2000). Examining the factor structure of participant reactions to training: A multidimensional approach. *Human Resource Development Quarterly*, *11*, 301-317.
- Noe, R. A. (1986). Trainee's attributes and attitudes: Neglected influences on training effectiveness. *Academy of Management Review*, *11*, 736-749.
- Noe, R. A., & Schmitt, N. (1986). The influence of trainee attitudes on training effectiveness: Test of a model. *Personnel Psychology*, *39*, 497-523.
- Noe, R. A., & Wilk, S. L. (1993). Investigation of the factors that influence employees' participation in development activities. *Journal of Applied Psychology*, *78*, 291-302.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.
- Ortet-Fabregat, G., Perez, J., & Lewis, R. (1993). Measuring attitudes toward prisoners. *Criminal Justice and Behavior*, *20*, 190-198.
- Parker, T. C. (1976). Statistical Methods for Measuring Training Results, in R. Craig (Ed.), *The Training and Development Handbook: A Guide to Human Resource Development* (2nd Ed.) (pp. 19-1-19-23). New York, NY: McGraw-Hill.
- Paulhaus, D.L., (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*, 598-609.
- Peters, L. H., O'Connor, E. J., & Eulberg, J. R. (1985). Situational constraints: Sources, consequences and future considerations, in K. M. Rowland and G. R. Ferris (Eds.),

- Research in personnel and human resource management* (Vol. 3) (pp. 79-114).
Greenwich CT: JAI Press.
- Phillips, J. J. (1996). Measuring the Results of Training, in R. Craig (Ed.), *The ASTD Training — Development Handbook* (pp. 314-341). New York, NY: McGraw-Hill.
- Phillips, J. J. (1997). *Handbook of training evaluation and measurement methods* (3rd Ed.). Houston, TX: Gulf Publishing Company.
- Phillips, P. P., & Phillips, J. J. (2007). *The value of learning: How organizations capture value and ROI and translate them into support, improvement and funds*. San Francisco, CA: John Wiley and Sons, Inc.
- Phillips, J. J., & Stone, R. D. (2002). *How to measure training results: A practical guide to tracking the six key indicators*. New York, NY: McGraw Hill.
- Phillips, P. P. & Stawarski, C. A. (2008). *Data Collection: Planning for and Collecting all Types of Data*. San Francisco, CA: Pfeiffer.
- Preacher, K. J., Curran, P. J., & Bauer, D. J. (2012). Simple intercepts, simple slopes, and regions of significance in MLR 2-way interactions. Retrieved January 8, 2012 from <http://www.psych.ku.edu/preacher/interact/mlr2.htm>
- Ramesh, G., & Ramesh, M. (2010). *The ACE of soft skills: Attitude, communication and etiquette for success*. New Delhi, India: Pearson Education.
- Ree, M. J., & Earles, J. A. (1991). Predicting training success: Not much more than g. *Personnel Psychology*, 44, 321-332.
- Richman-Hirsch, W. L. (2001). Posttraining interventions to enhance transfer: The moderating effects of work environments. *Human Resource Development Quarterly*, 12, 105-120.

- Robinson, D., Porporino, F., & Simourd, L. (1992). *Staff commitment in the correctional service of Canada*. Ottawa, Canada: Canada Correctional Service. NCJ Document Number: 148402.
- Robinson, D., Porporino, F. J., & Simourd, L. (1993). The influence of career orientation on support for rehabilitation among correctional staff. *The Prison Journal*, 73, 162-177.
- Robinson, D., Porporino, F. J., & Simourd, L. (1996). Do different occupational groups vary on attitudes and work adjustment in corrections? *Federal Probation*, 60, 45-53.
- Rouiller, J. Z., & Goldstein, I. L. (1993). The relationship between organizational transfer climate and positive transfer of training. *Human Resources Development Quarterly*, 4, 377-390.
- Ruona, W. E. A., Leimbach, M., Holton, E. F., & Bates, R. (2002). The relationship between learner utility reactions and predicted learning transfer among trainees. *International Journal of Training and Development*, 6, 218-228.
- Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology*, 52, 471-499.
- Seyler, D. L., Holton, E. F., Bates, R. A., Burnett, M. F., & Carvalho, M. A. (1998). Factors affecting motivation to use training. *International Journal of Training and Development*, 2, 2-16.
- Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability,

Psychological Bulletin, 86, 420-428.

Simourd, L. (1997). *Staff attitudes towards inmates and correctional work: An exploration of the attitude-work outcome relationship*. Unpublished document.

Carleton University.

Spector, P. (1982). Behavior in organizations as a function of employee's locus of control. *Psychological Bulletin*, 91, 482-497.

Stufflebeam, D. L. (1973). *Toward a science of educational evaluation*. Englewood Cliffs, NJ: Educational Technology Publications.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th Ed). Boston: Pearson.

Tan, J. A., Hall, R. J., & Boyce, C. (2003). The role of employee reactions in predicting training effectiveness. *Human Resource Development Quarterly*, 14, 397-411.

Tannenbaum, S. I., & Yukl, G. (1992). Training and development in work organizations. *Annual Review of Psychology*, 43, 399-341.

Tannenbaum, S. I., Cannon-Bowers, J. A., Salas, E., & Mathieu, J. E. (1993). Factors that influence training effectiveness: A conceptual model and longitudinal analysis (NAWCTSD Tech. Rep. No. 93-011). Orlando, FL: Naval Air Warfare Center.

Thayer, P. W., & Teachout, M. S. (1995). *A climate for transfer model* (Interim Rep. No. ADA3170570). Houston, TX: Brooks Air Force Base.

Thorndike, E. L. (1935). *The psychology of wants, interests, and attitudes*. New York, NY: Appleton-Century.

- Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review*, 8, 247-261.
- Tracey, J. B., Tannenbaum, S. I., & Kavanagh, M. J. (1995). Applying trained skills on the job: The importance of the work environment. *Journal of Applied Psychology*, 80, 239-252.
- Underwood, B. J. (1951). Associative transfer in verbal learning as a function of response similarity and degree of first-line learning. *Journal of Experimental Psychology*, 42, 44-53.
- Van Buren, M E., & Erskine, W. (2003). *The ASTD state of the industry report*. Alexandria, VA: American Society for Training and Development.
- Van der Klink, M. R., Gielen, E., & Nauta, C. (2001). Supervisory support as a major condition to enhance transfer. *International Journal of Training and Development*, 5, 52-63.
- van Merriënboer, J. (1997). *Training complex cognitive skills: A four component instructional design model for technical training*. Englewood Cliff, NJ: Educational Technology Publications.
- Velada, R., Caetano, A., Michel, J. W., Lyons, B. D., & Kavanagh M. J. (2007). The effects of training design, individual characteristics and work environment on transfer of training. *International Journal of Training and Development*, 11, 282-294.
- Wagner, R. K. (1987). Tacit knowledge in everyday intelligent behavior. *Journal of Personality and Social Psychology*, 52, 1236–1247.

- Warr, P., Allan, C., & Birdi, K. (1999). Predicting three levels of training outcome. *Journal of Occupational and Organizational Psychology, 72*, 351-375.
- Warr, P., & Bunce, D. (1995). Trainee characteristics and the outcomes of open learning. *Personnel Psychology, 48*, 347-376.
- Warr, P., Bird, M., & Rackham, N. (1970). *Evaluation of management training*. London, England: Gower Press.
- Werner, J. M., & DeSimone R. L. (2012). *Human resource development* (6th ed.). Mason, OH: Cengage/South-Western.
- Wexley, K. N., & Baldwin, T. T. (1986). Posttraining strategies for facilitating positive transfer: An empirical exploration. *Academy of Management Journal, 29*, 503-520.
- Xiao, J. (1996). The relationship between organizational factors and the transfer of training in the electronics industry in Shenzhen, China. *Human Resource Development Quarterly 7*, 55-73.
- Yamhill, S. (2001). *Factors affecting transfer of training in Thailand*. Unpublished Doctoral Dissertation, University of Minnesota, Minnesota.
- Zedeck, S. (1971). Problems with the use of "moderator" variables. *Psychological Bulletin, 76*, 295-310.
- Zerbe, W. J., & Paulhus, D.L. (1987). Socially desirable responding in organizational behavior: A reconception. *Academy of Management Review, 12*, 250-264.

Appendix A

Overview of courses and subject areas covered in Stage 1 including an indication of the number of formal knowledge-based evaluations.

Course	Subject Area	Number of Knowledge-based Evaluations
1	Corrections in Canada	-
2	Law and Policy	1
3	Officer Roles and Responsibilities	-
4	CAPRA and Problem Based Learning	-
5	Workplace Issues	-
6	Suicide Prevention and Self Injury	1
7	Substance Abuse	1
8	WHMIS	1
9	Ethical Issues	-
10	The Offender's World	1
-	Stage 1 Theory Test	1

Appendix B

Overview of the courses and subject areas covered in Stage 3 including an indication of the number of formal knowledge and skills-based evaluations.

Course	Subject Area	Number of Knowledge-based Evaluations	Number of Skills-based Evaluations
2 ^a	Law and Policy	1	-
3	Static and Dynamic Security	-	-
4	Perimeter Security	-	-
5	Counts and Offender Movement	-	-
6	Key and Tool Control	-	-
7	Acquiring and Analyzing Information	-	-
8	Report Writing	-	5
9	Communication and Conflict Management Skills	-	-
10	Administrative Segregation	-	-
11	Sudden In-Custody Death Syndrome (SICDS)	-	-
12	Identification Cards and Passes	-	-
13	Preservation of Evidence	-	-
14	Communication Equipment	-	-
15	Video Camera Operation	-	-
16	First Officer on the Scene	-	-
17	Responding to Suicide and Self Injury Situations	-	-
18	Integration Session A	-	-
19	Introduction to Searching	-	-
20	Introduction to Restraint Equipment	-	-
21	Chemical and Inflammatory Agents	-	-
22	Practice: Searching and Restraint Equipment	-	-
23	Integration Session B	-	-
24	Offender Discipline	-	-
25	Appearing as a Witness	-	-
26	Practical Test: Searching and Restraint Equipment	-	8
27	Hostage Taking and Forcible Confinement	-	-
28	Escorts	-	2
29	Practical Test: Communication Equipment and Chemical and Inflammatory Agents including Gas Mask	-	7
30	Posts – Armed Tower and Internal Armed and Non-armed, Checkpoints	-	-
31	Posts – Mobile, Principle and Vehicle Entrance	-	-
32	Final Security Test	1	-
33	Integration Session C	-	-

34	Fire Safety Training	1	3
35	Firearms	4	8
36	Self Defence	1	1
37	Arrest and Control and Baton	1	2
38	Sidearm Retention	-	-
39	Closing Administration	-	-

Note. ^aStage 3 begins with session #2 because Stage 2 is considered to comprise session #1.

Appendix C

A List of Topics Included in the Stage 4 Institutional Orientation Checklists

- 1) Inoculations
- 2) Uniform sizing & issue
- 3) Staffing documentation
- 4) X-ray Machine operation
- 5) Walk through scanner
- 6) Inmate counts
- 7) Take over armed posts
- 8) Communication Equipment
- 9) Vehicle Entrance
- 10) Security and Fire Walks
- 11) Segregation
- 12) Key control
- 13) Inmate Movement Control
- 14) External Security Escort
- 15) Perimeter Security
- 16) Principle Entrance
- 17) Offender Discipline
- 18) Personal Safety

Appendix D

Participation Rates per Region and CTP Cohort for Each Phase of the Study

Region / CTP	Recruitment Sample Size (<i>n</i>)	Participation Rate (Recruitment Phase) (<i>n</i>)	Participation Rate (Training Phase) (<i>n</i>)	Participation Rate (Transfer Phase) (<i>n</i>)
	(<i>N</i> = 102)	(<i>N</i> = 102)	(<i>N</i> = 68)	(<i>N</i> = 53)
Pacific	37	37	34	26
CTP 1 ^a	7	7	6	6
CTP 5	8	8	7	5
CTP 7	9	9	8	3
CTP 9	9	9	9	8
CTP 10	4	4	4	4
Ontario	51	51	27	24
CTP 10	3	3	2	2
CTP 11	9	9	2	2
CTP 12	10	10	3	3
CTP 13	5	5	0	-
CTP 15	6	6	6	6
CTP 16	5	5	5	3
CTP 17	5	5	3	2
CTP 18	7	7	6	6
Quebec	14	14	7	3
CTP 108	8	8	5	3
CTP 109	2	2	2	0
CTP 110	4	4	0	-

Note. ^aEach region determines the CTP numbers, which are utilized for tracking along with the year of delivery.

Appendix E

Demographic Data for the Training and Transfer Phases

Training Phase

	All Regions % (<i>n</i>)	Pacific % (<i>n</i>)	Ontario % (<i>n</i>)	Quebec % (<i>n</i>)
	<i>N</i> = 68	<i>n</i> = 34	<i>n</i> = 27	<i>n</i> = 7
Gender				
Female	32 (22)	38 (13)	33 (9)	-
Male	68 (46)	62 (21)	67 (18)	7 (100)
Ethnicity^a				
Caucasian	75 (51)	77 (26)	70 (19)	6 (86)
Asian	7 (5)	6 (2)	11 (3)	-
African	3 (2)	6 (2)	-	-
Aboriginal	4 (3)	6 (2)	37 (1)	-
European	3 (2)	6 (2)	-	-
Age				
20-25 years	28 (19)	29 (10)	30 (8)	14 (1)
26-30 years	29 (20)	21 (7)	37 (10)	43 (3)
31-35 years	16 (11)	18 (6)	11 (3)	29 (2)
36-40 years	7 (5)	12 (4)	-	14 (1)
> 40 years	19 (13)	21 (7)	22 (6)	-
Marital Status				
Single	34 (23)	38 (13)	26 (7)	71 (3)
Married/common law	59 (40)	50 (17)	70 (19)	29 (4)
Separated/divorced	7 (5)	12 (4)	4 (1)	-

	All Regions % (n)	Pacific % (n)	Ontario % (n)	Quebec % (n)
	N = 68	n = 34	n = 27	n = 7
Previous Work Experience				
No experience	34 (23)	41 (14)	15 (4)	71 (5)
<5 years related experience	47 (32)	41 (14)	59 (16)	29 (2)
5-10 years related experience	16 (11)	18 (6)	19 (5)	-
>10 years related experience	3 (2)	-	7 (2)	-
Previous Volunteer Experience^b				
No experience	65 (44)	68 (23)	52 (14)	100 (7)
<5 years related experience	28 (19)	32 (11)	30 (8)	-
5-10 years related experience	3 (2)	-	7 (2)	-
Educational Background^c				
High School	31 (21)	35 (12)	19 (5)	57 (4)
Unrelated College Diploma	10 (7)	12 (4)	7 (2)	14 (1)
Related College Diploma	31 (21)	21 (7)	48 (13)	14 (1)
Unrelated University Degree	12 (8)	18 (6)	4 (1)	14 (1)
Related University Degree	10 (7)	15 (5)	7 (2)	-

Note. ^an = 5 missing (all regions), n = 5 missing (Ontario), n = 1 missing (Quebec). ^bn = 3 missing (all regions & Ontario). ^cn = 4 missing (all regions & Ontario).

Transfer Phase

	All Regions % (n)	Pacific % (n)	Ontario % (n)	Quebec % (n)
	<i>N</i> = 53	<i>n</i> = 26	<i>n</i> = 24	<i>n</i> = 3
Gender				
Female	17 (32)	39 (10)	29 (7)	-
Male	36 (68)	62 (16)	71 (17)	100 (3)
Ethnicity^a				
Caucasian	72 (38)	73 (19)	67 (16)	100 (3)
Asian	9 (5)	8 (2)	13 (3)	-
African	2 (1)	4 (1)	-	-
Aboriginal	6 (3)	8 (2)	4 (1)	-
European	4 (2)	8 (2)	-	-
Age				
20-25 years	28 (15)	31 (8)	29 (7)	-
26-30 years	32 (17)	27 (7)	38 (9)	33 (1)
31-35 years	15 (8)	15 (4)	13 (3)	33 (1)
36-40 years	8 (4)	12 (3)	-	33 (1)
> 40 years	17 (9)	25 (4)	21 (5)	-
Marital Status				
Single	30 (16)	39 (10)	21 (5)	33 (1)
Married/common law	62 (33)	50 (13)	75 (18)	68 (2)
Separated/divorced	8 (4)	12 (3)	4 (1)	-
Previous Work Experience				
No experience	28 (15)	39 (10)	13 (3)	67 (2)
<5 years related experience	53 (28)	46 (12)	63 (15)	33 (1)
5-10 years related experience	15 (8)	15 (4)	17 (4)	-
>10 years related experience	4 (2)	-	8 (2)	-

	All Regions % (n)	Pacific % (n)	Ontario % (n)	Quebec % (n)
	N = 53	n = 26	n = 24	n = 3
Previous Volunteer Experience^b				
No experience	62 (33)	65 (17)	54 (13)	100 (3)
<5 years related experience	30 (16)	35 (9)	29 (7)	-
5-10 years related experience	2 (1)	-	4 (1)	-
Educational Background^c				
High School	28 (15)	35 (9)	21 (5)	33 (1)
Unrelated College Diploma	13 (7)	15 (4)	8 (2)	33 (1)
Related College Diploma	28 (15)	15 (4)	42 (10)	33 (1)
Unrelated University Degree	9 (5)	15 (4)	4 (1)	-
Related University Degree	13 (7)	19 (5)	8 (2)	-

Note. ^an = 4 missing (all regions & Ontario). ^bn = 3 missing (all regions & Ontario). ^cn = 4 missing (all regions & Ontario).

Appendix F

INFORMED CONSENT FORM
(For the Trainees of CTP)

The purpose of an informed consent is to ensure that you understand the purpose of the study and the nature of your involvement. The informed consent must provide sufficient information such that you have the opportunity to determine whether you wish to participate in the study.

Research Title: An Evaluation of the Correctional Officer Training Program (CTP)

Research Personnel: The following people, from Carleton University, are involved in this research project and may be contacted for further information: Shevaun Corey (Principal Investigator, 613-520-2600x1728, scorey@connect.carleton.ca) and Dr. Craig Bennell (Faculty Supervisor, 613-520-2600x1769, cbennell@connect.carleton.ca). If you have any ethical concerns about this study, please contact Dr. Avi Parush (Chair, Ethics Committee for Psychological Research, 613-520-2600x6026, avi_parush@carleton.ca). Should you have any other concerns about the study, please contact Dr. Janet Mantler (Chair, Department of Psychology, 613-520-2600x4173, psychchair@carleton.ca).

Purpose: The purpose of this study is to determine the effectiveness of CTP.

Task Requirements and Duration: Should you choose to participate in this study, you will:

1. Be asked to participate in a 1-hour session during the recruitment phase of CTP wherein you will be asked to complete a demographic questionnaire, a pre-test assessing your knowledge of CTP-related issues, and a survey assessing your attitudes toward correctional work and inmates.
2. Allow Correctional Service of Canada (CSC) to provide us with your scores on the Skills and Abilities test you'll complete during the recruitment phase and all course evaluations as well as all formal theory and skills-based tests that you take throughout CTP.
3. Be asked to participate, immediately after the completion of CTP, in a 1.5-hour session wherein you will complete a post-test assessing your CTP-related knowledge and skills, and a survey assessing your attitudes toward correctional work and inmates.
4. Be asked to complete the same information as above (#3) as well as a survey about factors that may or may not influence the effectiveness of CTP, 4-months after the completion of CTP, but only if you are hired by a CSC institution. In addition, if hired by

a CSC institution, you will be asked to complete a survey pertaining to your on-the-job behaviour.

5. Allow your supervisor to complete a survey pertaining to your on-the-job behaviour 4-months after the completion of CTP, if you are hired by a CSC institution.

6. Allow CSC to provide us with written reports completed by you, but only if hired by a CSC institution. These reports will come from the Offender Management System, and will provide us with general performance information (e.g., the quality of the reports).

Location: The different phases of the study will take place over many months in different locations. The first phase of the study (#1) and the first follow-up session (#3) will take place at the Staff Training College. The second follow-up session (#4) will take place in the CSC institution you are employed in.

Potential Risk/Discomfort: There are some aspects of this study that may cause anxiety and/or embarrassment, especially in the pre-training phase. For example, you will be tested on matters that you may have little knowledge of during the pre-training session. We would like to stress that this part of the study is purely for research purposes to determine your baseline level of knowledge before participating in CTP and will in no way affect your standing in CTP. In addition, apart from the formal theory and skills-based tests that you will undertake as part of CTP, your responses on the other evaluation instruments (including the demographic questionnaire and the correctional work and inmates survey) will have no bearing on your progress in CTP, nor will it impact whether or not you are given a job offer with CSC. Furthermore, the on-the-job surveys that you and your supervisor complete will in no way influence your standing within the institution where you work. We would like to remind you that your participation in this study is entirely voluntary. You may choose not to answer any questions and you may withdraw from the study at any time without penalty.

Anonymity/Confidentiality: The data gathered in this study are completely confidential. While your name will appear on the forms that you complete, this is simply to ensure that the Principal Researcher can match all of your data (e.g., your pre-test scores with your post-test scores, etc.). This consent form will be kept separate from your responses to ensure confidentiality and none of the recruiters, trainers, researchers, or institutional staff employed by CSC will have access to identifiable raw data collected for the purpose of this study, although aggregate data will be used by CSC to make improvements to CTP. The data will also be used for teaching and research purposes, but it will only be presented in aggregate form (i.e., no individual's data will be identified).

Signatures: I have read the above form and understand the conditions of my participation. My participation in this study is voluntary, and if for any reason, at any time, I wish to stop participating in the study I may do so without having to give an explanation. Furthermore, I am also aware that the data gathered in this study are confidential.

My signature indicates I agree to participate in this study.

Participant's Name: _____

Participant's Signature: _____

Witnesses Name: _____

Witnesses Signature: _____

Date: _____

Appendix G

INFORMED CONSENT FORM
(For the Supervisors)

The purpose of an informed consent is to ensure that you understand the purpose of the study and the nature of your involvement. The informed consent must provide sufficient information such that you have the opportunity to determine whether you wish to participate in the study.

Research Title: An Evaluation of the Correctional Officer Training Program (CTP)

Research Personnel: The following people, from Carleton University, are involved in this research project and may be contacted for further information: Shevaun Corey (Principal Investigator, 613-520-2600x1728,) and Dr. Craig Bennell (Faculty Supervisor, 613-520-2600x1769, cbennell@connect.carleton.ca). If you have any ethical concerns about this study, please contact Dr. Avi Parush (Chair, Ethics Committee for Psychological Research, 613-520-2600x6026, avi_parush@carleton.ca). Should you have any other concerns about the study, please contact Dr. Janet Mantler (Chair, Department of Psychology, 613-520-2600x4173, psychchair@carleton.ca).

Purpose: The purpose of this study is to determine the degree to which behaviours and skills gained during CTP are used on-the-job by trainees.

Task Requirements and Duration: Should you choose to participate in this study, you will be asked to complete a survey assessing on-the-job behaviours displayed by employees that you supervise who have recently graduated from CTP 2008. The survey will take approximately 15 minutes to complete.

Location: The survey will be emailed to you and should be emailed back to the Principal Researcher upon completion.

Potential Risk/Discomfort: There are some aspects of the study that may cause anxiety because you will be asked to evaluate another employee. These evaluations, however, will not be viewed by any other institutional staff, nor will they be used for any job-related evaluations. I would like to remind you that your participation is entirely voluntary. You may choose not to answer any question and you may withdraw from the study at any time without penalty.

Anonymity/Confidentiality: The data gathered in this study are confidential. The name of the trainee that you are evaluating will appear on the survey that you complete, but this is simply to ensure that the Principal Researcher can match all of your data to that of the correctional officer. This consent form will be kept separate from your responses to

ensure confidentiality and none of the recruiters, trainers, researchers, or institutional staff employed by CSC will have access to identifiable raw data collected for the purpose of this study, although aggregate data will be used by CSC to make improvements to CTP. The data will also be used for teaching and research purposes, but it will only be presented in aggregate form (i.e., no individual's data will be identified).

Signatures: I have read the above form and understand the conditions of my participation. My participation in this study is voluntary, and if for any reason, at any time, I wish to stop participating in the study I may do so without having to give an explanation. Furthermore, I am also aware that the data gathered in this study are confidential.

My signature indicates I agree to participate in this study.

Participant's Name: _____

Participant's Signature: _____

Witnesses Name: _____

Witnesses Signature: _____

Date: _____

Appendix H

Demographic Questionnaire

Please answer the following questions about yourself.

1. Gender:

- Female
- Male

2. Ethnicity: _____

3. Age at time of training:

- < 20 years
- 20-25 years
- 26-30 years
- 31-35 years
- 36-40 years
- > 40 years

4. Marital status at time of training:

- single
- married/common law
- separated/divorced

5. Amount and type of previous work experience at time of training (related means previous experience working in a criminal justice field):

- no experience
- under 5 years related experience
- 5-10 years related experience
- over 10 years related experience

6. If you indicated related work experience, please specify what that experience was and who it was for (e.g., police officer, RCMP):

7. Amount and type of previous volunteer experience at time of training (related means previous experience in a criminal justice field):

- no experience
- under 5 years related experience
- 5-10 years related experience
- over 10 years related experience

8. If you indicated related volunteer experience, please specify what that experience was and who it was for (e.g., police officer, RCMP):

9. Educational background at time of training (related means security, corrections, criminal justice, or policing):

- high school
- unrelated college diploma
- related college diploma
- unrelated university degree
- related university degree

Appendix I

Completion Rates of all Evaluation Tools by Participants Overall (N = 102) and per Region, Presented for each Phase of the Study

	All Regions % (N)	Pacific % (n)	Ontario % (n)	Quebec % (n)
Recruitment Phase				
Pre-Test (including Demographic Questionnaire, Knowledge Test & Attitudes Surveys – ATCW & ATI)	100 (102)	100 (37)	100 (50)	100 (14)
Soft Skills Assessment	100 (102)	100 (37)	100 (50)	100 (14)
Training Phase				
Trainee Reaction Questionnaire	56 (57)	84 (31)	52 (26)	-
Knowledge Test	80 (82)	92 (34)	76 (38)	71 (10)
Soft Skills Assessment	54 (55)	73 (27)	42 (21)	50 (7)
Hard Skills	67 (68)	92 (34)	54 (27)	50 (7)
Attitudes Surveys (ATCW & ATI)	67 (68)	92 (34)	54 (27)	50 (7)
Transfer Phase				
Knowledge Test	52 (53)	70 (26)	48 (24)	21 (3)
Soft Skills Assessment	50 (51)	68 (25)	46 (23)	21 (3)
Hard Skills	51 (52)	68 (25)	48 (24)	21 (3)
Attitudes Surveys (ATCW & ATI)	52 (53)	70 (26)	48 (24)	21 (3)
Trainee Behavioural Checklist	52 (53)	70 (26)	48 (24)	21 (3)
LTSI	52 (53)	70 (26)	48 (24)	21 (3)
Supervisor Behavioural Checklist	25 (25)	43 (16)	18 (9)	-

Appendix J

Level 1 Reaction Questionnaire Given to Trainees at the End of Stage 3

Evaluation of CTP 2008

Thinking about the stages of CTP that you have completed, please indicate how you feel about each of the statements below. Unless otherwise indicated, think about the training program as a whole when responding. Please indicate your response (circle a number) according to the following scale:

	1 Strongly Disagree	2 Disagree	3 Neither	4 Agree	5 Strongly Agree	6 Not Applicable (N/A)	
Training Program Materials							
	Strongly	Disagree	Disagree	Neither	Agree	Strongly Agree	N/A
1) I felt the training material was interesting. *	1	2	3	4	5	6	
2) I felt motivated by the training material. *	1	2	3	4	5	6	
3) I felt it was easy to follow the material contained in CTP. ***	1	2	3	4	5	6	
4) I felt that the training material was easy to understand. ***	1	2	3	4	5	6	
5) I did not feel overwhelmed by the training material. ***	1	2	3	4	5	6	
6) Overall, I felt that the training content was not too difficult to learn. ***	1	2	3	4	5	6	
7) The training material provided in CTP prepared me to be successful on any formal evaluations. **	1	2	3	4	5	6	
8) I felt that the training material provided me with knowledge and skills that will be useful on the job. **	1	2	3	4	5	6	

9) I feel that the training material has real practical value. **	1	2	3	4	5	6
Training Program Overall	Strongly Disagree	Disagree	Neither	Agree	Strongly Agree	N/A
10) Overall, I think the Correctional Officer Training Program (CTP) met the stated objectives. *	1	2	3	4	5	6
11) The training program was well structured. *	1	2	3	4	5	6
12) The training program flowed logically. *	1	2	3	4	5	6
13) In general, the evaluation methods were fair. *	1	2	3	4	5	6
14) I felt the evaluations effectively tested my knowledge of the training material. *	1	2	3	4	5	6
15) Overall, I felt that the tests were not too difficult. ***	1	2	3	4	5	6
16) The different types of learning activities (e.g., online modules, classroom-based learning, practical applications, etc.) enhanced my comprehension of key content areas. **	1	2	3	4	5	6
17) I felt able to self evaluate my learning through the different learning activities. *	1	2	3	4	5	6
18) I felt that the pace of the training program was just right. *	1	2	3	4	5	6
19) I feel that the quality of my work on the job will improve as a result of completing this training program. **	1	2	3	4	5	6
20) I feel motivated to use what I have learned in this training program in the future, when	1	2	3	4	5	6

I am on the job. **

Stages of CTP	Strongly Disagree	Disagree	Neither	Agree	Strongly Agree	N/A
21) In general, I am satisfied with the following elements of Stage 1 of CTP (the online component):						
Length *	1	2	3	4	5	6
Quality *	1	2	3	4	5	6
Content *	1	2	3	4	5	6
Time I Invested *	1	2	3	4	5	6
22) As a whole, the online component (Stage 1) was a valuable part of CTP. **	1	2	3	4	5	6
23) In general, I am satisfied with the following elements of Stage 3 of CTP:						
Length *	1	2	3	4	5	6
Quality *	1	2	3	4	5	6
Content *	1	2	3	4	5	6
Time I Invested *	1	2	3	4	5	6
24) As a whole, the theory-based classroom component of Stage 3 was a valuable part of CTP. **	1	2	3	4	5	6
25) As a whole, the practical application component of Stage 3 was a valuable part of CTP. **	1	2	3	4	5	6
26) I felt that I was able to link the different topic areas throughout the different Stages within CTP. *	1	2	3	4	5	6
27) The estimated time allotted to complete Stage 1 of CTP was realistic. *	1	2	3	4	5	6
28) The estimated time allotted to complete Stage 2 of CTP was realistic. *	1	2	3	4	5	6
29) The estimated time allotted to complete Stage 3 of CTP was realistic. *	1	2	3	4	5	6

Facilitator	Strongly Disagree	Disagree	Neither	Agree	Strongly Agree	N/A
30) In general, facilitators' were well prepared and organized. *	1	2	3	4	5	6
31) Overall, the facilitators' communicated in a clear and concise manner (e.g., when answering questions, providing direction/instruction, etc.). *	1	2	3	4	5	6
32) In general, I felt that the facilitators' possessed a lot of practical knowledge. **	1	2	3	4	5	6
33) In general, facilitators' created and maintained a positive learning environment. *	1	2	3	4	5	6
34) In general, the facilitators' served as positive and professional role models. *	1	2	3	4	5	6
35) Overall, I felt that the facilitators' provided feedback that was timely, specific and helped me identify areas in need of development. **	1	2	3	4	5	6

Note. * = general satisfaction with training; ** = satisfaction with training utility; *** = satisfaction with training difficulty

Open Ended Questions

1. Please indicate any issues or concerns you may have had with each Stage of CTP as well as CTP Overall and offer any suggestions you may have to remedy these concerns in the spaces provided below.

Stage 1 (Online)

Issues/Concerns:

Suggestions for Improvement:

Stage 2

Issues/Concerns:

Suggestions for Improvement:

Stage 3

Issues/Concerns:

Suggestions for Improvement:

CTP Overall:

Issues/Concerns:

Suggestions for Improvement:

Any other Comments:

Appendix K

Hard Skills Selection Questionnaire

Input on the Assessment of Hard Skills Acquisition and Retention (CTP 2008)

L&D (NHQ) in conjunction with Carleton University is conducting a large-scale evaluation of CTP 2008. As part of this evaluation we would like to include an assessment of the degree to which recruits acquire the **hard skills** taught in Stage 3 and the degree to which they retain them after being employed 4 months on the job. We would like to ask for your help in developing this evaluation tool.

This evaluation tool will be administered to the recruits at the staff colleges after they complete Stage 3 of CTP 2008 and again 4 months post-CTP 2008, either at the institution or at the staff colleges. This tool will be one of several others that encompass the toolkit for the evaluation of CTP 2008 (i.e., several surveys, the Personal Skills and Suitability test and a knowledge test). The time allotted for the administration of the entire evaluation toolkit is 2.5-3 hours per recruit and the time allotted specifically for the hard skills evaluation tool will be 45 minutes maximum. As part of the National Evaluation Team you will be administering the toolkit at these two time periods. It is estimated that there could be up to 200 recruits that need to be evaluated after Stage 3, however, it is expected that there will be approximately 60 recruits that will need to be assessed at 4 months post-CTP 2008.

The following is an outline of the process for developing the hard skills evaluation tool:

1. The first step in this process will be to identify which evaluation techniques (i.e., hard skills taught during CTP 2008) should be included in the evaluation of CTP 2008, taking into consideration the timeframe and context in which the recruits will be tested. Specifically, the recruits will be tested after working on the job for only 4 months, and therefore, it will be important to identify the skills that they will be utilizing most commonly within this time period. This survey will be used to narrow down the options and should be viewed as a first cut at the skills that will be included in the evaluation. Below you will be asked to identify, in your opinion, 20 of the most common and most important skills to be assessed.
2. Once this is done and the responses are collected from all members of the National Evaluation Team we will meet in person or via video/teleconference to narrow down the list even further. The actual number will be determined based on your responses to this survey.
3. Lastly, after the skills have been selected the evaluation tool will be designed by Learning & Development (NHQ) and then the team will reconvene for a training session to validate the instrument and to determine the level agreement when rating the skills.

Could you please answer the following questions and then send your feedback to Shevaun Corey (Project Officer, L&D, NHQ) at coreysa@csc-scc.gc.ca. It would be greatly appreciated if you could submit your feedback on or before August 28th, 2009.

Thank you very much in advance for your time.

Sincerely,
Shevaun Corey

1. How many skills could we realistically evaluate in the timeframe given for the evaluation of CTP 2008 (i.e., how many skills will they have had the opportunity to use 4 months on the job)?

2. Given that this evaluation will be conducted 4 months on the job, which of the following skills do you feel are the most important? Please check off 20 skills.

Self defence:

- Shuffle Pivot
- Progressive Pivot
- Side Step
- Breakfall (Forward, Backward, Side)
- Straight Arm Bar
- Wrist Throw
- Roundhouse Punch Defence
- Front Choke
- Body Pivot
- Neck Nerve/Knee Lift Takedown
- Nasal Nerve Peel Release
- Leg Trap
- Garrotte Strangulation
- Hair Pull or Shoulder Grab
- Arm Bar
- Knife Attack Situation
- Club Attack Situation

Baton Skills:

- First 3-Count (Poke, Overhand, & Underhand Thrusts)
- Second 3-Count (Forearm Smash, Overhand & Underhand Thrust)
- Third 3-Count (Forearm Smash, Shin & Leg Strike)

- 5-Count Thrust (Poke, Poke, Forearm Smash, Overhand & Underhand)
- Offensive Turn
- Defensive Turn
- 2-Count Forearm Slash
- 2-Count Power Slash
- 2-Count Leg Sweep

Arrest & Control:

- Twistlock Control Handcuffing Technique
- Prone Control Handcuffing Technique
- Low Profile Rear Wristlock Escort
- High Profile Rear Wristlock Escort
- Shuffle Pivot to Outside Deflection
- Progressive Pivot to Outside Deflection
- Shuffle Side Step to Outside Deflection
- Progressive Side Step to Outside Deflection
- Basic Straight Arm Bar
- Straight Arm Bar and Wristlock
- Elevated Arm Bar and Wristlock 2
- Standing Aikido Wristlock
- Kneeling Aikido Wristlock
- Prone Control
- Hammerlock Spin to Prone Control
- Front Wristlock
- Thumblock Front Wristlock
- Twistlock
- High Profile Rear Wristlock
- Assaultive Subject - Wrist Throw Takedown
- Assaultive Subject - Neck Nerve/Cross Face/Hair Pull Takedown
- Assaultive Subject - Arm Bar Takedown

Other Skills:

- Pat Down Frisk Search
- Strip Search
- MK 3 & 4 OC Projector
- MK 9 OC Projector
- Gas Mask
- Grenade
- Application of Leg Irons
- Removal of Leg Irons
- Application of Transport Chain
- Removal of Transport Chain
- Two-Way Radio
- Loudhailer

- Other, please specify:

3. Given that this evaluation will be conducted 4 months on the job, which of the following skills do you feel are the most commonly used? Please check off 20 skills. If you feel they are the same as those checked above for Question #2 – please write same as above in the box below.

Self defence:

- Shuffle Pivot
- Progressive Pivot
- Side Step
- Breakfall (Forward, Backward, Side)
- Straight Arm Bar
- Wrist Throw
- Roundhouse Punch Defence
- Front Choke
- Body Pivot
- Neck Nerve/Knee Lift Takedown
- Nasal Nerve Peel Release
- Leg Trap
- Garrote Strangulation
- Hair Pull or Shoulder Grab
- Arm Bar
- Knife Attack Situation
- Club Attack Situation

Baton Skills:

- First 3-Count (Poke, Overhand, & Underhand Thrusts)
- Second 3-Count (Forearm Smash, Overhand & Underhand Thrust)
- Third 3-Count (Forearm Smash, Shin & Leg Strike)
- 5-Count Thrust (Poke, Poke, Forearm Smash, Overhand & Underhand)
- Offensive Turn
- Defensive Turn
- 2-Count Forearm Slash
- 2-Count Power Slash
- 2-Count Leg Sweep

Arrest & Control:

- Twistlock Control Handcuffing Technique
- Prone Control Handcuffing Technique
- Low Profile Rear Wristlock Escort
- High Profile Rear Wristlock Escort
- Shuffle Pivot to Outside Deflection
- Progressive Pivot to Outside Deflection
- Shuffle Side Step to Outside Deflection
- Progressive Side Step to Outside Deflection
- Basic Straight Arm Bar

- Straight Arm Bar and Wristlock
- Elevated Arm Bar and Wristlock 2
- Standing Aikido Wristlock
- Kneeling Aikido Wristlock
- Prone Control
- Hammerlock Spin to Prone Control
- Front Wristlock
- Thumblock Front Wristlock
- Twistlock
- High Profile Rear Wristlock
- Assaultive Subject - Wrist Throw Takedown
- Assaultive Subject - Neck Nerve/Cross Face/Hair Pull Takedown
- Assaultive Subject - Arm Bar Takedown

Other Skills:

- Frisk Search
- Strip Search
- MK 3 & 4 OC Projector
- MK 9 OC Projector
- Gas Mask
- Grenade
- Application of the Body Belt
- Removal of the Body Belt
- Application of Leg Irons
- Removal of Leg Irons
- Two-Way Radio
- Loudhailer

Other, please specify:

4. What is the maximum length of time you feel should be used for a test of hard skills once the recruits have completed Stage 3 (the recruits would be tested before leaving the Staff College after Stage 3) and after they are already employed at an institution (they would have to be taken off the job to complete it). In other words, how much time is enough to test the recruits on the sample of hard skills you suggested above in question #1?

5. Do you have any other comments, concerns, or suggestions?

Appendix L

Definitions of the Hard Skills Assessed in the Current Study

Skill	Definition
(1) Frisk Search	The Corrections and Conditional Release Act (CCRA) defines a frisk search as “a manual search or a search by technical means, of the clothed body.” The Corrections and Conditional Release Regulations (CCRR) states “a frisk search shall be carried out from head to foot, down the front and rear of the body, around the legs and inside clothing folds, pockets and footwear.” The Checklist tests the officer’s ability to conduct a frisk search in a technically correct manner, in accordance with law and policy. It also tests that officers conduct the search in a respectful manner, and with due regard for Officer Safety.
(2) MK 3 & 4 OC Projector	The MK 3 and 4 OC projectors are canisters of Oleoresin Capsicum (Pepper Spray) that are designed to be carried on the officer’s person. The Checklist tests the officer’s ability to determine the correct use OC Spray within law and policy. It also tests their ability to deploy the OC in a way that is likely to assist in subduing an inmate.
(3) Two-Way Radio	Two- way radios are the primary form of mobile communication within the institutions. The Checklist tests the officer’s ability to follow proper radio procedures and to communicate effectively over the radio.
(4) Twistlock Handcuffing	Twistlock handcuffing is a controlled handcuffing technique for cuffing inmates behind the back. Inmates are handcuffed behind the back when their risk is not manageable by handcuffing in front. Twistlock handcuffing is used for inmates whose behaviour ranges from cooperative to verbally resistive, but not physically uncooperative. The checklist tests the officer’s ability to apply the Twistlock control and apply the handcuffs. It also tests that the officer applies the handcuffs with due regard for inmate and officer safety.
(5) Low Profile Rear Wristlock Escort	The Low Profile Rear Wristlock Escort is used to escort inmates who are handcuffed to the rear, and are cooperative. The Checklist tests the officers ability to apply the wristlock in a technically correct manner, with due regard for inmate and officer safety.

<p>(6) High Profile Rear Wristlock Escort</p>	<p>The High Profile Rear Wristlock Escort is used to escort inmates who are handcuffed to the rear, and are physically uncooperative or assaultive. The Checklist tests the officers ability to apply the wristlock in a technically correct manner, with due regard for inmate and officer safety.</p>
<p>(7) Application and Removal of Leg Irons</p>	<p>Leg Irons are used for both internal and external escorts. They are restraints that are applied around the ankles, similar to handcuffs. The Checklist tests the officers ability to apply and remove the restraint equipment in a technically correct manner, with due regard for inmate and officer safety.</p>
<p>(8) Application and Removal of Body Belt</p>	<p>Body Belts are used for external escorts. It is restraint equipment made up of handcuffs and a chain that surrounds the body. The Checklist tests the officers ability to apply and remove the restraint equipment in a technically correct manner, with due regard for inmate and officer safety.</p>

Appendix M

Hard Skills Scenarios

Scenario #1

Location: Men's Institution. Medium Security. Shop Area. Segregation Unit.		
Roles: 1 officer. 1 offender.	Officer equipment: Duty belt with cuffs, cuff keys, gloves, and flashlight. Inert OC Canister. Two-way radio.	Props/equipment: Sharp instrument resembling a knife.
Brief officer(s) with this information. Review officer equipment. Time: 1630 hours. You are conducting routine frisk searches of inmates as they are leaving the shop area. You have a radio and your call sign is K-1. The searches are going fine until you reach Offender Bennell, C; FPS 123334C. You start the search and nearing completion you find a potential weapon. Offender Bennell has a 6-inch sharpened piece of plastic taped to his right leg/ankle underneath his sock. Offender Bennell becomes agitated once you discover the weapon and you will need to control him using the OC spray. You will react appropriately in this situation taking into consideration the offender's demeanor. You may be prompted by the facilitator to perform certain techniques. You will need to use your two-way radio to call for backup after you discover the weapon and after you get the offender under control (NOTE: you will progress through the rest of the scenario without any backup). Ultimately you will need to handcuff Offender Bennell using the Twistlock Control Handcuffing Technique. Next, you will need to escort him to segregation using the Low Profile Wristlock Escort. The offender will become agitated again and try to run away; therefore you will need to use the High Profile Wristlock Escort.		
Provide this information to the offender. Time: 1630 hrs. You are offender Craig Bennell. You are serving your second sentence for drug trafficking. You did not have a great experience your first time being incarcerated and although this is a different institution, you are scared. Everything about this place scares you. To add to your situation, some of the other inmates are making you feel less than welcome. Overall, you do not feel very safe. During your time in the shop area you find a piece of plastic that you are able to modify to resemble a knife. You are not intent on using it but you just want to be prepared. You just want the sense of control. You have hidden your makeshift knife by taping it to the inside of your right leg/ankle and have concealed it with your sock. You are about to be searched by an officer. Once the officer finds the knife you are initially angry and behave in a way that makes the officer draw his or her OC canister and instruct you not to move. To ensure that the officer draws the OC you could start to move around, saying something like "Woah,		

woah, I wasn't going to use it on anyone" etc. If the officer does not draw the OC, the facilitator will call time out and prompt the officer to do so.

The officer will then use his/her radio to call for help and use the Twistlock Control Handcuffing Technique to restrain you. During this time, as long as the officer is not behaving in a way that escalates you, you can act cooperatively. The officer will use the Low Profile Rear Wristlock Escort to take you to the segregation unit. Once you are in the Low Profile Rear Escort on your way to segregation you will start to act out forcing the officer to use the High Profile Wrist Lock. Specifically, you will start to run away with your hands cuffed behind your back.

Stop/Time out guidelines:

- **Time out:** Call time out if the officers significantly jeopardize personal safety while searching, applying restraints or escorting offender Bennell or if you need to prompt the officer to perform a certain skill/technique.
- **Stop scenario:** The facilitator/evaluator will prompt the officer throughout the scenario, when needed, to demonstrate certain skills. Stop the scenario as soon as the hand cuffs are removed.

Skills to be demonstrated by Officer: As the scenario plays out if at any time the Officer chooses a technique not on the list you must call Time Out and prompt them to perform the appropriate technique.

- **Pat Down Frisk Search**
- **MK 3 & 4 OC Projector**
- **Two-way Radio**
- **Twistlock Control Handcuffing Technique**
- **Low Profile Rear Wristlock Escort**
- **High Profile Rear Wristlock Escort**

Tips for the Offender:

- The offender is initially angry towards the officer for finding his makeshift knife but defuses after the OC is utilized if the officer uses good defusing skills. The offender will get agitated again once put in the Low Profile Rear Wristlock Escort while being taken to segregation but defuses after being put in the High Profile Rear Wristlock Escort.
- If the officer adopts an interrogation style of communication, the offender could become more upset and agitated or become withdrawn and non-communicative.
- If the officer is more empathetic, the offender will be more willing to be cooperative (i.e., talk to the officer and reveal a bit of the problem driving this behaviour, be willing to listen to and be influenced by the officer.)
- If the officer is not clear with instructions, the offender should start to get agitated.

Points to Remember for the Facilitator/Evaluator:

- Prompt the officers when necessary (especially if they do not decide to use the OC spray).
- Fill out your evaluation tools as the scenario progresses & record any use of prompts.

Scenario #2

Location: Men's Institution. Medium Security. Segregation Unit. Transport to Maximum Security Institution.		
Roles: 1 officer. 1 offender.	Officer equipment: Duty belt with cuffs, cuff keys, gloves, and flashlight. Inert OC Canister. Two-way radio. Body belt. Leg Irons.	Props/equipment:
Brief officer(s) with this information. Review officer equipment. Time: 1700 hours. This is a continuation of the scenario completed above. You had previously conducted a routine frisk search in the shop area finding a concealed weapon (a plastic knife) on Offender Bennell. During the search he was agitated and you had to threaten/use your OC spray. After Offender Bennell calmed down you proceeded to restrain him and escort him to the segregation unit where you strip searched him. Now you are preparing to move Offender Bennell to a maximum security institution. You will need place transport chains and leg irons on the offender and upon arrival at the new institution remove the transport chains and leg irons.		
Provide this information to the offender. Time: 1700 hrs. You are offender Craig Bennell. You are currently in the segregation unit and are waiting to be transferred to a maximum security institution. Prior to this you were found concealing a makeshift knife when leaving the shop area. You were restrained and taken to the segregation unit and then strip searched. You have been agitated throughout some of the process; however, you have since calmed down and are actually looking forward to moving to a different institution since you were not feeling very welcome or safe here. You are cooperative with the officer as he/she prepares you for travel by placing transport chains and leg irons on you. You remain cooperative upon arrival at the new institution where the officer removes the transport chain and leg irons.		

Stop/Time out guidelines:

- **Time out:** Call time out if the officers significantly jeopardize personal safety while searching offender Bennell and applying restraints or if you need to prompt the officer to perform a certain skill/technique.
- **Stop scenario:** The facilitator/evaluator will prompt the officer throughout the scenario and demonstration of skills. Under the facilitator's direction, the transport chain and leg irons are placed on offender Bennell. Stop the scenario once the transport chain and leg irons are removed upon arrival at the new institution.

Skills to be demonstrated by Officer: As the scenario plays out if at any time the Officer chooses a technique not on the list you must call Time Out and prompt them to perform the appropriate technique.

- **Application of Leg Irons**
- **Application of Body Belt**
- **Removal of Body Belt**
- **Removal of Leg Irons**

Tips for the Offender:

- The offender is cooperative throughout the scenario.
- If the officer is not clear with instructions, the offender should start to get agitated.

Points to Remember for the Facilitator/Evaluator:

- Prompt the Officers when necessary.
- Fill out your evaluation tools as the scenario progresses and record any use of prompts.

Appendix N

Attitudes Toward Correctional Work Scale (Recruitment Phase)

Developed by Robinson et al. (1992) and adapted by the author to be delivered to potential trainees prior to training.

The statements listed below describe different attitudes toward correctional work. Please read each statement below and express *your feelings* about each statement by indicating whether you think the statement is *true* or *false*. Please circle the option that best describes your personal attitude. Please answer *every* item.

- | | | |
|--|------|-------|
| 1) For good reasons, the type of work done in corrections has a bad image with the public.* | True | False |
| 2) One of the most rewarding elements of correctional work will be that it is challenging. | True | False |
| 3) If I had the choice, I'd much prefer to work with non-offenders than offenders.* | True | False |
| 4) If it wasn't for the good pay, I would probably not choose a career in the field of corrections.* | True | False |
| 5) While every job has its rewards, offenders are the most interesting and challenging types of people to work with. | True | False |
| 6) In general, there are more good things than bad things about having a career in corrections. | True | False |
| 7) Being involved in the field of corrections will give me a personal sense of pride and accomplishment. | True | False |
| 8) I will not be very proud to tell people I earn my living working with offenders.* | True | False |
| 9) Generally, I would prefer to have a job in a different field than corrections.* | True | False |
| 10) What most attracts me to corrections is the type of work I will do, and not the pay, fringe benefits, or working conditions. | True | False |
| 11) I would have to agree that work in corrections is not a very respectable kind of job to have.* | True | False |
| 12) Working in corrections would be OK as long as you didn't have to | True | False |

deal with offenders directly.*

* indicates reverse scored items

Appendix O

Attitudes Toward Inmates Scale

Used by Bensimon (2005) and adapted from the Attitudes Toward Prisoners Scale developed by Melvin, Gramling, and Gardner (1985).

The statements listed below describe different attitudes toward inmates in institutions and prisons in Canada. There are no right or wrong answers, only opinions. Please read each statement below and express *your feelings* about each statement by indicating whether you (1) Strongly Disagree, (2) Disagree, (3) Undecided, (4) Agree, or (5) Strongly Agree. Indicate your opinion by circling the number that best describes your personal attitude in the left-hand margin. Please answer *every* item.

- | | | | | | |
|---|---|---|---|---|---|
| 1) Inmates are different from most people.* | 1 | 2 | 3 | 4 | 5 |
| 2) Only a few inmates are really dangerous. | 1 | 2 | 3 | 4 | 5 |
| 3) Inmates never change.* | 1 | 2 | 3 | 4 | 5 |
| 4) Most inmates are victims of circumstance and deserve to be helped. | 1 | 2 | 3 | 4 | 5 |
| 5) Inmates have feelings like the rest of us. | 1 | 2 | 3 | 4 | 5 |
| 6) It is not wise to trust an inmate too far.* | 1 | 2 | 3 | 4 | 5 |
| 7) I think I would like a lot of inmates. | 1 | 2 | 3 | 4 | 5 |
| 8) Bad institutional conditions just make an inmate more bitter. | 1 | 2 | 3 | 4 | 5 |
| 9) Give an inmate an inch and they will take a mile.* | 1 | 2 | 3 | 4 | 5 |
| 10) Most inmates are stupid.* | 1 | 2 | 3 | 4 | 5 |
| 11) Inmates need affection and praise just like anybody else. | 1 | 2 | 3 | 4 | 5 |
| 12) You should not expect too much from an inmate.* | 1 | 2 | 3 | 4 | 5 |
| 13) Trying to rehabilitate inmates is a waste of time and money.* | 1 | 2 | 3 | 4 | 5 |
| 14) You never know when an inmate is telling the truth.* | 1 | 2 | 3 | 4 | 5 |
| 15) Inmates are no better or worse than other people. | 1 | 2 | 3 | 4 | 5 |
| 16) You have to be constantly on your guard with inmates.* | 1 | 2 | 3 | 4 | 5 |

- | | | | | | |
|---|---|---|---|---|---|
| 17) In general, inmates think and act alike.* | 1 | 2 | 3 | 4 | 5 |
| 18) If you give an inmate your respect, they'll give you the same. | 1 | 2 | 3 | 4 | 5 |
| 19) Inmates only think about themselves.* | 1 | 2 | 3 | 4 | 5 |
| 20) There are some inmates I would trust with my life. | 1 | 2 | 3 | 4 | 5 |
| 21) Inmates will listen to reason. | 1 | 2 | 3 | 4 | 5 |
| 22) Most inmates are too lazy to earn an honest living.* | 1 | 2 | 3 | 4 | 5 |
| 23) I wouldn't mind living next door to an ex-inmate. | 1 | 2 | 3 | 4 | 5 |
| 24) Inmates are just plain mean at heart.* | 1 | 2 | 3 | 4 | 5 |
| 25) Inmates are always trying to get something out of somebody.* | 1 | 2 | 3 | 4 | 5 |
| 26) The values of most inmates are about the same as the rest of us? | 1 | 2 | 3 | 4 | 5 |
| 27) I would never want one of my children dating an ex-inmate.* | 1 | 2 | 3 | 4 | 5 |
| 28) Most inmates have the capacity to love. | 1 | 2 | 3 | 4 | 5 |
| 29) Inmates are just plain immoral.* | 1 | 2 | 3 | 4 | 5 |
| 30) Inmates should be under strict, hard discipline.* | 1 | 2 | 3 | 4 | 5 |
| 31) In general, inmates are basically bad people.* | 1 | 2 | 3 | 4 | 5 |
| 32) Most inmates can be rehabilitated. | 1 | 2 | 3 | 4 | 5 |
| 33) Some inmates are pretty nice people. | 1 | 2 | 3 | 4 | 5 |
| 34) I would like associating with some inmates. | 1 | 2 | 3 | 4 | 5 |
| 35) Inmates respect only brute force.* | 1 | 2 | 3 | 4 | 5 |
| 36) If a person does well in the institution, they should be let out on parole. | 1 | 2 | 3 | 4 | 5 |

* indicates reverse scored items

Appendix P

Attitudes Toward Correctional Work Scale (Transfer Phase)

Developed by Robinson et al. (1992) and to be delivered to trainees four to six months on the job.

The statements listed below describe different attitudes toward correctional work. Please read each statement below and express *your feelings* about each statement by indicating whether you think the statement is *true* or *false*. Please circle the option that best describes your personal attitude. Please answer *every* item.

- | | | |
|---|------|-------|
| 1) For good reasons, the type of work we do in corrections has a bad image with the public.* | True | False |
| 2) One of the most rewarding elements of correctional work is that it is challenging. | True | False |
| 3) If I had the choice, I'd much prefer to work with non-offenders than offenders.* | True | False |
| 4) If it wasn't for the good pay, I would probably not choose a career in the field of corrections.* | True | False |
| 5) While every job has its rewards, offenders are the most interesting and challenging types of people to work with. | True | False |
| 6) In general, there are more good things than bad things about having a career in corrections. | True | False |
| 7) Being involved in the field of corrections gives me a personal sense of pride and accomplishment. | True | False |
| 8) Usually, I am not very proud to tell people I earn my living working with offenders.* | True | False |
| 9) Generally, I would prefer to have a job in a different field than corrections.* | True | False |
| 10) What most attracts me to corrections is the type of work I do, and not the pay, fringe benefits, or working conditions. | True | False |
| 11) I would have to agree that work in corrections is not a very respectable kind of job to have.* | True | False |
| 12) Working in corrections would be OK as long as you didn't have to | True | False |

deal with offenders directly.*

* indicates reverse scored items

Appendix Q

Trainee Behavioural Checklist for CTP 2008

Instructions

1. Please complete this questionnaire as promptly as possible and send it to the following email address: shevauncorey@rogers.com. To provide responses, you will need to reflect on CTP and think about specific ways in which you have applied what you learned in CTP now that you are on the job.
2. Please take your time as you provide responses. Please answer every question on the checklists. Accurate and complete responses are very important. You should be able to provide thorough responses in about 15-20 minutes.
3. Please be objective in providing responses. In no way will your name be linked to your input. Your name is only required initially so as to enable the Principal Researcher to match your responses to the other data you have provided throughout the completion of CTP. Your questionnaire will only be viewed by the Principal Researcher from Carleton University. Specific responses or comments related to any individual will not be communicated to anyone at your institution (e.g., supervisors, etc.) or to your employer. Only the aggregated data without any identifying information will be provided to CSC.
4. Your responses will help determine the impact of CTP. In exchange for your participation in this evaluation, a copy of a report summarizing the results can be provided to you upon request once the evaluation is completed.
5. Should you need clarification or more information, for a copy of the report, or if you have any concerns, please contact the Principal Researcher, Shevaun Corey from Carleton University at shevauncorey@rogers.com.

1. Please indicate the ***frequency*** with which you engage in the following behaviours or utilize the following skills on the job ***on average per month***.

Behaviours/Skills	Never	1-5 Times	5-10 Times	10-15 Times	More than 15 Times	No Opportunity to Use Skill
Assessing Problematic Situations	1	2	3	4	5	6
Identifying Possible Solutions	1	2	3	4	5	6
Taking Necessary Measures	1	2	3	4	5	6
Motivating and Influencing Individuals	1	2	3	4	5	6
Communicating Effectively Orally	1	2	3	4	5	6
Communicating Effectively in Writing	1	2	3	4	5	6
Summarizing Information	1	2	3	4	5	6
Writing Reports						
• OSOR	1	2	3	4	5	6
• Post Search	1	2	3	4	5	6
• Use of Force	1	2	3	4	5	6
• Offence	1	2	3	4	5	6
Using Communication Equipment (e.g., two-way radio, loudhailer, etc.)	1	2	3	4	5	6
Using Restraint Equipment	1	2	3	4	5	6
Engaging in a Search (e.g., area, frisk, strip, etc.)	1	2	3	4	5	6
Escorts (e.g., internal and outside)	1	2	3	4	5	6
Using Chemical and Inflammatory Agents	1	2	3	4	5	6
Utilizing Fire Safety Measures	1	2	3	4	5	6
Using Self Defence Skills	1	2	3	4	5	6
Using Arrest and Control Skills	1	2	3	4	5	6
Using Baton Skills	1	2	3	4	5	6

2. Please indicate the **quality** of the skills you display ***in general*** on the job. Please use the following information as a guide for helping you to fill out the following checklist. If you ***always*** execute the behaviour or skill in a manner consistent with the way you were trained in CTP you should choose ***Above Average***. If you execute the skill in a manner consistent with the way you were trained in CTP ***most of the time*** you should choose ***Average***. If you ***rarely*** execute the skill in a manner consistent with the way you were trained in CTP you should choose ***Below Average***. If you have not had the opportunity to utilize a skill or engage in a particular behaviour choose ***No Opportunity to Use Skill***.

Behaviours/Skills	Below Average	Average	Above Average	No Opportunity to Use Skill
Assessing Problematic Situations	1	2	3	4
Identifying Possible Solutions	1	2	3	4
Taking Necessary Measures	1	2	3	4
Motivating and Influencing Individuals	1	2	3	4
Communicating Effectively Orally	1	2	3	4
Communicating Effectively in Writing	1	2	3	4
Summarizing Information	1	2	3	4
Writing Reports				
• OSOR	1	2	3	4
• Post Search	1	2	3	4
• Use of Force	1	2	3	4
• Offence	1	2	3	4
Using Communication Equipment (e.g., two-way radio, loudhailer, etc.)	1	2	3	4
Using Restraint Equipment	1	2	3	4
Engaging in a Search (e.g., area, frisk, strip, etc.)	1	2	3	4
Escorts (e.g., internal and outside)	1	2	3	4
Using Chemical and Inflammatory Agents	1	2	3	4
Utilizing Fire Safety Measures	1	2	3	4
Using Self Defence Skills	1	2	3	4
Using Arrest and Control Skills	1	2	3	4
Using Baton Skills	1	2	3	4

3. Have you used/reviewed the training materials you received for CTP 2008 since you finished the training program?

yes no

Please explain why or why not:

4. List the 3 behaviours you have used most frequently on the job as a result of CTP.

5. Please identify any specific accomplishments you've achieved on the job that you can link to your training in CTP.

6. What barriers, if any, have you encountered that have prevented you from using the skills or knowledge gained in CTP?

7. What specific suggestions do you have for improving the training delivered in CTP?

Appendix R

Supervisor Behavioural Checklist for CTP 2008

Instructions

1. Please complete this questionnaire as promptly as possible and send it to the following email address: shevauncorey@rogers.com. To provide responses, you will need to reflect on the particular trainee that you have supervised and think about the specific ways in which they have applied what they have learned from CTP 2008.
2. Please take your time as you provide responses. Please answer every item on the checklists. Accurate and complete responses are very important. You should be able to provide thorough responses in about 15 minutes.
3. Please be objective in providing responses. In no way will your name be linked to your input. The name of the trainee you are evaluating is required to enable us to match the data you are providing with the other information provided by that particular individual. Your questionnaire will only be viewed by the Principal Researcher from Carleton University. Specific responses or comments related to any individual will not be communicated to anyone at your institution (e.g., the trainee, etc.) or to your employer. Only the aggregated data without any identifying information will be provided to CSC. In turn, you are asked not to discuss the results of your evaluation with anyone, including the trainee themselves, other supervisors or coworkers, etc.
4. Your responses will help determine the impact of CTP. In exchange for your participation in this evaluation, a copy of a report summarizing the results can be provided to you upon request once the evaluation is completed.
5. Should you need clarification or more information, for a copy of the report, or if you have any concerns, please contact the Principal Researcher, Shevaun Corey from Carleton University at shevauncorey@rogers.com.

Trainee's name being evaluated: _____

When filling out the following checklist please think about the trainee's on-the-job behaviour in general, try not to focus on extreme cases. Use the scale provided to assess the frequency and quality of the trainee's on-the-job behaviour. Please try to assess each behaviour or skill honestly and to the best of your ability. If you are unable to assess the trainee's behaviour because you have not had the opportunity to observe it then please choose the *Not Observed* option on the scale.

1. Please indicate the ***frequency*** with which this trainee engages in the following behaviours or utilizes the following skills on the job ***on average per month***.

Behaviours/Skills	Never	1-5 Times	5-10 Times	10-15 Times	More than 15 Times	Not Observed
Assessing Problematic Situations	1	2	3	4	5	6
Identifying Possible Solutions	1	2	3	4	5	6
Taking Necessary Measures	1	2	3	4	5	6
Motivating and Influencing Individuals	1	2	3	4	5	6
Communicating Effectively Orally	1	2	3	4	5	6
Communicating Effectively in Writing	1	2	3	4	5	6
Summarizing Information	1	2	3	4	5	6
Writing Reports						
• OSOR	1	2	3	4	5	6
• Post Search	1	2	3	4	5	6
• Use of Force	1	2	3	4	5	6
• Offence	1	2	3	4	5	6
Using Communication Equipment (e.g., two-way radio, loudhailer, etc.)	1	2	3	4	5	6
Using Restraint Equipment	1	2	3	4	5	6
Engaging in a Search (e.g., area, frisk, strip, etc.)	1	2	3	4	5	6
Escorts (e.g., internal and outside)	1	2	3	4	5	6
Using Chemical and Inflammatory Agents	1	2	3	4	5	6
Utilizing Fire Safety Measures	1	2	3	4	5	6
Using Self Defence Skills	1	2	3	4	5	6

Using Arrest and Control Skills	1	2	3	4	5	6
Using Baton Skills	1	2	3	4	5	6

2. Please indicate the **quality** of the skills this trainee displays ***in general*** on the job. Please use the following information as a guide for helping you to fill out the following checklist. If the trainee executes the behaviour or skill in a manner consistent with how you expect them to be trained in CTP, ***all of the time***, you should choose ***Above Average***. If they execute the skill in a manner consistent with how you would expect them to be trained in CTP, ***most of the time***, you should choose ***Average***. If they ***rarely*** execute the skill in a manner consistent with how you would expect them to be trained in CTP, you should choose ***Below Average***. If you have not had the opportunity to observe the quality of the skill or behaviour choose ***Not Observed***.

Behaviours/Skills	Below Average	Average	Above Average	Not Observed
Assessing Problematic Situations	1	2	3	4
Identifying Possible Solutions	1	2	3	4
Taking Necessary Measures	1	2	3	4
Motivating and Influencing Individuals	1	2	3	4
Communicating Effectively Orally	1	2	3	4
Communicating Effectively in Writing	1	2	3	4
Summarizing Information	1	2	3	4
Writing Reports				
• OSOR	1	2	3	4
• Post Search	1	2	3	4
• Use of Force	1	2	3	4
• Offence	1	2	3	4
Using Communication Equipment (e.g., two-way radio, loudhailer, etc.)	1	2	3	4
Using Restraint Equipment	1	2	3	4
Engaging in a Search (e.g., area, frisk, strip, etc.)	1	2	3	4
Escorts (e.g., internal and outside)	1	2	3	4
Using Chemical and	1	2	3	4

Inflammatory Agents				
Utilizing Fire Safety Measures	1	2	3	4
Using Self Defence Skills	1	2	3	4
Using Arrest and Control Skills	1	2	3	4
Using Baton Skills	1	2	3	4

Appendix S

DEBRIEFING FORM
(For Trainees)*Research Title*

An Evaluation of the Correctional Officer Training Program (CTP)

What are we trying to learn with this research?

Although it is of great interest to know what trainees think about their training, and whether or not trainees are learning what is taught in training, of equal if not greater importance is determining whether or not knowledge and skills are retained and utilized on the job. In other words, are trainees able to transfer what they learn during training to the job? The ultimate goal of any training program should be to provide trainees with skills and knowledge that will help them perform their job more effectively and efficiently (Baldwin & Ford, 1988; Broad & Newstrom, 1992). Thus, one of the main reasons why organizations evaluate their training programs is to investigate whether transfer of training has occurred. Other reasons evaluations are done are to provide an indication of how training fits in to the broader goals of the organization and to make improvements on future training programs (Kirkpatrick, 1994). A large scale evaluation of CTP will provide CSC with this information.

Why is this research important?

The results emerging from this study will help to determine how you, the trainee, feels about the training program, whether involvement in CTP has had its intended impact on you with respect to the knowledge, skills, and attitudes that you have gained, whether you have transferred what you have learned to your job, what factors help and hinder transfer of training to the job, and ultimately, what impact your training has on CSC. This information will provide the curriculum designers, trainers, and researchers in the Learning & Development Branch of CSC with valuable information that will allow them to modify CTP so that it can reach its maximum potential. Ultimately, this research will help to ensure that correctional officers receive the best training possible so that you can perform your job in an effective and efficient manner.

Where can I learn more?

For more information about training evaluation please see:

Kirkpatrick, D. L. (1994). *Evaluating training programs: The four levels*. San Francisco, CA: Berrett-Koehler.

What if I have questions later?

We would like to thank you for your participation in this study. As well, we would like to remind you that your information will remain confidential and that you may ask us to remove your data from our study if you wish. If you have any questions about the research please contact the Principal Investigator, Shevaun Corey (613-520-2600X1728,) or the Faculty Supervisor, Craig

Bennell (613-520-2600X1769, cbennell@connect.carleton.ca). Should you have any ethical concerns about this study, please contact Dr. Avi Parush (Chair, Carleton University Ethics Committee for Psychological Research, 613-520-2600x6026, avi_parush@carleton.ca) or Christa Gillis (Director, Operational Research at CSC, 613-943-8844). Should you have any other concerns about the study, please contact Dr. Janet Mantler (Chair, Department of Psychology, Carleton University, 613-520-2600x4173, psychchair@carleton.ca).

Appendix T

DEBRIEFING FORM
(For Supervisors)*Research Title*

An Evaluation of the Correctional Officer Training Program (CTP)

What are we trying to learn with this research?

Although it is of great interest to know what trainees think about their training, and whether or not trainees are learning what is taught in training, of equal if not greater importance is determining whether or not knowledge and skills are retained and utilized on the job. In other words, are trainees able to transfer what they learn during training to the job? The ultimate goal of any training program should be to provide trainees with skills and knowledge that will help them perform their job more effectively and efficiently (Baldwin & Ford, 1988; Broad & Newstrom, 1992). Thus, one of the main reasons why organizations evaluate their training programs is to investigate whether transfer of training has occurred. Other reasons evaluations are done are to provide an indication of how training fits in to the broader goals of the organization and to make improvements on future training programs (Kirkpatrick, 1994). A large scale evaluation of CTP will provide CSC with this information.

Why is this research important?

The results emerging from this study will help to determine whether involvement in CTP has its intended impact on trainees with respect to the knowledge, skills, and attitudes they gain, whether they have transferred what they have learned to their job, what factors help and hinder transfer of training to the job, and ultimately, what impact the training has on CSC. This information will provide the curriculum designers, trainers, and researchers in the Learning & Development Branch of CSC with valuable information that will allow them to modify CTP so that it can reach its maximum potential. Ultimately, this research will help to ensure that correctional officers receive the best training possible so that they can perform their job in an effective and efficient manner.

Where can I learn more?

For more information about training evaluation please see:

Kirkpatrick, D. L. (1994). *Evaluating training programs: The four levels*. San Francisco, CA: Berrett-Koehler.

What if I have questions later?

We would like to thank you for your participation in this study. As well, we would like to remind you that your information will remain confidential and that you may ask us to remove your data from our study if you wish. If you have any questions about the research please contact the Principal Investigator, Shevaun Corey (613-355-7496, or the Faculty Supervisor, Craig Bennell (613-520-2600X1769, cbennell@connect.carleton.ca). Should you have any

ethical concerns about this study, please contact Dr. Avi Parush (Chair, Carleton University Ethics Committee for Psychological Research, 613-520-2600x6026, avi_parush@carleton.ca) or Christa Gillis (Director, Operational Research at CSC, 613-943-8844). Should you have any other concerns about the study, please contact Dr. Janet Mantler (Chair, Department of Psychology, Carleton University, 613-520-2600x4173, psychchair@carleton.ca).

Appendix U
Data Collection Schedule

CTP Evaluation Schedule - FY 09/10

<u>Region</u>	<u>CTP</u>	<u>April</u>	<u>May</u>	<u>June</u>	<u>July</u>	<u>August</u>	<u>September</u>	<u>October</u>	<u>November</u>	<u>December</u>	<u>January</u>	<u>February</u>	<u>March</u>
QUEBEC	108												
	109												
ONTARIO	10			PHASE 1					PHASE 2				
	11			PHASE 1						PHASE 2			
	12					PHASE 1					PHASE 2		
	13						PHASE 1						PHASE 2
	15												
	16												
PACIFIC	5						PHASE 1			PHASE 2			
	7							PHASE 1			PHASE 2		
	9								PHASE 1				PHASE 2
	10									PHASE 1			PHASE 2

CTP Evaluation Schedule - FY 10/11

<u>Region</u>	<u>CTP</u>	<u>April</u>	<u>May</u>	<u>June</u>	<u>July</u>	<u>August</u>	<u>September</u>	<u>October</u>	<u>November</u>	<u>December</u>	<u>January</u>	<u>February</u>	<u>March</u>
QUEBEC	108		PHASE 2					PHASE 3					
	109	PHASE 2			PHASE 2				PHASE 3				
	110			PHASE 2					PHASE 2				PHASE 3
ONTARIO	10		PHASE 3										
	11		PHASE 3										
	12			PHASE 3									
	13					PHASE 3							
	15	PHASE 2				PHASE 2				PHASE 3			
	16	PHASE 2				PHASE 2				PHASE 3			
	17		PHASE 2				PHASE 2				PHASE 3		
	18				PHASE 2				PHASE 2				PHASE 3
PACIFIC	5		PHASE 3										
	7			PHASE 3									
	9				PHASE 3								
	10					PHASE 3							
	1			PHASE 2			PHASE 2				PHASE 3		

PHASE 1 = Recruitment Phase

PHASE 2 = Training Phase

PHASE 3 = Transfer Phase

Appendix V

Descriptive Statistics per Item of the Trainee Reaction Questionnaire Ontario & Pacific Regions Combined

Sections & Items	<i>n</i>	<i>Range</i>	<i>Mean (SD)</i>
Training Program Materials			
1) I felt the training material was interesting.	67	4-5	4.65 (.48)
2) I felt motivated by the training material.	67	4-5	4.63 (.49)
3) I felt it was easy to follow the material contained in CTP.	67	2-5	4.49 (.63)
4) I felt that the training material was easy to understand.	67	2-5	4.56 (.60)
5) I did not feel overwhelmed by the training material.	67	1-5	4.30 (.89)
6) Overall, I felt that the training content was not too difficult to learn.	67	3-5	4.54 (.57)
7) The training material provided in CTP prepared me to be successful on any formal evaluations.	67	3-5	4.54 (.54)
8) I felt that the training material provided me with knowledge and skills that will be useful on the job.	67	3-5	4.63 (.52)
9) I feel that the training material has real practical value.	67	3-5	4.68 (.51)
Training Program Overall			
10) Overall, I think the Correctional Officer Training Program (CTP) met the stated objectives.	67	2-5	4.67 (.58)
11) The training program was well structured.	67	2-5	4.61 (.68)

12) The training program flowed logically.	67	1-5	4.58 (.73)
13) In general, the evaluation methods were fair.	67	2-5	4.54 (.66)
14) I felt the evaluations effectively tested my knowledge of the training material.	67	2-5	4.63 (.62)
15) Overall, I felt that the tests were not too difficult.	67	1-5	4.51 (.76)
16) The different types of learning activities (e.g., online modules, classroom-based learning, practical applications, etc.) enhanced my comprehension of key content areas.	67	3-5	4.61 (.53)
17) I felt able to self evaluate my learning through the different learning activities.	67	3-5	4.58 (.60)
18) I felt that the pace of the training program was just right.	67	1-5	4.32 (.83)
19) I feel that the quality of my work on the job will improve as a result of completing this training program.	67	3-5	4.65 (.55)
20) I feel motivated to use what I have learned in this training program in the future, when I am on the job.	67	4-5	4.72 (.45)

Stages of CTP

21) In general, I am satisfied with the following elements of Stage 1 of CTP (the online component):			
Length	67	1-5	4.25 (.99)
Quality	67	1-5	4.35 (.81)
Content	67	1-5	4.35 (.81)
Time I Invested	67	1-5	4.37 (.79)
22) As a whole, the online component (Stage 1) was a valuable part of CTP.	67	1-5	4.37 (.82)

23) In general, I am satisfied with the following elements of Stage 3 of CTP:			
Length	67	1-5	4.30 (.93)
Quality	67	4-5	4.65 (.48)
Content	67	2-5	4.61 (.59)
Time I Invested	67	2-5	4.60 (.62)
24) As a whole, the theory-based classroom component of Stage 3 was a valuable part of CTP.	67	2-5	4.65 (.58)
25) As a whole, the practical application component of Stage 3 was a valuable part of CTP.	67	4-5	4.72 (.45)
26) I felt that I was able to link the different topic areas throughout the different Stages within CTP.	67	3-5	4.65 (.52)
27) The estimated time allotted to complete Stage 1 of CTP was realistic.	67	1-5	4.56 (.71)
28) The estimated time allotted to complete Stage 2 of CTP was realistic.	67	1-5	4.32 (.97)
29) The estimated time allotted to complete Stage 3 of CTP was realistic.	67	1-5	4.60 (.75)
<hr/>			
Facilitator			
<hr/>			
30) In general, facilitators' were well prepared and organized.	67	2-5	4.68 (.57)
31) Overall, the facilitators' communicated in a clear and concise manner (e.g., when answering questions, providing direction/instruction, etc.).	67	2-5	4.72 (.56)
32) In general, I felt that the facilitators' possessed a lot of practical knowledge.	67	4-5	4.84 (.37)
33) In general, facilitators' created and maintained a positive learning environment.	67	1-5	4.68 (.69)
34) In general, the facilitators' served as positive and professional role models.	67	1-5	4.75 (.64)

35) Overall, I felt that the facilitators' provided feedback that was timely, specific and helped me identify areas in need of development.	67	2-5	4.72 (.59)
---	----	-----	------------

Appendix W

Descriptive Statistics per Item of the Trainee Reaction Questionnaire for the Pacific Region

Sections & Items	<i>n</i>	<i>Range</i>	<i>Mean (SD)</i>
Training Program Materials			
1) I felt the training material was interesting.	34	4-5	4.61 (.50)
2) I felt motivated by the training material.	34	4-5	4.61 (.50)
3) I felt it was easy to follow the material contained in CTP.	34	4-5	4.52 (.51)
4) I felt that the training material was easy to understand.	34	4-5	4.58 (.50)
5) I did not feel overwhelmed by the training material.	34	3-5	4.42 (.56)
6) Overall, I felt that the training content was not too difficult to learn.	34	3-5	4.52 (.57)
7) The training material provided in CTP prepared me to be successful on any formal evaluations.	34	4-5	4.52 (.51)
8) I felt that the training material provided me with knowledge and skills that will be useful on the job.	34	4-5	4.65 (.49)
9) I feel that the training material has real practical value.	34	4-5	4.68 (.48)
Training Program Overall			
10) Overall, I think the Correctional Officer Training Program (CTP) met the stated objectives.	34	4-5	4.71 (.46)
11) The training program was well structured.	34	4-5	4.68 (.48)

12) The training program flowed logically.	34	4-5	4.68 (.48)
13) In general, the evaluation methods were fair.	34	4-5	4.61 (.50)
14) I felt the evaluations effectively tested my knowledge of the training material.	34	4-5	4.68 (.48)
15) Overall, I felt that the tests were not too difficult.	34	3-5	4.65 (.55)
16) The different types of learning activities (e.g., online modules, classroom-based learning, practical applications, etc.) enhanced my comprehension of key content areas.	34	4-5	4.61 (.50)
17) I felt able to self evaluate my learning through the different learning activities.	34	3-5	4.55 (.57)
18) I felt that the pace of the training program was just right.	34	3-5	4.39 (.67)
19) I feel that the quality of my work on the job will improve as a result of completing this training program.	34	3-5	4.61 (.56)
20) I feel motivated to use what I have learned in this training program in the future, when I am on the job.	34	4-5	4.74 (.45)

Stages of CTP

21) In general, I am satisfied with the following elements of Stage 1 of CTP (the online component):			
Length	34	1-5	4.29 (.94)
Quality	34	1-5	4.19 (.95)
Content	34	1-5	4.19 (.95)
Time I Invested	34	1-5	4.23 (.85)
22) As a whole, the online component (Stage 1) was a valuable part of CTP.	34	1-5	4.19 (.95)

23) In general, I am satisfied with the following elements of Stage 3 of CTP:			
Length	34	1-5	4.23 (.96)
Quality	34	4-5	4.65 (.49)
Content	34	4-5	4.65 (.49)
Time I Invested	34	2-5	4.55 (.68)
24) As a whole, the theory-based classroom component of Stage 3 was a valuable part of CTP.	34	4-5	4.65 (.49)
25) As a whole, the practical application component of Stage 3 was a valuable part of CTP.	34	4-5	4.65 (.49)
26) I felt that I was able to link the different topic areas throughout the different Stages within CTP.	34	3-5	4.58 (.56)
27) The estimated time allotted to complete Stage 1 of CTP was realistic.	34	1-5	4.42 (.85)
28) The estimated time allotted to complete Stage 2 of CTP was realistic.	34	1-5	4.10 (1.11)
29) The estimated time allotted to complete Stage 3 of CTP was realistic.	34	1-5	4.48 (.81)

Facilitator

30) In general, facilitators' were well prepared and organized.	34	4-5	4.74 (.45)
31) Overall, the facilitators' communicated in a clear and concise manner (e.g., when answering questions, providing direction/instruction, etc.).	34	4-5	4.81 (.40)
32) In general, I felt that the facilitators' possessed a lot of practical knowledge.	34	4-5	4.90 (.30)
33) In general, facilitators' created and maintained a positive learning environment.	34	4-5	4.77 (.43)
34) In general, the facilitators' served as positive and professional role models.	34	4-5	4.84 (.37)

35) Overall, I felt that the facilitators' provided feedback that was timely, specific and helped me identify areas in need of development.	34	3-5	4.74 (.51)
---	----	-----	------------

Appendix X

Descriptive Statistics per Item of the Trainee Reaction Questionnaire for the Ontario Region

Sections & Items	<i>n</i>	<i>Range</i>	<i>Mean (SD)</i>
Training Program Materials			
1) I felt the training material was interesting.	33	4-5	4.69 (.47)
2) I felt motivated by the training material.	33	4-5	4.65 (.49)
3) I felt it was easy to follow the material contained in CTP.	33	2-5	4.46 (.76)
4) I felt that the training material was easy to understand.	33	2-5	4.54 (.71)
5) I did not feel overwhelmed by the training material.	33	1-5	4.15 (1.16)
6) Overall, I felt that the training content was not too difficult to learn.	33	3-5	4.58 (.58)
7) The training material provided in CTP prepared me to be successful on any formal evaluations.	33	3-5	4.58 (.58)
8) I felt that the training material provided me with knowledge and skills that will be useful on the job.	33	3-5	4.62 (.57)
9) I feel that the training material has real practical value.	33	3-5	4.69 (.55)
Training Program Overall			
10) Overall, I think the Correctional Officer Training Program (CTP) met the stated objectives.	33	2-5	4.62 (.70)
11) The training program was well structured.	33	2-5	4.54 (.86)

12) The training program flowed logically.	33	1-5	4.46 (.65)
13) In general, the evaluation methods were fair.	33	2-5	4.46 (.81)
14) I felt the evaluations effectively tested my knowledge of the training material.	33	2-5	4.58 (.76)
15) Overall, I felt that the tests were not too difficult.	33	1-5	4.35 (.94)
16) The different types of learning activities (e.g., online modules, classroom-based learning, practical applications, etc.) enhanced my comprehension of key content areas.	33	3-5	4.62 (.57)
17) I felt able to self evaluate my learning through the different learning activities.	33	3-5	4.62 (.64)
18) I felt that the pace of the training program was just right.	33	1-5	4.23 (.99)
19) I feel that the quality of my work on the job will improve as a result of completing this training program.	33	3-5	4.69 (.55)
20) I feel motivated to use what I have learned in this training program in the future, when I am on the job.	33	4-5	4.69 (.47)

Stages of CTP

21) In general, I am satisfied with the following elements of Stage 1 of CTP (the online component):			
Length	33	1-5	4.19 (1.06)
Quality	33	3-5	4.54 (.58)
Content	33	3-5	4.54 (.58)
Time I Invested	33	2-5	4.54 (.71)
22) As a whole, the online component (Stage 1) was a valuable part of CTP.	33	3-5	4.58 (.58)

23) In general, I am satisfied with the following elements of Stage 3 of CTP:			
Length	33	2-5	4.38 (.90)
Quality	33	4-5	4.65 (.49)
Content	33	2-5	4.58 (.70)
Time I Invested	33	3-5	4.65 (.56)
24) As a whole, the theory-based classroom component of Stage 3 was a valuable part of CTP.	33	2-5	4.65 (.69)
25) As a whole, the practical application component of Stage 3 was a valuable part of CTP.	33	4-5	4.81 (.40)
26) I felt that I was able to link the different topic areas throughout the different Stages within CTP.	33	4-5	4.73 (.45)
27) The estimated time allotted to complete Stage 1 of CTP was realistic.	33	4-5	4.73 (.45)
28) The estimated time allotted to complete Stage 2 of CTP was realistic.	33	3-5	4.58 (.70)
29) The estimated time allotted to complete Stage 3 of CTP was realistic.	33	2-5	4.73 (.67)

Facilitator

30) In general, facilitators' were well prepared and organized.	33	2-5	4.62 (.70)
31) Overall, the facilitators' communicated in a clear and concise manner (e.g., when answering questions, providing direction/instruction, etc.).	33	2-5	4.62 (.70)
32) In general, I felt that the facilitators' possessed a lot of practical knowledge.	33	4-5	4.77 (.43)
33) In general, facilitators' created and maintained a positive learning environment.	33	1-5	4.58 (.90)
34) In general, the facilitators' served as positive and professional role models.	33	1-5	4.65 (.85)

35) Overall, I felt that the facilitators' provided feedback that was timely, specific and helped me identify areas in need of development.	33	2-5	4.69 (.68)
---	----	-----	------------

Appendix Y

Qualitative Analysis of Trainee Reaction Comments

The table below outlines the categories, created during the content analyses that were used for the qualitative analysis.

Category	Definition	Examples
1. Instructors	Comments in this category included anything related to the instructors of the courses.	<ul style="list-style-type: none"> – “The instructors were always available to help us and answer our questions” – “Administrators/facilitators were outstanding professionals and are a huge asset to the program. Thanks to everyone!”
2. Content	Comments in this category included anything related to the content of the course, including timing issues.	<ul style="list-style-type: none"> – “Time to complete online portions of CTP stage 1 & 2. Suggest more time for online portion and session 9.” – “No Aboriginal content?”
3. Compensation	Comments in this category included anything related to the compensation (or lack thereof) during the program.	<ul style="list-style-type: none"> – “Compensation for all stages would help.” – “10 weeks is a long time without pay. Pay recruits some type of allowance. Even if it is \$100 a week.”
4. Technical Difficulties	Comments in this category included anything related to difficulties involving technology when accessing the program.	<ul style="list-style-type: none"> – “Could not open program with my MAC laptop.” – “Had some server issues on stage 1 – tests not working properly”
5. Other Comments	Comments in this category were placed here because they were not classifiable into any of the other categories.	<ul style="list-style-type: none"> – “I would be more critical about the performed skills. There should be higher standards to pass.” – “More contact with your classmates in stage 1 to

		help each other”
6. Overall Comments	Comments in this category included anything about the overall nature of the program.	<ul style="list-style-type: none"> – “Great program, education, eye opening.” – “Overall – great experience. Gained a lot of knowledge and practical skills that have prepared me for my career with CSC.”

The number of comments made by trainees is provided in the table below. Note that a single trainee could be associated with a number of comments; therefore the number of comments does not equal the sample sizes. In addition, all comments were kept in the original language of the trainee so as to eliminate any guesses on the part of the researcher and to provide all relevant information to CSC for formative feedback purposes.

Number of Trainee Comments from the Trainee Reaction Questionnaire

Region	Overall (n = 25)					Pacific (n = 17)					Ontario (n = 8)				
Category	S 1	S 2	S 3	CTP Overall	Total	S 1	S 2	S 3	CTP Overall	Total	S 1	S 2	S 3	CTP Overall	Total
1. Instructors	0	0	2	5	7	0	0	1	4	5	0	0	1	1	2
2. Content	4	7	7	6	24	3	7	4	5	19	1	0	3	1	5
3. Compensation	2	1	2	4	9	1	1	2	2	6	1	0	0	2	3
4. Technical Difficulties	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0
5. Other Comments	2	0	0	1	3	0	0	0	1	1	2	0	0	0	2
6. Overall Comments	6	6	4	10	26	4	4	4	7	19	2	2	0	3	7
Total	14	15	15	26	70	8	13	11	19	51	6	2	4	7	19

Note. S1 refers to Stage 1, S2 to Stage 2 and S3 to Stage 3.

The following table outlines all of the specific comments provided by trainees organized by Region and comment category. Overall, the majority of the comments concerned CTP content or were general comments about the program as a whole (e.g., Self-defence and arrest & control should have more time allotted to learn skills; Excellent program – hard, intense but rewarding.). This comment structure was consistent across both regions. Proportionally; more comments were made by more people in the Pacific region (approximately 3 per person) compared to the Ontario region, which had fewer comments by fewer people (approximately 2 per person). The main purpose of these comments is to provide formative feedback to CSC, as indicated above, as well as some context to the quantitative results.

Pacific Region

1. Instructor Comments

- | | |
|-------------|--|
| Stage 3 | <ul style="list-style-type: none"> - Not enough access to other facilitators. Have at least one available during the 15 minutes prior to class beginning. |
| CTP Overall | <ul style="list-style-type: none"> - Facilitators were great with course delivery. Good mentors. - Facilitators were exceptional at their job - Not enough facilitators allocated at certain times. Better scheduling. There were times we sat and waited instead of being tested. - Thanks to great instructors |
-

2. Content Comments

- | | |
|---------|--|
| Stage 1 | <ul style="list-style-type: none"> - Provide recruits with printouts of all materials; completion of course on paper rather than online would be more reader friendly - Lots of work is not memorable or useful. Reduce quantity improve quality |
|---------|--|

-
- The material was not presented in a concise manner. Streamline/reduce the content of stage 1. The material was only accessible via internet connection. For applicants with difficulty accessing internet, provide a hard copy. The material presented a very large commitment. Condense the volume of material and place stage 1 as part of the CTP at the college setting.
- Stage 2
- Too much time allotted
 - Have a review of the stage 2 material before stage 3 exams
 - Never knew how we were doing regarding assignment. No verification of being on the right track. Online quizzes/tests to verify you are on the right track.
 - Online quizzes to ascertain if we are still absorbing the right information and/or if we are misinterpreting the information.
 - A little more time.
 - Workbook instructions did not match online.
 - The material should be condensed and part of the CTP college time.
- Stage 3
- Designate more hours to arrest and control, practical application throughout course would help with retention. More time for firearms.
 - More time for scenarios.
 - Feel very rushed at the range. Arrest & Control and Self Defence you can really hurt individuals more too. More time at the range, extra week. More time in the PSA, extra week.
 - The content should be presented in a timely manner, shortening the time required for the material. To improve the delivery of the CTP, make an effort to reward participants for working hard by shortening the day or an incentive to raise the level of professionalism and work ethic. Example would be: instructors able to shorten day/workload if behaviour is exhibited at a high level of professionalism.
- CTP Overall
- I would prefer having short 10-15 min review sessions of practical skills at the end of the day (instructor led) rather than having some shortened days. Keeps skills fresh and ensures what we practice at home is done correctly. More time for practical application of arrest and control and firearms.
 - Would add two more weeks to overall course in practical areas.
 - Narrow Stage 1
 - A couple more institutional tours would be a really great element. It helps to make it real for us.
 - More realistic scenario-based training methods, including interactive video simulation. Video based simulation training would be helpful in preparing response options/ideas for
-

incidents in the institutions. Hiring actors for scenarios. Help prepare participants for the realities of the situations they may encounter. More focus on preparing participants to know their authority and what they can and cannot do legally. Very little discussion on stress management techniques for working as a correctional officer. More emphasis needed on how to integrate CTP with what staff in institutions are teaching new officers. It would be helpful to interact/interview a CO1 who has been working for a 6-month period and learn from senior officers on how to integrate CTP with transitioning into the institution.

3. Compensation Comments

- | | |
|-------------|---|
| Stage 1 | - Needs pay |
| Stage 2 | - Needs pay |
| Stage 3 | - Pay us during training
- Needs pay |
| CTP Overall | - Pay
- It would be great if CSC and Human Resources Canada can put together a training allowance program that will reduce recruits financial stress burden. |

4. Technical Difficulties Comments

- | | |
|---------|---|
| Stage 2 | - Minor technical glitches. For example, FireFox could not access the online training and when the cursor is placed over a box the title appeared over it and covered other text too. |
|---------|---|

5. Other Comments

- | | |
|-------------|---|
| CTP Overall | - My online application was submitted at the end of February 2009. When I complete my OJT it would almost be 1 year's time. It's too long to commit or expected to wait. Timely decision was not, to my opinion, executed. I heard some recruits only had one interview instead of 2 like me. |
|-------------|---|

Stage 1	<ul style="list-style-type: none">- No suggestions for improvement.- No concerns or suggestions for improvement.- I am very satisfied with level of training.- Very straight forward.
Stage 2	<ul style="list-style-type: none">- No concerns or suggestions for improvement.- No concerns or suggestions for improvement.- No concerns or suggestions for improvement.- Straight forward
Stage 3	<ul style="list-style-type: none">- No concerns or suggestions for improvement.- No concerns or suggestions for improvement.- No concerns or suggestions for improvement.- A bit challenging in some areas.
CTP Overall	<ul style="list-style-type: none">- The CTP definitely surpassed my expectations of what I thought the program would consist of. I really enjoyed the reality based training.- Excellent program.- Excellent program – hard, intense but rewarding.- Good with exception of Stage 1 Content- No concerns or suggestions for improvement.- I was incredibly impressed by the whole program. The instructors were amazing at consistently providing a safe environment for us to learn. It was so well organized and I feel ready for OJT. Thank you!- Great program, education, eye opening

Ontario Region

1. Instructor Comments

- Stage 3 - General consensus among instructors for marking criteria
- CTP Overall - Instructors were absolutely amazing. Hope they keep up the great work with all other recruits.
-

2. Content Comments

- Stage 1 - More direct contact with the Stage 1 facilitators
- Stage 3 - Self-defence and arrest & control should have more time allotted to learn skills.
 - More time needed for firearms training as 2 weeks is not enough time.
 - Institution specific.
- CTP Overall - Maybe reduce the time allotted for Stage 1, 2, or 3 and have a longer OJT. More institution specific training would be valuable.
-

3. Compensation Comments

- Stage 1 - Compensation
- CTP Overall - Compensation for all stages would help
 - Would be nice to get paid during training
-

5. Other Comments

- Stage 1 - More contact with your classmates in stage 1 to help each other
 - Online material hard to find
-

6. Overall Comments

-
- | | |
|-------------|---|
| Stage 1 | <ul style="list-style-type: none">- It was great. No suggestions for improvement- Good. Gave a good basis for the entire program and prepared us for topics discussed on core |
| Stage 2 | <ul style="list-style-type: none">- Great!- Good |
| CTP Overall | <ul style="list-style-type: none">- Great experience, learned a lot of new skills- Good training.- Good training program and awesome instructors/facilitators.
Thank you :) |
-

Appendix Z

Descriptive Statistics for the Level 2 and 3 Training Outcomes as well as the Training Effectiveness Factors for the Quebec Region

Knowledge

Region/Knowledge Test and Sections	<i>n</i>	<i>M</i> % (<i>SD</i>)	<i>Range</i>
Knowledge Test (Recruitment Phase)	14	57.37 (12.06)	30.53-70.53
Stage 1	14	85.17 (7.21)	73.08-96.15
Stage 2	14	72.62 (9.49)	58.33-91.67
Stage 3	14	40.12 (16.64)	8.84-57.38
Knowledge Test (Training Phase)	10	81.92 (2.63)	75.76-84.85
Stage 1	10	84.62 (3.63)	80.77-88.46
Stage 2	10	85.83 (5.62)	75.00-91.67
Stage 3	10	80.00 (4.15)	70.49-85.25
Knowledge Test (Transfer Phase)	3	73.68 (3.65)	69.47-75.79
Stage 1	3	83.33 (5.89)	76.92-88.46
Stage 2	3	72.822(9.62)	66.67-83.33
Stage 3	3	69.59 (2.67)	66.67-71.93

Soft Skills

Soft Skills Measure	<i>n</i>	<i>M</i> (<i>SD</i>)	<i>Range</i>
Skills & Abilities Assessment (Recruitment Phase)	14	73.71 (8.11)	60-92
Role Play	14	71.43 (10.27)	60-93.33
Written Component	14	77.14 (11.39)	60-90
Skills & Abilities Assessment (Training Phase)	7	87.43 (8.14)	72-92
Role Play	7	89.52 (6.51)	80-100
Written Component	7	84.29 (17.18)	50-100
Skills & Abilities Assessment (Transfer Phase)	3	89.33 (8.33)	80-96
Role Play	3	91.11 (10.18)	80-100
Written Component	3	86.67 (11.55)	80-100

Hard Skills

Hard Skills Measure	<i>n</i>	<i>M % (SD)</i>	<i>Range</i>
Training Phase			
Hard Skills (Training Phase)	7	98.57 (.75)	97.48-99.26
Frisk	7	100.00 (0.00)	-
OC	7	100.00 (0.00)	-
Two-way Radio	7	100.00 (0.00)	-
Twistlock Handcuffing	7	99.11 (2.36)	93.75-100.00
Low Profile Escort	7	100.00 (0.00)	-
High Profile Escort	7	97.96 (5.40)	85.71-100.00
Application & Removal of Leg Irons	7	93.28 (2.22)	88.24-94.12
Application & Removal of Body Belt	7	98.21 (3.05)	93.75-100.00
Transfer Phase			
Hard Skills (Transfer Phase)	3	87.23 (6.97)	79.63-93.33
Frisk	3	77.78 (11.11)	66.67-88.89
OC	3	100.00 (0.00)	-
Two-way Radio	3	100.00 (0.00)	-
Twistlock Handcuffing	3	87.50 (6.25)	81.25-93.75
Low Profile Escort	3	83.33 (28.87)	50.00-100.00
High Profile Escort	3	85.71 (14.29)	71.43-100.00
Application & Removal of Leg Irons	3	84.31 (8.99)	76.47-94.12
Application & Removal of Body Belt	3	79.17 (14.43)	62.50-87.50

Attitudes

Attitudes Measure	<i>n</i>	<i>M (SD)</i>	<i>Range</i>
ATCW (Recruitment Phase)	14	86.31 (10.13)	66.67-100
ATCW (Training Phase)	7	92.86 (7.50)	83.33-100
ATCW (Transfer Phase)	3	88.89 (9.62)	83.33-100
ATI (Recruitment Phase)	14	68.21 (8.80)	51.67-86.11
ATI (Training Phase)	7	66.19 (3.56)	61.11-72.22
ATI (Transfer Phase)	3	57.96 (2.10)	55.56-59.44

Trainee Behavioural Checklist – Frequency Items

Behaviour/Skill	<i>n</i>	<i>No Opportunity to Use (n)^a</i>	<i>Median</i>	<i>Mode</i>	<i>Range</i>
Assessing Problematic Situations	3	0	3	2	2-5
Identifying Possible Solutions	3	0	3	2	2-5
Taking Necessary Measures	3	0	2	2	2-5
Motivating and Influencing Individuals	3	0	3	2	2-4
Communicate Effectively Orally	3	0	5	5	4-5
Communicate Effectively in Writing	3	0	3	2	2-4
Summarizing Information	3	0	3	2	2-4
Report Writing:					
• OSOR	3	0	2	2	1-2
• Post Search	3	0	2	2	2-2
• Use of Force	3	1	1	1	1-1
• Offence	3	0	2	2	2-3
Using Communication Equipment (e.g., two-way radio, loudhailer, etc.)	3	0	5	5	5-5
Using Restraint Equipment	3	0	3	3	2-3
Engaging in a Search (e.g., area, frisk, strip, etc.)	3	0	4	2	2-5
Escorts (e.g., internal and outside)	3	0	2	2	2-3
Using Chemical and Inflammatory Agents	3	2	1	1	1-1
Utilizing Fire Safety Measures	3	2	1	1	1-1
Using Self Defence Skills	3	2	1	1	1-1
Using Arrest and Control Skills	3	2	1	1	1-2
Using Baton Skills	3	2	1	1	1-1

Note. ^aTrainees were given the option to indicate that there was no opportunity to use this skill. When this response was selected it was obviously not included in the calculation of the descriptive statistics; however, the number of trainees who provided this response is provided.

Trainee Behavioural Checklist – Quality Items

Behaviour/Skill	<i>n</i>	<i>No Opportunity to Use (n)^a</i>	<i>Mean</i>	<i>SD</i>	<i>Range</i>
Assessing Problematic Situations	3	0	2.00	.00	2-3
Identifying Possible Solutions	3	0	2.33	.58	2-3
Taking Necessary Measures	3	0	2.00	.00	2-2
Motivating and Influencing Individuals	3	0	2.00	.00	2-2
Communicate Effectively Orally	3	0	2.33	.58	2-3
Communicate Effectively in Writing	3	0	2.33	.58	2-3
Summarizing Information	3	0	2.67	.58	2-3
Report Writing:					
• OSOR	3	0	2.67	.58	2-3
• Post Search	3	0	2.00	.00	2-2
• Use of Force	3	1	2.33	1.53	1-4
• Offence	3	0	2.33	.58	2-3
Using Communication Equipment (e.g., two-way radio, loudhailer, etc.)	3	0	2.67	.58	2-3
Using Restraint Equipment	3	0	2.00	.00	2-2
Engaging in a Search (e.g., area, frisk, strip, etc.)	3	0	2.33	.58	2-3
Escorts (e.g., internal and outside)	3	0	2.67	.58	2-3
Using Chemical and Inflammatory Agents	3	2	3.00	1.73	1-4
Utilizing Fire Safety Measures	3	2	3.00	1.73	1-4
Using Self Defence Skills	3	2	3.00	1.73	1-4
Using Arrest and Control Skills	3	2	3.00	1.73	1-4
Using Baton Skills	3	2	3.00	1.73	1-4

Note. ^aTrainees were given the option to indicate that there was no opportunity to use this skill. When this response was selected it was obviously not included in the calculation of the mean; however, the number of trainees who provided this response is provided.

LTSI

LTSI Scale	<i>M (SD)</i>	<i>Range</i>
Learner Readiness	3.75 (.25)	3.50-4.00
Motivation to Transfer	4.50 (.66)	3.75-5.00
Positive Personal Outcomes	2.89 (.51)	2.33-3.33
Negative Personal Outcomes	2.75 (.43)	2.25-3.00
Personal Capacity for Transfer	3.50 (.66)	3.00-4.25
Peer Support	3.17 (.38)	2.75-3.50
Supervisor Support	2.72 (.48)	2.17-3.00
Supervisor Opposition	2.56 (.19)	2.33-2.67
Perceived Content Validity	3.73 (.31)	2.80-4.40
Transfer Design	3.92 (.72)	3.50-4.75
Opportunity to Use	3.08 (.52)	2.50-3.50
Transfer Effort-Performance Expectations	3.83 (.52)	3.25-4.25
Performance-Outcomes Expectations	3.73 (.31)	3.40-4.00
Resistance-Openness to Change	2.61 (.48)	2.33-3.17
Performance Self-Efficacy	3.92 (.14)	3.75-4.00
Performance Coaching	3.50 (.25)	3.25-3.75
Intent to Transfer	4.08 (.14)	4.00-4.25

Appendix AA

Correlation Matrices for the Individual Hard Skills

Training Phase

Variables	1	2	3	4	5	6	7	8
1. Frisk Search	-							
2. OC Spray	.05	-						
3. Radio	.06	-.13	-					
4. Handcuffing	.17	.02	.22	-				
5. Low Profile Escort	-.16	-.11	.12	.39*	-			
6. High Profile Escort	.05	.01	.11	.47*	.45*	-		
7. Leg Irons	.17	.25	.19	.29	.31	.04	-	
8. Body Belt	.34	-.07	.07	.06	.02	.10	.35	-

Note. Bonferroni correction = $.05/28 = .002$.

*Correlations significant at the corrected alpha level.

Transfer Phase

Variables	1	2	3	4	5	6	7	8
1. Frisk Search	-							
2. OC Spray	.07	-						
3. Radio	.11	.15	-					
4. Handcuffing	.34	.11	.08	-				
5. Low Profile Escort	-.13	-.01	-.15	.23	-			
6. High Profile Escort	.25	.11	.01	.23	.43*	-		
7. Leg Irons	.40	.08	.22	.34	-.02	.03	-	
8. Body Belt	-.10	-.07	.20	.09	-.22	-.16	.09	-

Note. Bonferroni correction = $.05/28 = .002$.

*Correlations significant at the corrected alpha level.

Appendix BB

Reponses to the Open-Ended Items from the Trainee Behavioural Checklist

Question 3:

Question three asked trainees, “Have you used/reviewed the training materials you received for CTP 2008 since you finished the training program?” Officers responded by checking one of two boxes, which either indicated yes they had used/reviewed the training materials or no they had not. The results are displayed below.

Region	Yes	No	No Answer
Pacific (<i>n</i> = 25)	11	12	2
Ontario (<i>n</i> = 23)	11	12	0
Quebec (<i>n</i> = 3)	1	2	0
Total	23	26	2

As a follow up to question three trainees were asked “Please explain why or why not”. This open ended portion of question was coded using the following categories:

	Category	Explanation	Example
If No...	No relevance	Trainee did not find a need to read/use the material because it was not considered relevant.	- “no need”
	No time	Trainee did not have the time to either review or use the training materials	- “no time”
	Other	Not classifiable into other 2 categories	- “constant learning on the job”
If Yes...	A general review	Trainee reviewed/ used material in a general sense	- “I like to review my material to ensure that my skills and knowledge are consistent”
	Specific review	Trainee reviewed/used specific material	- “Review of Inflammatory Spray”
	Other	Not classifiable into other 2 categories	- “A few times since September”

The following table provides the frequency of comments in each category, organized by region:

Region	No			Yes			No Answer
	Relevance	Time	Other	General	Specific	Other	No Answer
Pacific	8	1	3	7	4	0	2
Ontario	4	3	4	5	3	1	3
Quebec	1	0	1	1	0	0	0
Total	13	4	8	13	7	1	51

Of those officers who indicated that they did not use or review the training materials since completing the training program, most indicated no relevance as their reason for not doing so, followed by other reasons and no time. Of those who indicated that they had used or reviewed the training material, most reported doing so for general review purposes. The second most common reason was specific review followed by other reasons. The same overall trends were observed within the Pacific region. Within the Ontario region, trainees who had not reported using or reviewing the training materials equally reported not making use of them because of lack of relevance and other reasons. Otherwise, the same trends were observed in the Ontario Region.

Question 4:

Question 4 asked trainees to: "List the 3 behaviours you have used most frequently on the job as a result of CTP". This open –ended question was coded using the following categories. Most of the categories are self-explanatory, and many trainees responded with one word answers. The following table provides examples of comments coded into each category.

Category	Example
Communication/ Listening	– “Oral Communication” – “Listening Actively”
Security	– “Static security” – “Dynamic security”
Firearms	– “Firearms” – “Weapons checks”
Searches	– “Searching” – “Frisk searching”
Personal Safety	– “Self defence/personal safety” – “Personal safety”
Problem Solving	– “Finding solutions” – “Assessing”
Teamwork/ Cooperation	– “Teamwork” – “Collaboration”
Reports/ Writing	– “Writing” – “OSOR/Post search reports”
Observation	– “Observing” – “Awareness”
Use of equipment	– “Using hand held radio” – “Hand cuffing”
Other not classifiable	– “Complying” – “Active learning”

Category	Pacific (<i>n</i> = 25)	Ontario (<i>n</i> = 23)	Quebec (<i>n</i> = 3)	Total
Communication/ Listening	9	15	1	25
Security	9	1	0	10
Firearms	2	0	1	2
Searches	5	3	0	9
Personal Safety	3	2	0	5
Problem Solving	7	7	1	15
Teamwork/Cooperation	3	0	0	3
Reports/Writing	3	0	2	5
Observation	4	5	0	9
Use of equipment	6	2	2	10
Other not classifiable	3	5	2	10

Collapsed across all regions, communication/ listening was reported as the behaviour used most commonly on the job as a result of CTP. Over half of Ontario trainees reported engaging in communication/ listening behaviours. The second most commonly used behaviour was problem solving. Use of equipment, security, and other not classifiable behaviours were reported as the third most commonly used behaviours. Firearm behaviour was the least reportedly used behaviour. Trainees from the Pacific region equally reported engaging in communication/ listening behaviours and security behaviours. No trainees from the Ontario region reported use of firearms, teamwork/ cooperation, and reports/writing as frequently engaged in behaviours.

Question 5:

Question 5 asked officers, “What barriers, if any, have you encountered that have prevented you from using the skills or knowledge gained in CTP?” This open-ended question was coded using the following categories. Examples are provided.

Category	Example
None – no barriers encountered	<ul style="list-style-type: none"> – “Nothing holding me back” – “None”
Coworkers/ Peers/ Staff	<ul style="list-style-type: none"> – “Staff ideas of CTP being useless or not very good at training unless trained at institution” – “Fellow employees that have a negative attitude towards work and inmates and seem to thrive on being confrontational. Also people that think you should never help an inmate”
Supervisors/ Managers	<ul style="list-style-type: none"> – “I was told by a CM that in regard to Use of force that we do not follow the Criminal Code, but only the CCRA’s. And told to do a quickie pat down”
CTP Training (not useful or not realistic)	<ul style="list-style-type: none"> – “Many scenarios in CTP were very unlike to the real thing. In CTP it was usually defused by talking easily” – “Dealing with inmates is much different than scenarios at the college”
No opportunity to use skills	<ul style="list-style-type: none"> – “Have not been in many situations to use them” – “Have not been posted to areas that would require the skills to be used”
Other, not classifiable	<ul style="list-style-type: none"> – No comments were found that were not classifiable into one of the listed categories.

The following table provides the frequency of comments in each category, organized by region:

Region	No Barriers	Coworkers	Supervisors	CTP training	No opportunity
Pacific	3	8	1	2	6
Ontario	4	3	0	7	4
Quebec	0	1	0	1	1
Total	7	12	1	10	11

The barrier most often reported as preventing officers from using their skills or knowledge gained in CTP was their coworkers. No opportunity to use skill was the second most common barrier and the third was CTP training (not preparing them or not being realistic). These overall trends were apparent within the Pacific region. On the other hand, Ontario trainees reported CTP training (not preparing them or not being realistic) as their most common barrier. No opportunity to use knowledge or skills was their second most reported barrier, but an equal number of Ontario trainees reported not having encountered any barriers.

Question 6:

Question 6 asked officers, “What specific suggestions do you have for improving the training delivered in CTP?” This open-ended question was coded using the following categories. Examples of types of comments coded into those categories are provided in the table below.

Category	Example
Instructors	– “The instructors – some were amazing, others real self-centered, all about them”
	– “Instructors should be stricter and require more of recruits”
Content	– “Some of the lessons were too lengthy. The material was covered and sometimes over explained”
	– “Institution specific training”
Compensation	– “Compensation for the training would be a plus for the overall success of this training”
	– “Being paid for the training would take a lot of stress out of the first 6 months on job. Playing catch up is hard with a family”
Other not classifiable	– “Keep reading materials”
General	– “Keep it up. Train them to be perfect and hope for the best”
Program	– “Very good training in general”
None	– “None, great experience”

The following table provides the frequency of comments in each category, organized by region:

Region	Instructors	Content	Compensation	Other	General	None
Pacific	2	13	1	1	3	4
Ontario	2	9	2	2	3	2
Quebec	0	1	1	0	2	0
Total	4	23	4	3	8	6

Overall, officers primarily provided content related suggestions for improving the training delivered in CTP. Over half of the respondents from Pacific and less than half of Ontario respondents provided content related suggestions. Comments about the general program were the second most common types of comments in this section, followed by comments which indicated that the officer had no suggestions for improving the training. For the Pacific region, no suggestions for improvement were the second most common type of comments and general comments were the third most common. An equal proportion of Ontario trainees provided suggestions about instructors, compensation, other not classifiable, and no suggestions for improvement.

Appendix CC

Correlation Matrix for the Scales of the LTSI

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Learner Readiness	-																
2. Motivation to Transfer	.17	-															
3. Personal Outcome Positive	.35	.33	-														
4. Personal Outcome Negative	.51*	.20	.53*	-													
5. Personal Capacity	.18	.23	.30	.27	-												
6. Peer Support	.18	.24	.45	.37	.35	-											
7. Supervisor Support	.36	.25	.52*	.43	.20	.59*	-										
8. Supervisor Opposition	-.25	-.14	-.08	-.42	-.06	-.26	-.16	-									
9. Content Validity	.31	.39	.29	.27	.26	.35	.44	.00	-								
10. Opportunity to use	.02	.07	.20	.19	.25	.36	.32	-.19	.47*	-							
11. Transfer Effort	.24	.57*	.39	.41	.29	.50*	.58*	-.29	.60*	.41	-						
12. Performance Outcome	.26	.33	.42	.39	.20	.53*	.65*	-.35	.47	.43	.56*	-					
13. Resistance-Openness to Change	-.06	.16	.28	.12	.13	.57*	.61*	-.09	.30	.44	.42	.53*	-				
14. Personal Self-Efficacy	.17	.36	.25	.21	.22	.38	.41	.04	.49*	.46	.50*	.34	.44	-			

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
15. Performance Coaching	.20	.21	.30	.16	.11	.28	.39	.14	.47*	.14	.24	.20	.47*	.44	-		
16. Intent to Transfer	.28	.34	-.10	.12	.20	.10	.15	-.23	.41	.27	.40	.13	-.14	.39	-.13	-	
17. Transfer Design	.33	.23	.12	.18	.04	.20	.38	.01	.58*	.36	.35	.34	.10	.40	.08	.55*	-

Note. Bonferroni correction = $.05/136 = .0004$.

*Correlations significant at the corrected alpha level.