

Technical Report: September 2019 CHRP ELE

Human Resources Professionals Association

3 October 2019



Contents

Executive Summary	5
Administration	6
Form Setting	6
Testing Window	7
Analysis	8
Data Cleaning and Integrity Checks.....	8
Post-Examination Survey.....	10
Initial Analysis	13
Key Validation.....	13
Establishing the Pass Mark: Equating	16
Scoring	26
Key Examination Metrics	28
Related Development Activities	29
Appendix	30
Blueprint	30

List of Tables

Table 1: Domain fit at administration	6
Table 2: Cognitive level fit at administration	7
Table 3: Administration-related post-examination survey questions*	11
Table 4: Content-related post-examination survey questions*	11
Table 5: Preference regarding independent and case-based items.....	12
Table 6: Preference regarding number of response options	12
Table 7: Preference regarding computer-based testing versus pencil-and-paper	12
Table 8: Value of access to legislation	12
Table 9: Initial examination statistics	13
Table 10: CHRP Examination Validation Committee – Key validation	14
Table 11: Domain fit for final scored items	15
Table 12: Cognitive level fit for final scored items.....	15
Table 13: Item type fit for final scored items	15
Table 14: Anchor item fit to blueprint – To January 2019	17
Table 15: Equating parameter table – To May 2019.....	17
Table 16: Equating outcome table – To May 2019	18
Table 17: Anchor item fit to blueprint – To May 2018	19
Table 18: Equating parameter table – To May 2018.....	19
Table 19: Equating outcome table – To May 2018	20
Table 20: Anchor item fit to blueprint – To January 2019	21
Table 21: Equating parameter table – To January 2019.....	21
Table 22: Equating outcome table – To January 2019	22
Table 23: Anchor item fit to blueprint – To September 2018.....	23
Table 24: Equating parameter table – To September 2018	23
Table 25: Equating outcome table – To September 2018.....	24
Table 26: Equating outcome table – Combined results	25
Table 27: Historical pass rates	25
Table 28: CHRP Examination Validation Committee – Pass mark approval.....	26
Table 29: Total and domain scores for all candidates.....	27
Table 30: Correlations between functional area scores for all candidates	27

Table 31: Key examination metrics – Candidates included in analysis only.....28

Table 32: CHRP Employment Law Examination Blueprint Structural Variables.....30

Table 33: CHRP Employment Law Examination Blueprint Content Weights32

Table 34: CHRP Employment Law Examination Blueprint Cognitive Level Weights.....33

List of Figures

Figure 1: Examination time distribution for all candidates..... 9

Figure 2: Candidate volume and score trends across testing window 9

Figure 3: Score distribution for all candidates.....27

Executive Summary¹

Note that this technical report covers only the primary new form or forms administered during an administration, and not detailed results for all forms used (which may include previously used forms, scrambled forms, and other modifications to maintain exam and score integrity).

The CHRP Employment Law Exam (CHRP ELE) was administered to 114 candidates using computer-based testing at Prometric test centres September 9–23, 2019, inclusive. The examination comprised 110 three-option multiple choice items and had a 3½-hour time limit.

As per the CHRP ELE blueprint, the exam was scored using the 98–102 best-performing items (while adhering to the prescribed distribution across topics). The mean score for first-time candidates² ($n=98$) was 72.7 (72.7%), and for all candidates it was 71.1 (71.1%), out of 100 validated items for scoring. Reliability was marginal at .77 (noting that there is substantial range restriction with these candidates). The final set of scored items adhered to the blueprint parameters.

The pass mark was set using equating back to the May 2018, September 2018, January 2019, and May 2019 administrations, yielding an integer pass mark of 57. Equating was conducted to compensate for minor changes in exam form difficulty so that any given candidate has an equivalent hurdle regardless of when they write the CHRP ELE. This pass mark resulted in a pass rate for first-time candidates of 95.9% and a pass rate for all candidates of 89.5%.

This report, the analyses performed, and the processes followed are consistent with NCCA standards³ and ISO 17024 standards.⁴

¹ This technical report is an abbreviated version of the full report. Information has been excluded that if known to candidates could negatively affect the validity of future candidate test score interpretations. This includes item-level statistics, some information about the construction of test forms, and some specific details concerning equating.

² Excludes those who had failed an HRP A employment law examination in the past, who were identified as being statistical outliers, or who had written an alternative test form.

³ National Commission for Certifying Agencies (2014). *Standards for the accreditation of certification programs*. Washington, DC: Institute for Credentialing Excellence.

⁴ International Organization for Standardization (2012). *ISO/IEC 17024:2012 Conformity assessment – General requirements for bodies operating certification of persons*. Geneva: International Organization for Standardization.

Administration

Form Setting

Using only validated test items, Wickett Measurement Systems prepared one 110-item test form. Wickett constructed the final test form according to the following parameters:

1. Including only items validated by the validation panel in the past 2 years
2. Fitting the total item count of 110
3. Excluding enemy items
4. Matching the blueprint weights
5. Maximizing spread across subtopics as per the blueprint weights
6. Reducing item exposure
7. Selecting items with perceived psychometric effectiveness, using statistics from previous administrations as available

The final form was reviewed for currency and enemy items by Roxanne Chartrand and Karen Weiler (CHRP Examination Validation Committee members) in a remote session held June 17, 2019.

The final form composition for the September 2019 CHRP ELE is shown in Table 1 (domain weighting) and Table 2 (cognitive level weighting). The form reflected the examination blueprint (see Appendix for full CHRP ELE blueprint).

Note that at any administration, HRPAs make use of previously validated and administered test forms along with new test forms, in addition to employing other mechanisms to maintain the integrity of the exams and candidate scores.

Table 1: Domain fit at administration

Domain	Actual Items	Target Range	Target Items	Variance
A Employment Contracts and Terminations	50	46% ± 5%	46–56	—
B Employer Obligations	38	33% ± 4%	32–40	—
C Regulations and Legislation	22	21% ± 3%	20–26	—
TOTAL	110		110	—

Table 2: Cognitive level fit at administration

Cognitive Level	Actual Items	Target Range	Target Items	Variance
Knowledge	11	10% ± 3%	8–14	—
Application	63	60% ± 10%	55–77	—
Critical thinking	36	30% ± 10%	22–44	—
TOTAL	110		110	—

The test form adhered to the blueprint for content domain and cognitive level.

Testing Window

The examination was administered via computer-based testing at Prometric test sites primarily in Ontario. The testing window was September 9–23, 2019, inclusive, and 114 candidates wrote the exam.

Candidates had access to a basic-function calculator on screen and access via PDF to 10 pieces of searchable legislation (compiled into 2 documents):

Provincial

- AODA – *Accessibility for Ontarians with Disabilities Act, 2005*
- ESA – *Employment Standards Act, 2000*
- LRA – *Labour Relations Act, 1995*
- OHRC – *Human Rights Code*
- OHSA – *Occupational Health and Safety Act*
- PEA – *Pay Equity Act*
- WSIA – *Workplace Safety and Insurance Act, 1997*

Federal

- CHRA – *Canadian Human Rights Act*
- CLC – *Canada Labour Code*
- PIPEDA – *Personal Information Protection and Electronic Documents Act*

The versions of the legislation were as accessed on June 24, 2019.

Analysis

Data Cleaning and Integrity Checks

Prometric provided data in .xml format via a secure FTP site. Candidate files were provided as candidates completed the examination throughout the testing window. These files were extracted to Microsoft Excel for processing. They contained identifying information for each candidate, form information, start and stop times, answer string, key string, candidate total score, item comments if the candidate made any, and time spent per item.

The data files received were reconciled against the roster provided by Prometric to ensure that all .xml files had been received. Further, each candidate total score as computed by Prometric was reconciled with that computed by Wickett for the full set of 110 items to verify key accuracy. Comments on items were also reviewed to identify any specific item-level issues. No problems were encountered.

Due to an administration scheduling error, 3 candidates wrote outside the testing window and were not included in the preliminary analyses.

The average time taken by all candidates was assessed to detect potential examination timing concerns. The distribution is shown in Figure 1. The mean was 2 hours, 59 minutes (up 1 minute from May 2019). The time limit on the CHRP ELE was 3½ hours, suggesting that time may have been a factor for at least some candidates.

Fifteen candidates (14%) took the full 3½ hours, suggesting that those candidates may have wanted more time, and 2 candidates (2%) left at least one item blank, suggesting that those candidates timed out of the exam before being able to complete it. These metrics will continue to be monitored, but presently they do not appear problematically high, though they are not as low as would be preferred. Compared with the May 2019 administration, there was little change in observed values. Note that because they have access to legislation, candidates may take more time than intended by researching more answers. This may skew time metrics higher.

The correlation between scores on the 110 items and time spent writing the examination was essentially zero at a value of .01, suggesting that time constraints were not generally related to candidate performance.

Candidate scores were computed across the window to look for any evidence of item exposure. As shown in Figure 2, there was little variation across the window, and the difference between the first 3 days and the last 3 days was a modest increase of 0.5 marks out of 110 (though there were few candidates overall so this analysis lacks power to identify a significant change).

As a matter of interest, candidate volumes were also examined across the window; these are also shown in Figure 2. The usual peak in volumes towards the end of the window was observed.

Figure 1: Examination time distribution for all candidates

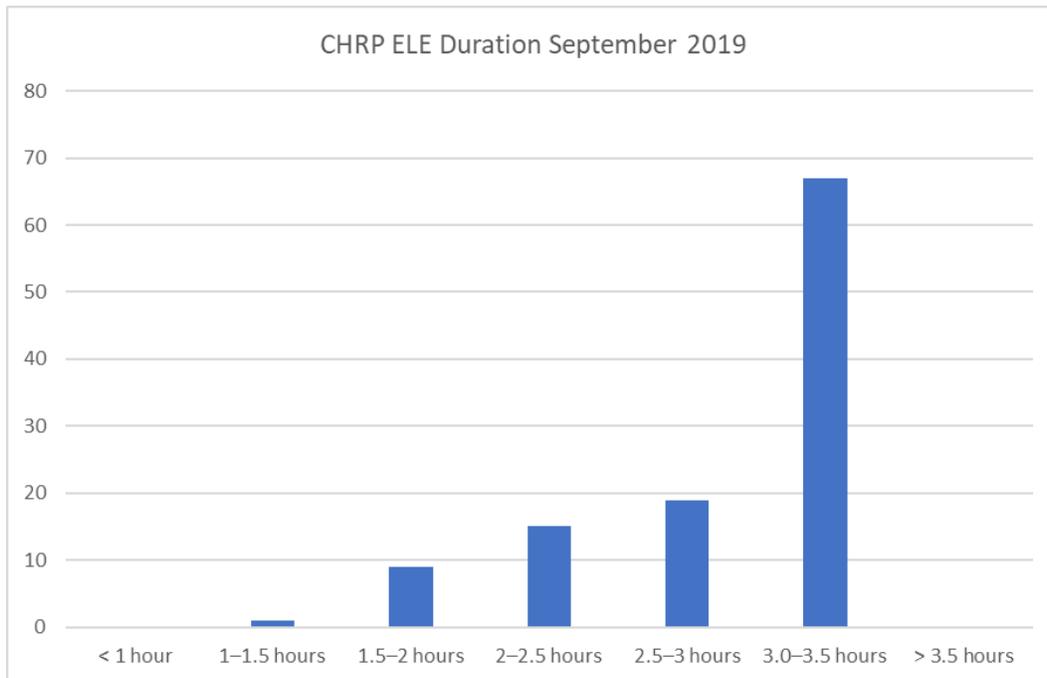
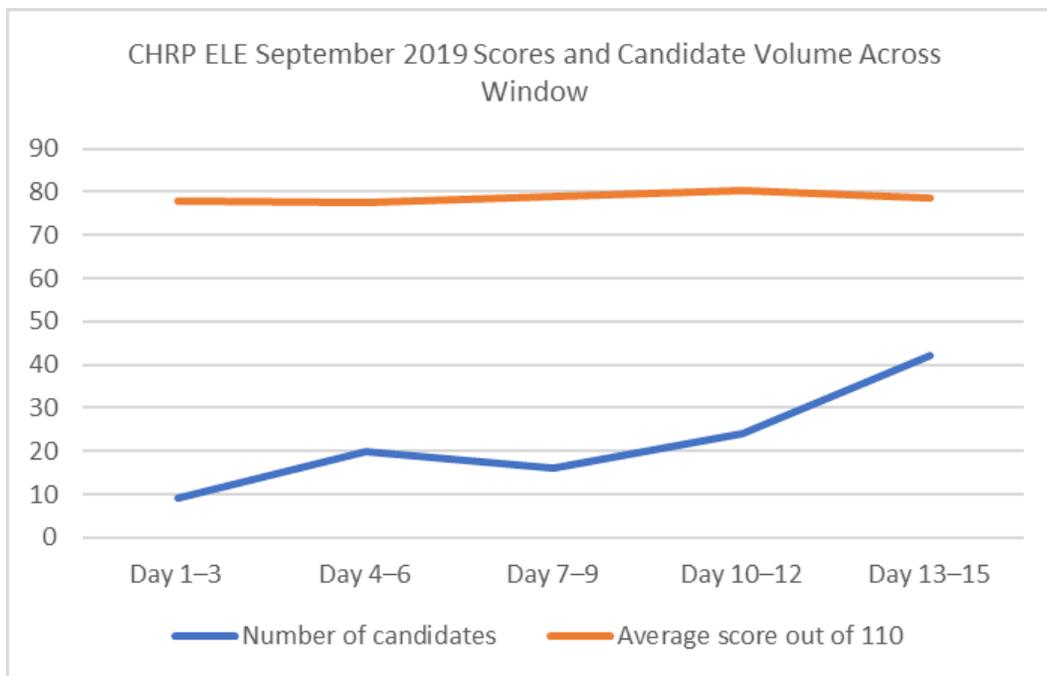


Figure 2: Candidate volume and score trends across testing window



After removing candidates who were administered a previously used test form (who were scored using the same decisions employed at the time that form was originally used), scores were calculated for all remaining candidates based on the full set of 110 items. One candidate who was unable to finish the test due to illness was removed from analysis. No other candidates

were flagged for an abnormally low or high score (z value outside ± 3.0). Also, the 110 items were arbitrarily broken into 4 blocks of 25 items for each candidate plus 1 final block of 10 items; the 5 resulting subscores for each candidate were evaluated for outliers as well. For candidates with any subscore more than 3 standard deviations (SD) from their average z -score, the .xml file was examined closely for any issues. No instances were identified and so no candidates were removed from analysis. Candidates with abnormal response patterns (such as having 5 or more blanks) were removed. To be conservative, candidates who had been granted a testing accommodation were also removed from the main analysis (simply because their testing conditions were not the same as those of the main group of candidates, even though each accommodation was granted on the premise that it would make the testing experience equivalent in terms of opportunity to demonstrate competence). As a result of these factors, 3 candidates were removed from analyses.

Candidates who had failed a previous employment law examination (CHRP ELE or CHRL ELE; $n=10$) scored lower than did those who had not (60.1 and 73.1, respectively, on the full exam of 110 items; $t(11)=5.19$, $p<.01$). In keeping with standard procedures, these candidates were removed from subsequent analyses. The CHRP ELE analysis proceeded with 98 candidates.

Owing to the modest number of candidates, all subsequent analyses were interpreted with caution.

Post-Examination Survey

Candidates were provided access to the post-examination survey immediately after submitting their responses to the CHRP ELE; 105 candidates responded (response rate, 95%).

Table 3 shows the responses to the administration-related questions. Note that candidates were generally positive about the administration experience. Table 4 shows the content-related questions; there was a tendency to more neutrality on these questions.

Table 3: Administration-related post-examination survey questions*

	Question	SA	A	N	D	SD	Score	Agreement
1.	I was able to book a seat to write the examination at a time that was convenient for me.	48	43	5	8	1	4.23	87%
2.	I was well informed about what documents to bring to the exam location.	63	38	1	0	2	4.54	97%
3.	Proctors enforced the exam-day rules and the security procedures at the test centre were what I expected.	65	35	1	1	0	4.61	98%
4.	Proctors were professional and courteous.	67	28	4	1	2	4.54	93%
5.	The tutorial helped me understand how to complete the examination on the computer.	58	36	6	1	0	4.50	93%
6.	The legislation and case texts were easy to access during the examination.	52	31	12	6	1	4.25	81%
7.	Navigation through the examination was easy and intuitive.	58	37	2	3	0	4.50	95%

*Response categories: SA = Strongly Agree; A = Agree; N = Neutral; D = Disagree; SD = Strongly Disagree.

Table 4: Content-related post-examination survey questions*

	Question	SA	A	N	D	SD	Score	Agreement
8.	The time allotted for this examination was sufficient.	38	44	7	9	3	4.04	81%
9.	Information available prior to exam day provided me with adequate details about the content and format of the exam.	29	41	15	12	3	3.81	70%
10.	I feel I was adequately prepared to write this examination.	16	47	26	10	1	3.67	63%
11.	The questions in the examination were clearly written.	16	46	20	14	4	3.56	62%
12.	The terminology used in the examination was accurate.	19	62	12	7	0	3.93	81%
13.	The situations presented in the examination were realistic.	24	65	8	3	0	4.10	89%
14.	The questions in the examination reflected the Employment Law Examination blueprint.	18	46	29	5	2	3.73	64%
15.	The examination was a fair assessment of my ability.	11	43	29	13	2	3.49	55%

*Response categories: SA = Strongly Agree; A = Agree; N = Neutral; D = Disagree; SD = Strongly Disagree.

Candidates were asked their opinions regarding several structural variables; these results appear in Table 5 through Table 8. Candidates reported a preference for independent items and a strong preference for 3-option multiple choice items. Most indicated that taking the test on a computer likely had no effect on their performance or that it was a benefit. Most indicated that access to the legislation was necessary, whether they consulted it a few times or often.

Table 5: Preference regarding independent and case-based items

	Count	%
I preferred the independent items.	48	49%
I preferred the case-based items.	23	23%
I had no preference between independent and case-based items.	27	28%

Table 6: Preference regarding number of response options

	Count	%
I preferred having 3 options.	86	88%
I preferred having 4 options.	3	3%
It did not matter to me how many options were used.	9	9%

Table 7: Preference regarding computer-based testing versus pencil-and-paper

Question	Count	%
I feel that completing the examination on a computer improved my performance.	45	46%
I feel that completing the examination on a computer decreased my performance.	10	10%
I feel that completing the examination on a computer had no effect on my performance.	43	44%

Table 8: Value of access to legislation

	Count	%
Yes, it was essential to me in completing the examination.	37	38%
Yes, but I only consulted it a few times.	41	42%
No, I could not find the answers to questions I had.	9	9%
No, I did not need to consult it to complete the examination.	5	5%
No, it was more of a distraction than an aid.	6	6%

An open-ended question was also posed to candidates asking for any additional comments. Those comments were provided to HRP A for information and consideration. Nothing actionable with respect to scoring emerged in these comments.

Initial Analysis

The full CHRP ELE examination was 110 items, of which approximately 100 were to be scored. The remain 8–12 items were designated as experimental. However, because only 1 new form was administered, all items were potentially available for scoring and the focus of subsequent item analysis and key validation was on determining the best set of approximately 100 items that still reflected the examination blueprint.

The initial analysis summary statistics are presented in Table 9.

Table 9: Initial examination statistics

Index	CHRP ELE
Items	110
Total candidates	114
Candidates in analysis	98
Mean	80.3 (73.0%)
Standard deviation	8.53
Range	61–97 (55.5–88.2%)
Cronbach's alpha	.75
Disattenuated alpha	.85
Mean r_{pb}^*	.14

Standard classical test theory analysis was conducted to identify the following:

1. Item difficulty (percent obtaining correct result, p)
2. Item discrimination (corrected point-biserials, r_{pb}^*)
3. Distractor quality (based primarily on distractor discrimination)

Wickett compiled these statistics, along with any comments made by candidates concerning flagged items, to identify items that may have been keyed incorrectly or that were performing poorly. Most emphasis was placed on the corrected point-biserials as evidence of item quality, after removing items with difficulty values at the extremes. Items were ranked from worst performing to best performing accordingly.

Key Validation

Key validation was conducted via web meeting on September 27, 2019, using members of the CHRP Examination Validation Committee (EVC). The group (Table 10) was first reminded of the

methods used for key validation and was oriented to the main statistics used to evaluate the quality of the CHRP ELE.

Table 10: CHRP Examination Validation Committee – Key validation

Member	Credential	Years of Relevant Experience	Start on EVC	Industry
✓ Sunday Ajao	CHRL	15–20	2017	Banking/Finance
Roxanne Chartrand	CHRL	20–29	2018	Insurance
Claire Chester	CHRL	10–15	2017	Regulation/CPA
Tanya Gopaul	CHRL	10–15	2017	Banking
Jean Lazarus	CHRL	15–19	2017	Health services
Suman Seth	CHRL	15–19	2018	Public sector
Kriss Stone	CHRP	10–15	2017	Real estate
✓ Ielean Tait	CHRL	15–20	2017	Environmental
✓ Patricia Verkley	CHRL	10–15	2019	Not-for-profit
Karen Weiler	CHRL	20–29	2017	Software/ Communications
Alyssa Young	CHRL	5–9	2017	Non-profit

✓ Participated in the session.

The group was informed that test reliability, as measured by Cronbach’s alpha, was .75 based on the set of 110 potentially scored items and that this was below the generally accepted threshold of .80. The group was advised that restriction of range was considered the most likely basis for the lower value. They were also informed that part of the goal of the key validation review was to bring this value up if possible.

The group was walked through the flagged items one at a time, with the recommendation that the worst-performing items be removed from scoring, but were given less direction on those with borderline statistics. Where available, candidates’ comments about the items were also shown. Further, historic data on items was used to help in making decisions such that items with strong statistics on past administrations were more likely to be retained for scoring. The group made decisions based on content and the data through discussion; they removed 10 items that they felt were inappropriate to retain for scoring. Panel members’ comments about specific items were recorded for future item revision activities.

Not all remaining items were strong-performing, and several items were retained that were very easy or very hard or that had a low corrected point-biserial. Most were moderate to strong items, however. The final alpha for the set of 100 scored items was .77. The difficulties ranged from 38.8% to 94.9%, with a mean of 72.7%. The r_{pb}^* values ranged from $-.12$ to $.47$, with a

mean of .16. Note that with a small sample of candidates, negative point-biserial values are not necessarily a sign of a problematic item, and items that have performed well in the past were more likely to be retained even if showing a poor point-biserial in this candidate sample.

Table 11 shows the scored CHRP ELE's final fit to the domain weighting. Table 12 shows the same for cognitive level, and Table 13 shows the same for item type. The exam fit on all dimensions.

The group approved the final set of items for use in scoring the September 2019 CHRP ELE candidates.

Table 11: Domain fit for final scored items

Domain	Actual Items	Target Range	Target Items	Variance
A Employment Contracts and Terminations	46	46% ± 5%	41–51	—
B Employer Obligations	34	33% ± 4%	29–37	—
C Regulations and Legislation	20	21% ± 3%	18–24	—
TOTAL	100		100	—

Table 12: Cognitive level fit for final scored items

Cognitive Level	Actual Items	Target Range	Target Items	Variance
Knowledge	10	10% ± 3%	7–13	—
Application	58	60% ± 10%	50–70	—
Critical thinking	32	30% ± 10%	20–40	—
TOTAL	100		100	—

Table 13: Item type fit for final scored items

Item Type	Actual Items	Target Range	Target Items	Variance
Independent	25	25% ± 3%	22–28	—
Case	75	75% ± 3%	72–78	—
TOTAL	100		100	—

Establishing the Pass Mark: Equating

Equating, as per Kolen and Brennan (2014)⁵ and Livingston and Kim (2009),⁶ was used to establish the pass mark for the September 2019 CHRP ELE. The goal of this process was to set a pass mark for the September 2019 CHRP ELE that would be equivalent to that set for previous administrations; that is, to set a pass mark that would give each candidate the same probability of passing regardless of which form they took.

The passing standard for the CHRP ELE was last set after the January 2018 offering of the CHRP ELE using the Modified Angoff and Bookmark methods. Specific information on the standard-setting session is provided in the Technical Report issued for the January 2018 administration.

Four equating procedures were conducted back to different administrations (May 2018, September 2018, January 2019, and May 2019). The intention following these 4 equating runs was to average them to arrive at a final pass mark for the September 2019 CHRP ELE. These administrations were chosen as the most recent administration (May 2019), the administration from 1 year in the past (May 2018 was selected because of previously observed variations in the statistics derived from the September 2018 administration), and the remaining two selected because of modest variance in the first two equating runs.

Note that because the September 2019 administration was analyzed with 98 candidates, it falls just under the usual threshold of 100 for opting for a linear equating solution (usually the Tucker method). Because of the proximity to that threshold, and historic tendency for Tucker to show the optimal solution, Tucker continued to be favoured in the ensuing analyses though all methods were assessed for viability.

Equating Back to the May 2019 Administration

Linear equating (Tucker) was the chosen method for setting the pass mark and it was conducted once key validation was complete. Linear equating is the primary method considered with more than 100 candidates; equipercentile equating would have been considered with more than 1,000 candidates. With candidate samples of fewer than 100, mean or circle arc equating is most prudent. As already noted, though the September sample was just under 100 candidates, linear equating continued to be favoured.

All candidates in the analysis (i.e., no repeat candidates or outliers) were used in the equating process. Delta-plot analysis was used to identify anchor items showing substantial deviations (generally, although not exclusively, greater than 3 *SD* units) from expected difficulty values, with an emphasis on establishing an anchor set with difficulty equivalent to that of the full form that adhered to the blueprint. Further, items with very high or low difficulty values and those with low corrected point-biserials were also flagged for potential removal from the anchor set. The goal was a strong midi-test (i.e., moderate range of difficulty, moderate to high discrimination, fit to blueprint) of sufficient length to estimate candidate ability.

⁵ Kolen, M.J., & Brennan, R.L. (2014). *Test equating, scaling, and linking*. New York, NY: Springer.

⁶ Livingston, S.A., & Kim, S. (2009). The circle-arc method for equating in small samples. *Journal of Educational Measurement*, 46, 330-343.

The selected set of anchor items had a mean difficulty of 0.72 and a mean corrected point-biserial of .22.

Table 14 shows the fit of the set of anchor items to the blueprint, as percentages. The actual counts are reasonably aligned with the targets and reflect the scope and approximate weighting across the full exam.

Table 14: Anchor item fit to blueprint – To January 2019

	Area	Actual	Target
A	Employment Contracts and Terminations	45%	46%
B	Employer Obligations	34%	33%
C	Regulations and Legislation	21%	21%

The mean, Tucker, Levine observed-score, circle arc, and Braun-Holland methods were computed to ascertain concordance of solutions. Given the sample sizes and similarities of test parameters, Tucker was considered the primary method.

Table 15 shows some of the parameters used to derive the equating estimates, along with other parameters describing the test forms. Of note is that on the anchor items, the population taking the September 2019 CHRP ELE scored lower than the population taking the May 2019 CHRP ELE (72.4% vs. 73.9%, respectively; $t(254)=0.96, ns$). Because the September 2019 CHRP ELE candidates were of modestly lower ability (based on the anchors, non-significance notwithstanding), they should have a modestly lower pass rate (though with small samples the tails of the distribution will not be regular).

The equating analysis generally did bear this out (Table 16). The methods show a pass mark of 56 to 59. Given the sample sizes involved, Tucker would be the primary method under consideration and the equated value of 57.06 was carried forward in the analysis. Note that with the smaller sample of candidates in September 2019, the circle arc value of 55.26 was also considered relevant from this equating run.

Table 15: Equating parameter table – To May 2019

		May 2019	Sep. 2019
N		158	98
Scored items		102	100
Mean score	Total	73.2%	72.7%
	Anchors	73.9%	72.4%

Table 16: Equating outcome table – To May 2019

Method	Pass Mark		Pass Rate	
	Precise	Integer	All	First Time
Equated May 2019	55.97	56	95.9%	96.8%
Tucker	57.06	58	87.4%	93.9%
Levine observed	58.12	59	87.4%	93.9%
Mean	54.99	55	95.5%	98.0%
Circle Arc 1	55.26	56	91.0%	95.9%
Circle Arc 2	55.26	56	91.0%	95.9%
Braun-Holland	56.97	57	89.2%	95.9%

Equating Back to the May 2018 Administration

Linear equating (Tucker) was the chosen method for setting the pass mark and it was conducted once key validation was complete. Linear equating is the primary method considered with more than 100 candidates; equipercentile equating would have been considered with more than 1,000 candidates. With candidate samples of fewer than 100, mean or circle arc equating is most prudent.

All candidates in the analysis (i.e., no repeat candidates or outliers) were used in the equating process. Delta-plot analysis was used to identify anchor items showing substantial deviations (generally, although not exclusively, greater than 3 *SD* units) from expected difficulty values, with an emphasis on establishing an anchor set with difficulty equivalent to that of the full form that adhered to the blueprint. Further, items with very high or low difficulty values and those with low corrected point-biserials were also flagged for potential removal from the anchor set. The goal was a strong midi-test (i.e., moderate range of difficulty, moderate to high discrimination, fit to blueprint) of sufficient length to estimate candidate ability.

The selected set of anchor items had a mean difficulty of 0.71 and a mean corrected point-biserial of .21.

Table 17 shows the fit of the set of anchor items to the blueprint, as percentages. The actual counts are aligned with the targets and reflect the scope and weighting across the full exam.

Table 17: Anchor item fit to blueprint – To May 2018

	Area	Actual	Target
A	Employment Contracts and Terminations	50%	46%
B	Employer Obligations	27%	33%
C	Regulations and Legislation	23%	21%

The mean, Tucker, Levine observed-score, circle arc, and Braun-Holland methods were computed to ascertain concordance of solutions. Given the sample sizes and similarities of test parameters, Tucker was considered the primary method.

Table 18 shows some of the parameters used to derive the equating estimates, along with other parameters describing the test forms. Of note is that on the anchor items, the population taking the September 2019 CHRP ELE scored lower than the population taking the May 2018 CHRP ELE (70.7% vs. 72.6%, respectively; $t(209)=1.05$, *ns*). Because the September 2019 CHRP ELE candidates were of modestly lower ability (based on the anchors, non-significance notwithstanding), they should have a modestly lower pass rate (though, again, the tails of the distribution will be more erratic with small samples).

The equating analysis approximates this, with the pass rate being down modestly (Table 19). All methods show a pass mark of 57 to 60. Given the sample sizes involved, Tucker or Levine observed would be the primary methods under consideration, though again a case for circle arc could be made. The Tucker method showed the lowest estimated equating error and would be chosen based on general rules of thumb for deciding between the Tucker and Levine observed methods.

Table 18: Equating parameter table – To May 2018

	May 2018	Sep. 2019
n	113	98
Scored items	100	100
Mean score	Total	71.3%
	Anchors	72.6%
		70.7%

Table 19: Equating outcome table – To May 2018

Method	Pass Mark		Pass Rate	
	Precise	Integer	All	First Time
Equated May 2018	54.85	55	97.5%	97.3%
Tucker	56.75	57	89.2%	95.9%
Levine observed	59.57	60	86.5%	92.9%
Mean	57.16	58	87.4%	93.9%
Circle Arc 1	57.24	58	87.4%	93.9%
Circle Arc 2	57.15	58	87.4%	93.9%
Braun-Holland	56.47	57	89.2%	95.9%

Because the two planned equating runs showed modest variations across methods, and because of the smaller number of candidates, two additional equating runs were conducted to provide a more precise final pass mark. Note that the observed divergence is likely explained by differences in variance across administrations and the small number of candidates.

Equating Back to the January 2019 Administration

Linear equating (Tucker) was the chosen method for setting the pass mark and it was conducted once key validation was complete. Linear equating is the primary method considered with more than 100 candidates; equipercentile equating would have been considered with more than 1,000 candidates. With candidate samples of fewer than 100, mean or circle arc equating is most prudent.

All candidates in the analysis (i.e., no repeat candidates or outliers) were used in the equating process. Delta-plot analysis was used to identify anchor items showing substantial deviations (generally, although not exclusively, greater than 3 *SD* units) from expected difficulty values, with an emphasis on establishing an anchor set with difficulty equivalent to that of the full form that adhered to the blueprint. Further, items with very high or low difficulty values and those with low corrected point-biserials were also flagged for potential removal from the anchor set. The goal was a strong midi-test (i.e., moderate range of difficulty, moderate to high discrimination, fit to blueprint) of sufficient length to estimate candidate ability.

The selected set of anchor items had a mean difficulty of 0.72 and a mean corrected point-biserial of .19.

Table 20 shows the fit of the set of anchor items to the blueprint, as percentages. The actual counts are closely aligned with the targets and reflect the scope and weighting across the full exam.

Table 20: Anchor item fit to blueprint – To January 2019

	Area	Actual	Target
A	Employment Contracts and Terminations	50%	46%
B	Employer Obligations	30%	33%
C	Regulations and Legislation	20%	21%

The mean, Tucker, Levine observed-score, circle arc, and Braun-Holland methods were computed to ascertain concordance of solutions. Given the sample sizes and similarities of test parameters, Tucker was considered the primary method.

Table 21 shows some of the parameters used to derive the equating estimates, along with other parameters describing the test forms. Of note is that on the anchor items, the population taking the September 2019 CHRP ELE scored lower than the population taking the January 2019 CHRP ELE (72.2% vs. 73.8%, respectively; $t(215)=0.90$, *ns*). Because the September 2019 CHRP ELE candidates were of modestly lower ability (based on the anchors, non-significance notwithstanding), they should have a modestly lower pass rate (though, again, the tails of the distribution will be more erratic with small samples).

The equating analysis shows a lower pass rate as expected (Table 22). The methods show a pass mark range of 57 to 59. Given the sample sizes involved, Tucker or Levine observed would be the primary methods under consideration. The Tucker method showed the lowest estimated equating error and would be chosen based on general rules of thumb for deciding between the Tucker and Levine observed methods.

Table 21: Equating parameter table – To January 2019

		Jan. 2019	Sep. 2019
	N	119	98
	Scored items	101	100
Mean score	Total	71.6%	72.7%
	Anchors	73.8%	72.2%

Table 22: Equating outcome table – To January 2019

Method	Pass Mark		Pass Rate	
	Precise	Integer	All	First Time
Equated Jan. 2019	56.86	57	97.0%	98.3%
Tucker	57.37	58	87.4%	93.9%
Levine observed	58.83	59	87.4%	93.9%
Mean	58.06	59	87.4%	93.9%
Circle Arc 1	58.18	59	87.4%	93.9%
Circle Arc 2	58.10	59	87.4%	93.9%
Braun-Holland	56.39	57	89.2%	95.9%

Equating Back to the September 2018 Administration

Linear equating (Tucker) was the chosen method for setting the pass mark and it was conducted once key validation was complete. Linear equating is the primary method considered with more than 100 candidates; equipercetile equating would have been considered with more than 1,000 candidates. With candidate samples of fewer than 100, mean or circle arc equating is most prudent.

All candidates in the analysis (i.e., no repeat candidates or outliers) were used in the equating process. Delta-plot analysis was used to identify anchor items showing substantial deviations (generally, although not exclusively, greater than 3 *SD* units) from expected difficulty values, with an emphasis on establishing an anchor set with difficulty equivalent to that of the full form that adhered to the blueprint. Further, items with very high or low difficulty values and those with low corrected point-biserials were also flagged for potential removal from the anchor set. The goal was a strong midi-test (i.e., moderate range of difficulty, moderate to high discrimination, fit to blueprint) of sufficient length to estimate candidate ability.

The selected set of anchor items had a mean difficulty of 0.71 and a mean corrected point-biserial of .23.

Table 23 shows the fit of the set of anchor items to the blueprint, as percentages. The actual counts are closely aligned with the targets and reflect the scope and weighting across the full exam.

Table 23: Anchor item fit to blueprint – To September 2018

	Area	Actual	Target
A	Employment Contracts and Terminations	48%	46%
B	Employer Obligations	28%	33%
C	Regulations and Legislation	24%	21%

The mean, Tucker, Levine observed-score, circle arc, and Braun-Holland methods were computed to ascertain concordance of solutions. Given the sample sizes and similarities of test parameters, Tucker was considered the primary method.

Table 24 shows some of the parameters used to derive the equating estimates, along with other parameters describing the test forms. Of note is that on the anchor items, the population taking the September 2019 CHRP ELE scored higher than the population taking the September 2018 CHRP ELE (71.3% vs. 69.4%, respectively; $t(201)=1.01$, *ns*). Because the September 2019 CHRP ELE candidates were of modestly higher ability (based on the anchors, non-significance notwithstanding), they should have a modestly higher pass rate (though, again, the tails of the distribution will be more erratic with small samples).

The equating analysis shows a higher pass rate as expected (Table 25), though the difference was marginal. The methods show a pass mark range of 56 to 57. Given the sample sizes involved, Tucker or Levine observed would be the primary methods under consideration. The Tucker method showed the lowest estimated equating error and would be chosen based on general rules of thumb for deciding between the Tucker and Levine observed methods.

Table 24: Equating parameter table – To September 2018

		Sep. 2018	Sep. 2019
	N	105	98
	Scored items	98	100
Mean score	Total	71.7%	72.7%
	Anchors	69.4%	71.3%

Table 25: Equating outcome table – To September 2018

Method	Pass Mark		Pass Rate	
	Precise	Integer	All	First Time
Equated Sep. 2018	54.72	55	91.4%	95.2%
Tucker	55.55	56	91.0%	95.9%
Levine observed	55.76	56	91.0%	95.9%
Mean	56.15	57	89.2%	95.9%
Circle Arc 1	55.45	56	91.0%	95.9%
Circle Arc 2	55.45	56	91.0%	95.9%
Braun-Holland	56.02	57	89.2%	95.9%

Combined Results

Table 26 shows the pass mark values across the 4 equating runs. The value highlighted in green is the one that would be selected based on sample parameters at each equating run, historic precedent, and similarity other values after averaging. Though there are differences across equating runs, they converge on a value between 56 and 57, except for Levine observed which is higher with a weighted average just over 58 (weighing was by both candidate sample size and number of anchor items). The most prudent pass mark was identified by the weighted mean (56.721) of the four Tucker values.

Note that a legitimate case could be made for using the circle arc or mean equating methods. These yield values very close to that provided by the Tucker method, and their use would have no impact on the final pass mark as an integer value.

Using the established convention for this testing program, the averaged pass mark would be rounded up to a cut score of 57. The resulting pass rate for first-time candidates (95.9%) is slightly lower than that observed for the May 2018, January 2019 and May 2019 administrations, which is in line with the modestly lower scores for September 2019 candidates. It is marginally higher than was seen in September 2018, again as expected given the modestly higher performance for September 2019 candidates in that comparison. The pass rate for all candidates was 89.2% which was lower than what was seen on previous administrations, though this was mostly influenced by an increase in low scoring repeat candidates. See Table 27 for historical pass rates.

The final pass mark value, and the process used to derive it, was presented to the CHRP EVC (Table 28) via teleconference on September 30, 2019. No concerns were raised regarding the pass mark or pass rate. The panel formally approved the pass mark (which was presented along with the consequent pass rate data) for recommendation to HRP. The HRP Registrar accepted the recommended pass mark and so the pass mark was formally established.

Table 26: Equating outcome table – Combined results

	May 18	Sep. 18	Jan. 19	May 19
Tucker	56.8	55.5	57.4	57.1
Levine observed	59.6	55.8	58.8	58.1
Mean	57.2	56.2	58.1	55.0
Circle arc 1	57.2	55.5	58.2	55.3
Circle arc 2	57.2	55.5	58.1	55.3
Braun-Holland	56.5	56.0	56.4	57.0

Table 27: Historical pass rates

	Pass rate	
	All	First-time
Jan. 17	94.6%	95.7%
May 17	94.7%	95.2%
Sep. 17	95.6%	95.9%
Jan. 18	95.8%	97.3%
May 18	97.5%	97.3%
Sep. 18	91.4%	95.2%
Jan. 19	97.0%	98.3%
May 19	95.9%	96.8%
Sep. 19	89.2%	95.9%

Table 28: CHRP Examination Validation Committee – Pass mark approval

Member	Credential	Years of Relevant Experience	Start on EVC	Industry
Sunday Ajao	CHRL	15–20	2017	Banking/Finance
Roxanne Chartrand	CHRL	20–29	2018	Insurance
Claire Chester	CHRL	10–14	2017	Regulation/CPA
✓ Tanya Gopaul	CHRL	10–15	2017	Banking
Jean Lazarus	CHRL	15–19	2017	Health services
✓ Suman Seth	CHRL	15–19	2018	Public sector
Kriss Stone	CHRP	10–15	2017	Real estate
Ilelean Tait	CHRL	15–20	2017	Environmental
✓ Patricia Verkley	CHRL	10–15	2019	Not-for-profit
Karen Weiler	CHRL	20–29	2017	Software/ Communications
✓ Alyssa Young	CHRL	5–9	2017	Non-profit

✓ Participated in the session.

Scoring

To finalize the scoring, candidates who were not included in the item and form analyses were reinserted into the dataset. Scores for each of the 3 domain areas were also computed for each candidate. An Excel file with the final candidate results was provided to HRP.

Table 29 provides the means and standard deviations for the domains and for the total score, using all candidates who took the September 2019 CHRP ELE. Table 30 provides the correlations between each domain. Caution should be exercised in interpreting differences between correlations. Variation can be explained largely by the number of items making up each domain score. That is, domain C has fewer items and shows lower correlations with the other domains. Figure 3 shows the distribution of scores for all candidates, along with the pass mark.

Table 29: Total and domain scores for all candidates

Domain	Percentage	Mean	SD*
A Employment Contracts and Terminations	72%	33.1	5.3
B Employer Obligations	70%	23.8	4.1
C Regulations and Legislation	71%	14.2	2.6
Total score	71.1%	71.1	10.3

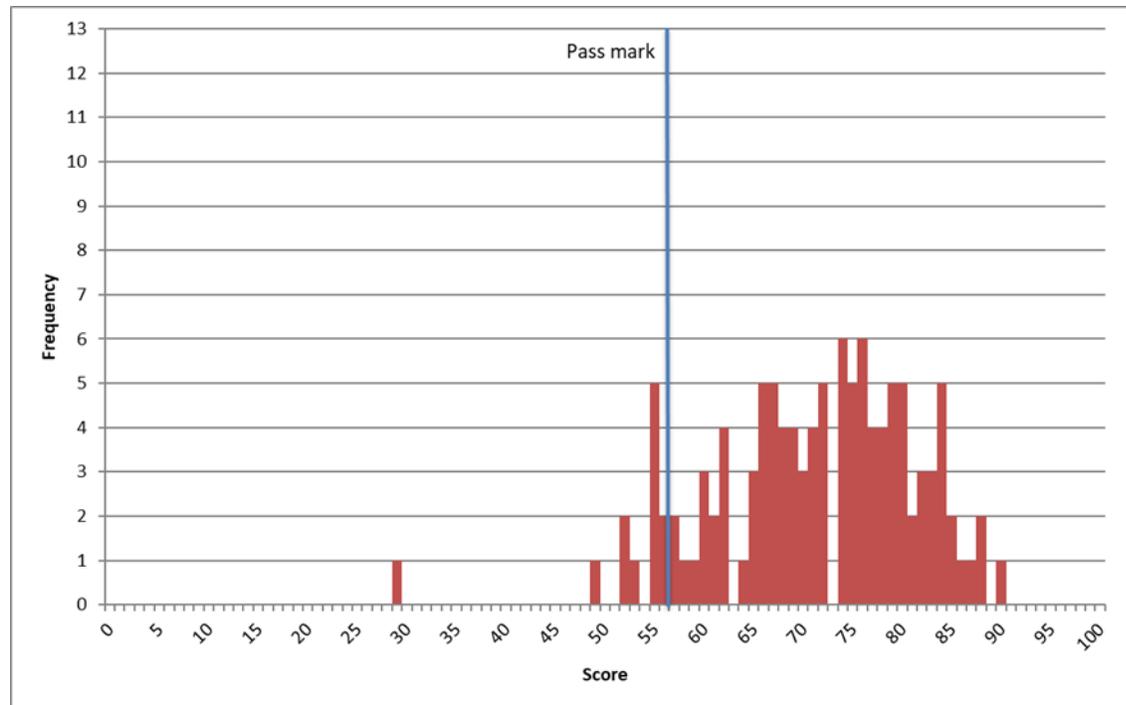
*SD = Standard deviation.

Table 30: Correlations between functional area scores for all candidates

Domain*	A	B	C
A		.62	.55
B			.47
C			

*See Table 29 for the full name of each functional area.

Figure 3: Score distribution for all candidates



Key Examination Metrics

Table 31 shows the key examination metrics for candidates included in the main analysis; that is, only first-time candidates, with outliers removed. Past metrics are provided for reference.

Table 31: Key examination metrics – Candidates included in analysis only

Index	September 2019	May 2019	January 2019	September 2018	May 2018
Scored items	100	102	101	98	100
Candidates	98	158	119	105	113
Mean	72.69 (72.7%)	74.62 (73.2%)	72.29 (71.6%)	70.30 (71.7%)	71.30 (71.3%)
Median	74 (74.0%)	75 (73.5%)	73 (72.3%)	72 (73.5%)	71 (71.0%)
Skewness	-0.259	-0.369	0.005	-0.593	-0.350
Kurtosis	-0.528	-0.122	-0.244	-0.121	0.281
Range	52–90 (52.0– 90.0%)	51–95 (50.0– 93.1%)	51–91 (50.5– 90.1%)	49–84 (50.0– 85.7%)	50–88 (50.0– 88.0%)
Standard deviation	8.66	9.05	8.01	8.04	7.80
Cronbach's alpha	.77	.80	.74	.75	.72
Mean r_{pb}^*	.16	.18	.15	.15	.14
SEM ⁱ	4.13	4.08	4.11	4.06	4.09
SEM at the pass mark	4.68	4.68	4.59	4.55	4.57
Decision consistency (uncorrected) ⁱⁱ	.94	.96	.95	.94	.95
Perceived fairness ⁱⁱⁱ	55%	57%	63%	64%	47%
Pass mark	56.721	55.971	56.857	54.718	54.846
Effective pass mark	57	56	57	55	55
Pass rate	95.9%	96.8%	98.3%	95.2%	97.3%

ⁱSEM = standard error of measurement.

ⁱⁱSubkoviac method.

ⁱⁱⁱBased on responses to the post-examination survey for all candidates.

Related Development Activities

Since the last administration of the CHRP ELE in May 2019, no development activities have taken place.

Appendix

Blueprint

CHRP Employment Law Examination Blueprint

Human Resources Professionals Association

Version 2.1

Approved by CHRP Exam Validation Committee April 10, 2018

Approved by HRP A Registrar April 11, 2018

Effective September 2018 administration

Credential

Passing the CHRP Employment Law Examination is a requirement for certification of CHRP candidates.

Purpose

The CHRP ELE assesses whether a candidate has the ability to make effective decisions when presented with HR situations where comprehension of laws and regulations is centrally relevant, at the CHRP level, in Ontario.

Structure

The structural variables provide high level guidance as to what the examination will look like. These appear in Table 32.

Table 32: CHRP Employment Law Examination Blueprint Structural Variables

Item types	75% Case-based 3-option multiple choice (15-20 single scenarios tied to 4-6 test items each)
	25% Independent 3-option multiple choice
Length	110 total items
	8–12 experimental items
Duration	Up to 3½ hours
Delivery mode	Computer based testing in proctored test centres
Frequency	3 windows per year

Content Weighting

The topic weights were set through a survey of employment lawyers on the most typical situations where employment-related issues are escalated to legal proceedings.

Categories are:

- A. Employment Contracts
- B. Employer Obligations
- C. Regulations and Legislation

Within each Category, the Topics are:

- A. Employment Contracts
 - A1 Termination
 - A2 Contracts
 - A3 Employee Benefits and Perquisites
- B. Employer Obligations
 - B1 Duty to Accommodate
 - B2 Misconduct in the Workplace
 - B3 Common Law
 - B4 Sale of Business
- C. Regulations and Legislation
 - C1 Employment Standards Act
 - C2 Occupational Health and Safety Act
 - C3 Jurisdiction
 - C4 Pay Equity Act
 - C5 Canada Labour Code

The full blueprinted list of Categories, Topics and Subtopics, along with their weighting, appears in Table 33.

Table 33: CHRP Employment Law Examination Blueprint Content Weights

Category Weight	Topic Weight	Topic	Subtopic Weight
46%	A. Employment Contracts and Terminations		
	28%	A1. Termination	
		A1.1 Termination with or without cause	8%
		A1.2 Termination pay, termination notice, and pay in lieu of notice	6%
		A1.3 Continuation of benefits to employee after termination	5%
		A1.4 Severance pay entitlements	5%
		A1.5 What type of income is considered part of terminated employee's salary	2%
		A1.6 Whether or not it is legal to lay off an employee	1%
		A1.7 When and how to lay off an employee	1%
	11%	A2. Contracts	
		A2.1 Contracts and employment agreements	9%
		A2.2 Collective bargaining contracts	2%
	7%	A3. Employee Benefits and Perquisites	
		A3.1 Vacation time, vacation pay and bonuses	5%
	A3.2 Overtime exemptions	2%	
33%	B. Employer Obligations		
	16%	B1. Duty to Accommodate	
		B1.1 Mental health or physical disabilities	9%
		B1.2 Discriminatory grounds (such as family status, age, marital status, etc.)	5%
		B1.3 The duty to accommodate until undue hardship (the threshold)	2%
	9%	B2. Misconduct in the Workplace	
		B2.1 Dealing with harassment and violence in the workplace	5%
		B2.2 HR professional approach to dealing with discipline	2%
		B2.3 Workplace investigations	2%
	6%	B3. Common Law	
		B3.1 Including consideration of Common Law principles	5%
		B3.2 Employers' obligations under Common Law	1%
	2%	B4. Sale of Business	
	B4.1 The effects of the sale of the business	2%	
21%	C. Regulations and Legislation		
	10%	C1. Employment Standards Act	
		C1.1 How to properly interpret the <i>Employment Standards Act, 2000</i>	5%
		C1.2 Probation period under <i>Employment Standards Act, 2000</i>	2%
		C1.3 Different leaves permitted under the <i>Employment Standards Act, 2000</i>	2%
	C1.4 Employers' obligations under <i>Employment Standards Act, 2000</i>	1%	

4%	C2. Occupational Health and Safety Act	
	C2.1 Making policies that are compliant with the <i>Occupational Health and Safety Act, 1990</i>	2%
	C2.2 Ministry of Labour's rights under the <i>Occupational Health and Safety Act, 1990</i>	2%
4%	C3. Jurisdiction	
	C3.1 The difference between federal and provincial legislations	2%
	C3.2 Determining governing legislation when the organization is interprovincial	2%
2%	C4. Pay Equity Act	
	C4.1 Application of <i>Pay Equity Act, 1990</i>	2%
1%	C5. Canada Labour Code	
	C5.1 Employers' obligations under <i>Canada Labour Code, 1985</i>	1%

Note: Reasonable ranges around the Topic weights are employed.

Cognitive Level

The cognitive level weights are based on Bloom's taxonomy. The purpose of this weighting is generally to ensure that an examination does not unintentionally over-focus on specific types of items, and to provide candidates with a range of items (in approximate proportion) that reflects the cognitive operations they must apply on the job. The weights appear in Table 34.

Table 34: CHRP Employment Law Examination Blueprint Cognitive Level Weights

Level	Weight	Range
Knowledge	10%	+/- 3%
Application	60%	+/- 10%
Critical Thinking	30%	+/- 10%

Miscellaneous Guidance

Guidance is not considered binding on the examination, but is used in item development and form development to help create balanced forms.

1. Where scenarios or test items include a workplace, the workplace allocation will be as follows:
 - a. For profit enterprise, 60% (+/- 10%)
 - b. Government, 20% (+/- 5%)
 - c. Not-for-profit, 20% (+/- 5%)
2. 20% (+/- 10%) of workplaces mentioned in scenarios and test items will be unionized.
3. 10% (+/- 5%) of employers mentioned in scenarios and test items will have physical locations in more than one Canadian province.
4. 10% (+/- 5%) of employers mentioned in scenarios and test items will have physical locations both inside and outside of Canada.