

Technical Report: January 2018 CHRP ELE

Human Resources Professionals Association

22 April 2018



Contents

Executive Summary	4
Administration.....	5
Form Setting	5
Testing Window	6
Analysis.....	7
Data Cleaning and Integrity Checks.....	7
Post-Examination Survey.....	9
Initial Analysis	12
Key Validation.....	12
Establishing the Pass Mark: Modified Angoff and Bookmark.....	14
Scoring	21
Key Examination Metrics.....	23
Related Development Activities	24
Item Revision.....	24
Item Review.....	25

List of Tables

Table 1: Domain fit at administration	5
Table 2: Cognitive level fit at administration	6
Table 3: Administration-related post-examination survey questions*	10
Table 4: Content-related post-examination survey questions*	10
Table 5: Preference regarding independent and case-based items.....	11
Table 6: Preference regarding number of response options	11
Table 7: Preference regarding computer-based testing versus pencil-and-paper	11
Table 8: Value of access to legislation	11
Table 9: Initial examination statistics	12
Table 10: Key validation panel members.....	13
Table 11: Domain fit for final scored items	14
Table 12: Cognitive level fit for final scored items.....	14
Table 13: Standard setting panel members.....	15
Table 14: Mean and standard deviations of final Angoff ratings	17
Table 15: Angoff summary statistics.....	18
Table 16: Bookmark results by panel member for the ELE 1	19
Table 17: Standard setting session feedback (for both ELE 1 and ELE 2 ratings)	20
Table 18: Total and domain scores for all candidates.....	21
Table 19: Correlations between functional area scores for all candidates	21
Table 20: Key examination metrics – Candidates included in analysis only.....	23
Table 21: Panel for item revision session	24
Table 22: Panel for item review session	25

List of Figures

Figure 1: Examination time distribution for all candidates	8
Figure 2: Candidate volume and score trends across testing window	8
Figure 3: Score distribution for all candidates.....	22

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 (ELE 1) was administered to 120 candidates using computer-based testing at Prometric test centres from January 8 to 22, 2018, inclusive. The examination comprised 110 three-option multiple choice items and had a 3½-hour time limit. Note that this administration marked an increase of 30 minutes in time allotment to candidates.

The exam was scored using 100 psychometrically-effective items (while adhering to the prescribed distribution across functional areas). The mean score for first-time candidates ($n=111^2$) was 73.0 (73.0%), and for all candidates it was 72.3 (72.3%). Reliability was acceptable at .76 (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 standard setting (combination of modified Angoff and bookmark methods), yielding an integer pass mark of 57. Standard setting was employed because of the increase in testing time available to candidates. The goal of the standard setting exercise was to produce a pass mark that would be equivalent to that used in September 2017, after accounting the change in administration conditions. The pass mark resulted in a pass rate for first-time candidates of 97.3% and a pass rate for all candidates of 95.8%. Note that only candidates who passed CKE 1 were eligible to attempt the ELE 1, and so a very high pass rate is expected.

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

¹ Note: 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 a statistical outlier, 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 forms. Wickett selected the final test forms according to the following parameters:

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

The final form was proofed by Wickett for text errors and detection of potential enemy items. Items flagged as enemies were replaced. This form was also reviewed for currency and enemy items by Alyssa Young and Claire Chester (CHRP EVC members) in a remote session held November 13, 2017.

The final form composition for the January 2018 ELE 1 is shown in Table 1 for domain weighting and Table 2 for cognitive level weighting.

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 candidates scores.

Table 1: Domain fit at administration

Domain	Actual Items	Target Range	Target Items	Variance
A Employment Contracts and Terminations	49	46% ± 5%	46–56	—
B Employer Obligations	36	33% ± 4%	32–40	—
C Regulations and Legislation	25	21% ± 3%	20–26	—
TOTAL	110		110	—

Table 2: Cognitive level fit at administration

Cognitive Level	Actual Items	Target Range	Target Items	Variance
Knowledge	13	10% ± 3%	8–14	—
Application	62	40% ± 10%	33–55	+7
Critical thinking	35	50% ± 10%	44–66	-9
TOTAL	110		110	—

The blueprint was adhered to for content domain but was off blueprint for cognitive level. This reflects the challenges of constructing critical thinking items for this exam content.

Testing Window

The examination was administered via computer-based testing at Prometric test sites primarily in Ontario. The testing window was from January 8 to 22, 2018, inclusive, and 120 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 two documents). Those pieces of legislation are listed below:

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 all of these legislations were as accessed on October 20, 2017.

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, the 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.

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, 56 minutes. The time limit on the ELE 1 was 3½ hours, suggesting that time may still have been a factor for at least some candidates, though with the same number and type of items and an additional 30 minutes, the time pressure must have been less.

Eighteen (15%) of candidates took the full 3½ hours suggesting that those candidates may have wanted more time, and 7 candidates (6%) left at least one item blank suggesting those candidates timed out of the exam before being able to complete it. These metrics will continue to be monitored, and at the present do not appear problematically high, though they are not as low as would be preferred. Note that because of having access to legislation, candidates may take additional time beyond that intended by spending time 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 $-.04$, suggesting that the time constraints were not generally an issue for candidate performance. (Note that two candidates exceeded the time limit; these candidates were granted additional time in advance of the administration as an accommodation.)

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 three days and the last three days was a modest increase of 3.8 marks out of 110 (that increase would be cause for some concern excepting that only 10 candidates wrote in the first 3 days).

As a matter of interest, candidate volumes were also examined across the window and these are also shown in Figure 2. Though not psychometrically meaningful, there is a clear pattern for candidates to prefer to book at the end of window rather than the start.

Figure 1: Examination time distribution for all candidates

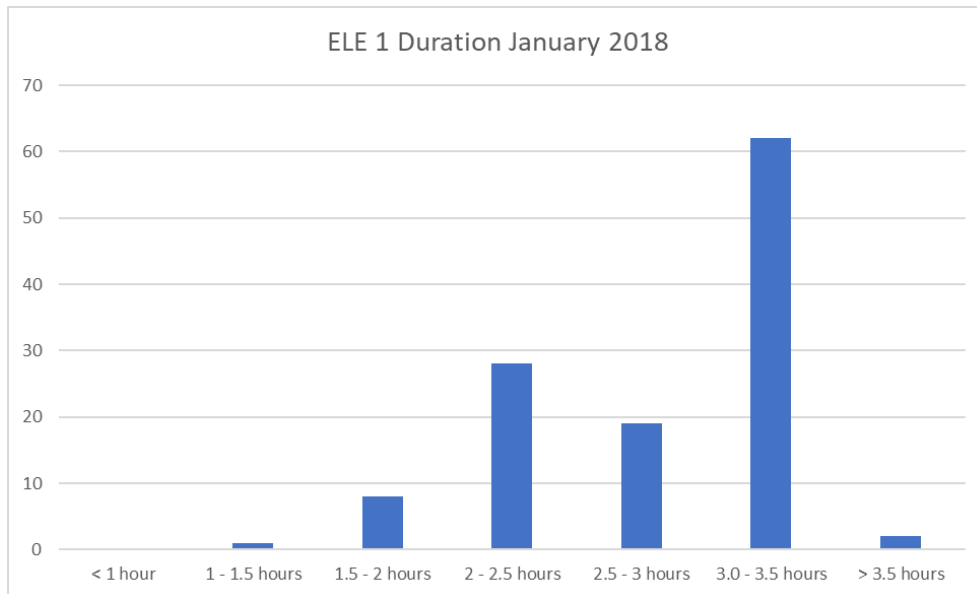
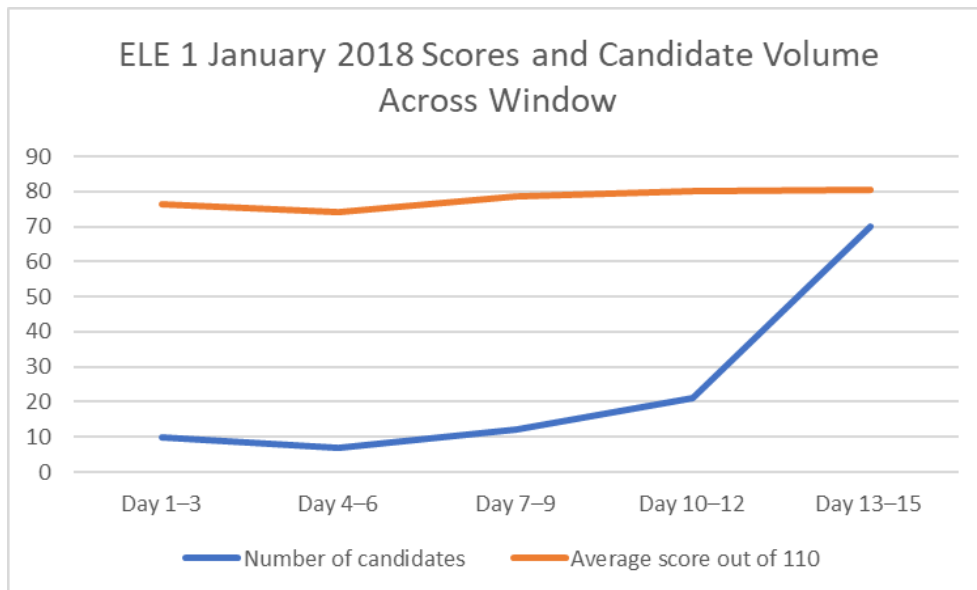


Figure 2: Candidate volume and score trends across testing window



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. No candidates

were flagged for an abnormally low or high score (z value outside ± 3.0). Also, the 110 items were arbitrarily broken into four blocks of 25 items for each candidate plus one final block of 10 items; the 5 resulting scores 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. All outliers were removed from initial analyses and candidates with abnormal response patterns (such as having 5 or more blanks) were also 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 the main group of candidates, and 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 all of these factors, five candidates were removed from scoring.

Candidates who had failed a previous employment law examination (ELE 1 or ELE 2) scored lower than did those who had not (66.0 and 79.9, respectively, on the full exam of 110 items). In keeping with standard procedures, these candidates were removed from subsequent analyses. The ELE 1 analysis proceeded with 111 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 ELE 1; 112 responses were obtained from candidates (response rate, 93%).

Table 3 shows the responses to the administration-related questions for ELE 1 candidates. Note that candidates were generally positive about the administration experience. Table 4 shows the content-related questions for ELE 1 candidates. There was a tendency to more neutrality on these questions. The question on time allotment (Question 8) was up from the 64% endorsement reported in September 2017, presumably owing to the increase in exam duration.

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.	62	38	2	7	3	4.33	89%
2.	I was well informed about what documents to bring to the exam location.	81	25	3	2	1	4.63	95%
3.	Proctors enforced the exam-day rules and the security procedures at the test centre were what I expected.	80	27	1	2	0	4.68	97%
4.	Proctors were professional and courteous.	80	25	4	0	0	4.70	96%
5.	The tutorial helped me understand how to complete the examination on the computer.	67	36	3	0	1	4.57	96%
6.	The legislation and case texts were easy to access during the examination.	48	37	12	7	2	4.15	80%
7.	Navigation through the examination was easy and intuitive.	59	38	5	2	2	4.42	92%

*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.	47	31	5	15	8	3.89	74%
9.	Information available prior to exam day provided me with adequate details about the content and format of the exam.	45	36	12	11	2	4.05	76%
10.	I feel I was adequately prepared to write this examination.	19	55	22	8	2	3.76	70%
11.	The questions in the examination were clearly written.	18	56	16	14	2	3.70	70%
12.	The terminology used in the examination was accurate.	22	68	11	4	1	4.00	85%
13.	The situations presented in the examination were realistic.	30	64	7	3	1	4.13	90%
14.	The questions in the examination reflected the Employment Law Examination blueprint.	26	40	32	6	1	3.80	63%
15.	The examination was a fair assessment of my ability.	11	51	28	13	2	3.53	59%

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

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

Table 5: Preference regarding independent and case-based items

	Count	%
I preferred the independent items.	52	50%
I preferred the case-based items.	34	32%
I had no preference between independent and case-based items.	19	18%

Table 6: Preference regarding number of response options

	Count	%
I preferred having 3 options.	93	88%
I preferred having 4 options.	3	3%
It did not matter to me how many options were used.	10	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.	43	41%
I feel that completing the examination on a computer decreased my performance.	15	14%
I feel that completing the examination on a computer had no effect on my performance.	48	45%

Table 8: Value of access to legislation

	Count	%
Yes, it was essential to me in completing the examination.	49	46%
Yes, but I only consulted it a few times.	40	38%
No, I could not find the answers to questions I had.	9	8%
No, I did not need to consult it to complete the examination.	3	3%
No, it was more of a distraction than an aid.	5	5%

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 was emerged in these comments.

Initial Analysis

The full ELE 1 examination was 110 items, of which 100 were to be scored. The other 10 items were designated as experimental items. However, because only one 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	ELE 1
Items	110
Candidates	120
Candidates in analysis	111
Mean	80.2 (72.9%)
Range	56–96 (50.9–87.3%)
Cronbach's alpha	.74
Disattenuated alpha	.85
Mean r_{pb}^*	.13

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. Items were ranked from worst performing to best performing accordingly.

Key Validation

Key validation was conducted via web meeting on February 6, 2018, using members of the CHRP EVC. The group (Table 10) was first trained in basic item and test analysis methods and

was oriented to the main statistics used to evaluate the quality of the ELE 1. Note that Kriss Stone was absent for approximately 30 minutes of this call.

Table 10: Key validation panel members

Member	Credentials	Years of Relevant Experience	Year on EVC	Industry
Alyssa Young	CHRL	5–9	1	Non-profit
Kriss Stone	CHRL	10–15	1	Real estate
Jean Lazarus	CHRL	15–19	1	Health services
Karen Weiler	CHRL	20–29	1	Software/ Communications

The group was informed that test reliability, as measured by Cronbach’s alpha, was .74 based on the set of 110 potentially scored items and that this was below the generally accepted threshold of .80. They were 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 less direction on those with borderline statistics. Where available, any comments made by candidates to the items were also shown. The group made the decision based on content and the data through discussion; they removed 10 items that they felt were least appropriate to retain for scoring. Comments made by the panel members 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. However, most were moderate to strong items. The final alpha for the set of 100 scored items was .76. The difficulties ranged from 30.6% to 99.1%, with a mean of 73.0%. The r_{pb}^* values ranged from $-.02$ to $.48$, with a mean of $.15$.

Table 11 shows the scored ELE 1’s final fit to the domain weighting. Table 12 shows the same for cognitive level. The exam fit based on content domain, but fell short on having as much critical thinking weighting as intended. This was not considered meaningful with respect to candidate score interpretations but would be monitored in the future.

The group endorsed the final set of items for use in scoring the January 2018 ELE 1 candidates.

Table 11: Domain fit for final scored items

Domain	Actual Items	Target Range	Target Items	Variance
A Employment Contracts and Terminations	43	46% ± 5%	41–51	—
B Employer Obligations	33	33% ± 4%	29–37	—
C Regulations and Legislation	24	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	11	10% ± 3%	7–13	—
Application	57	40% ± 10%	30–50	+7
Critical thinking	32	50% ± 10%	40–60	-8
TOTAL	100		100	—

Establishing the Pass Mark: Modified Angoff and Bookmark

Overview

Because 30 minutes of additional time was given to candidates as of this administration, standard setting was required to establish the pass mark. The modified Angoff method was used (see, e.g., Cizek & Bunch, 2007⁵; Plake & Cizek, 2012⁶; and Smith & Springer, 2009⁷) following a standard two-round protocol with individual and group feedback and information on item statistics in cases where substantial discrepancies were revealed. The bookmark method

⁵ Cizek, G.J., & Bunch, M.B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage Publications.

⁶ Plake, B.S., & Cizek, G.J. (2012). Variations on a theme: The modified Angoff, extended Angoff, and yes/no standard setting methods. In G.J. Cizek (Ed.), *Setting performance standards* (pp. 181–199). New York, NY: Routledge.

⁷ Smith, I.L., & Springer, C.C. (2009). Standard setting. In Institute for Credentialing Excellence, *Certification: The ICE handbook* (pp. 235–264). Washington, DC: Institute for Credentialing Excellence.

(see, e.g., Cizek & Bunch, 2007⁸; Lewis, Mitzel, Mercado & Schulz, 2012⁹) was also conducted with a standard two-round protocol.

The panel members are shown in Table 13. The session took place February 15–16, 2018 at the HRP office. The ELE 2 was also assessed during this session, and the panel members are a cross section from the CHRP EVC and CHRL EVC.

Table 13: Standard setting panel members

Member	Credentials	Years of Relevant Experience	Year on EVC	Industry
Sunday Ajao	CHRL	15–20	1	Banking/Finance
Jennifer Borges	CHRL	10–14	1	Manufacturing
Annette Dhanasar	CHRL	15–19	1	Transportation
Lorri Ann Kennis	CHRL	20–29	*	Provincially regulated
Jennifer King	CHRL	20–29	1	Business and professional services
Jean Lazarus	CHRL	15–19	1	Health services
Lisa Scian	CHRL	15–19	1	Information & communication technology
Kriss Stone	CHRP	10–15	1	Real estate
Karen Weiler	CHRL	20–29	1	Software/Communications
Alyssa Young	CHRL	5–9	1	Non-profit

*Not on an examination validation committee. Brought in to add more perspectives.

The panel members were provided with the following information in advance of the session:

- Overview of session process
- Descriptions of the CHRP and CHRL credential holders
- Links to current requirements for CHRP and CHRL designations
- Blueprint for the ELE
- Links to current legislation

⁸ Cizek, G.J., & Bunch, M.B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage Publications.

⁹ Lewis, D.M., Mitzel, H.C., Mercado, R.L., & Schulz, E.M. (2012). The bookmark standard setting procedure. In G.J. Cizek (Ed.), *Setting performance standards* (pp. 225-253). New York, NY: Routledge.

- Details on recent changes stemming from Bill 148
- Basic psychometric terms
- General information on the Angoff method

Just Qualified Candidate

At the session, after a general orientation of the group to the credentials, the panel spent time refining the definition of the just qualified candidate for both CHRP and CHRL designations. These descriptions built upon past work, and the two credentials were considered in contrast to each other to help in better delineating differences in expectations. The HRPAs registrar participated in the initial discussions on the just qualified candidate from the perspective of the intent of HRPAs as a regulator. The group expressed general consensus on the level of performance of the borderline candidates, even if there was some disagreement on specific details.

Modified Angoff

Once participants expressed a common understanding of the level at which the JQC functions, they were trained on the Angoff method. Before starting with the ratings, the panel members were asked to complete the first 10 items as a test candidate would to provide insight into the actual difficulty of these items.

For each item, the raters were to determine the likelihood of a correct response from a JQC (on a scale from 5% likelihood to 95% likelihood). The emphasis was on evaluating the likelihood that a JQC *would* get the item correct, rather than *should*. Further, the raters were encouraged to consider the entire scale; it was ensured that they understood that higher ratings meant relatively easier items.

For materials, panel members had a single book of the 54 items in common between the ELE 1 and ELE 2 as administered in January 2018. The intent was to Angoff only those items, and then extrapolate to the full test in proportion to the difficulty of the selected items. The judges also had a booklet of relevant case texts and rating sheets containing space for ratings for ELE 1 and for ELE 2.

Note that for each item, the panel members provided ratings for both ELE 1 and ELE 2 at the same time. In contrast to past standard setting studies, the decision was taken on this occasion to more formally recognize that the two credentials (CHRP and CHRL) exist in relation to each other and so ratings of Angoff expectations should be considered at the same time.

During calibration (using live items), after reading each item, the SMEs were asked to consider their rating and then, once everyone had written down their rating independently, provide it orally. For each item, the raters were to determine the likelihood of a correct response from a JQC (on a scale from 5% likelihood to 95% likelihood). The emphasis was on evaluating the likelihood that a JQC *would* get the item correct, rather than *should*. Further, the raters were encouraged to consider the entire scale; it was ensured that they understood that higher ratings meant relatively easier items.

Once all had provided their value, they were told how they did as a group and individually relative to the actual performance for candidates. For each item, the ratings were compiled using Microsoft Excel, and items for which the range in ratings was 30 points or more were identified for group discussion. For these discrepant items, the highest and lowest raters were asked to justify their ratings, and then all raters had the opportunity to contribute to the discussion. Similarly, if the group rated an easy item as being difficult (or vice versa), they were given this information and asked to consider why candidates had a different experience on the item. Once the discussion was complete, the raters had the option to change or not change their Angoff ratings for that item. This discussion constituted Round 2, at the end of which the ratings were finalized. Note that discussion was curtailed in Round 2 as the overall session time constrained fuller elaboration. These discussions were held concurrently for the ELE 1 and ELE 2 ratings (but, if an item was discrepant for one exam but not the other, it was only discussed with reference to the exam where the discrepancy was observed).

Note that even though the pass rate consequences were known at this time, the group was not provided with that information. It was, however, used throughout the process by the facilitator to monitor the general level of Angoff ratings.

The process continued until the 54 common items had been Angoffed for both ELE 1 and ELE 2. The final mean Angoff value was 55.72% for the ELE 1 items. The mean difficulty of these 54 items was 0.727 ($SD = 0.176$), almost the same as seen on the full test which had a mean difficulty of 0.730 ($SD = 0.169$). A simple extrapolation using the ratio of full form difficulty to Angoffed subform difficulty was used to adjust the Angoff up to 55.96%. The individual means and standard deviations for each rater are presented in Table 14. Table 15 shows the quality metrics for the Angoff procedure.

Table 14: Mean and standard deviations of final Angoff ratings

Alyssa	Karen	Lorri	Jennifer B	Lisa	Sunday	Annette	Jennifer K	Kriss	Jean	Group Mean
59.4 (12.5)	61.3 (12.8)	56.1 (17.1)	52.2 (9)	57.6 (12.7)	54.9 (8.8)	52.5 (5.5)	63.1 (13.9)	48.9 (9.8)	51.1 (14.9)	55.7 (8.2)

Table 15: Angoff summary statistics

	ELE 1
Items Angoffed	54
Mean Angoff	55.7%
Median	54.3%
Range	39.0–73.1%
SEj	1.397
SEi	1.109
Reliability*	.86
<i>r</i> with item difficulties	.68
SEM	0.081
Pass mark extrapolated to full form of 100 items	55.96

*Hoyt

Bookmark

The bookmark proceeded upon completion of the Angoff. The panel knew from feedback on the Angoff ratings that the pass mark may be a little lower than expected, though that did not form part of the training for the bookmark and they were not instructed to use that information in the bookmark exercise.

The training for the bookmark included the primary instruction to evaluate each item as to whether the just qualified CHRP candidate will answer the question correctly two-thirds of the time. The panel members were instructed to continue until the answer to this question was 'no', at which point they were to indicate the previous item as their 'bookmark'.

The panel members were provided with an ordered item booklet constructed based on a 1-parameter item response theory analysis of the candidate data. Items from both exams were placed on the same scale so that a single ordered item booklet could be produced. Some of the easiest items were removed to reduce reading time, and similarly the hardest items were also removed as the cut score would not conceivably be that high (the panel was informed of this and agreed with the decision).

The panel was broken into pairs so that discussion could be held while they worked through individual ratings for Round 1. After completion of Round 1, feedback was provided on pass rate consequence information and this was to be used by the panel members as they returned to the ordered item booklets for their Round 2 re-consideration of the placement of the bookmark. Both

the CHRP and CHRL were completed at the same time: the panel was instructed to first seek out the bookmark placement for CHRP, and then continue from that point for the borderline just-qualified candidate.

Table 16 shows the bookmark ratings for each panel member, along with the final mean value.

Table 16: Bookmark results by panel member for the ELE 1

Panel member	Page	Theta	Raw score equivalent
Alyssa	51	-1.553	60
Karen	43	-1.860	55
Lorri	56	-1.485	60
Jennifer B	49	-1.590	59
Lisa	56	-1.485	60
Sunday	55	-1.493	60
Annette	44	-1.690	58
Jennifer K	52	-1.538	60
Kriss	39	-1.949	54
Jean	43	-1.860	55
Mean	48.8	-1.650	58

Final Pass Mark Determination

To arrive at a final pass mark, the panel was provided with several pieces of information in addition to their Angoff and bookmark mean ratings.

First, linear equating (Tucker) was conducted to provide information on the lowest plausible pass mark. That is, because the provision of extra time could only be neutral or beneficial to scores, the equated value back to September 2017 provides the lowest pass mark logically possible. This value was 52.80.

Second, the panel was told that historically, the pass rates on the ELE 1 have ranged from 95.2% to 95.9% (for first-time candidates in the 2017 administrations of the ELE 1) and that there would not be sound reason to expect a substantial deviation from values of that magnitude. That is, the expectations have not changed for what candidates need to demonstrate, and on the assumption of a comparable candidate population in January 2018, the pass rate should be similar.

Third, and related to the second piece of information, the panel was told that pass mark of 58 on the January 2018 ELE 1 would yield a pass rate most similar to that seen in the past (the same value was closest for each of the 2017 administrations).

The panel was told that the mean Angoff value of 55.95 (rounded up to 56) would lead to a pass rate of 98.2%, and that the mean bookmark value of 57.67 (rounded up to 58) would lead to a pass rate of 96.4%. The panel was advised that in the absence of a clear reason to do otherwise, the simple average of these two values (56.82, rounded up to 57) would be recommended as the final cut score.

The panel discussed their relative confidence in the Angoff and bookmark procedures, and reported that while they felt more confidence in the Angoff values because of the volume of ratings they made, they felt better informed by the time they made their bookmark ratings. Accordingly, they were comfortable with the average of the two values. The resulting pass rate (97.3%) was marginally higher than that seen in the past, but not out of line with expectations. The panel was concerned about how high this value was, but with discussion about why it should be high accepted the value for recommendation to HRP.

Once this process was completed, the SMEs were provided with a post-standard setting survey to quantify their experience. Based on these ratings (see Table 17), the panel had confidence in the final recommendations.

Table 17: Standard setting session feedback (for both ELE 1 and ELE 2 ratings)

		Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	I was able to conceptualize the just qualified candidate.				1	9
2	I had enough training to complete the Angoffing exercise.				2	8
3	I had sufficient time to complete my Angoff ratings.				1	9
4	I am confident in the Angoff ratings I provided.				2	8
5	I had enough training to complete the Bookmark exercise.				2	8
6	I had sufficient time to complete my Bookmark ratings.				2	8
7	I am confident in the Bookmark ratings I provided.				4	6
8	I am confident in the final pass marks provided by the group.				1	9

Pass Mark Approval

Analyses were provided to the HRP A registrar for consideration on February 18, 2018. The registrar formally approved the pass mark for the January 2018 ELE 1 via email on February 19.

Scoring

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

Table 18 provides the means and standard deviations for the domains and for the total score, using all candidates who took the January 2018 ELE 1. Table 19 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 18: Total and domain scores for all candidates

Domain	Percentage	Mean	SD*
A Employment Contracts and Terminations	72%	31.0	4.5
B Employer Obligations	73%	24.0	3.3
C Regulations and Legislation	72%	17.3	2.7
Total score	72.3%	72.3	8.8

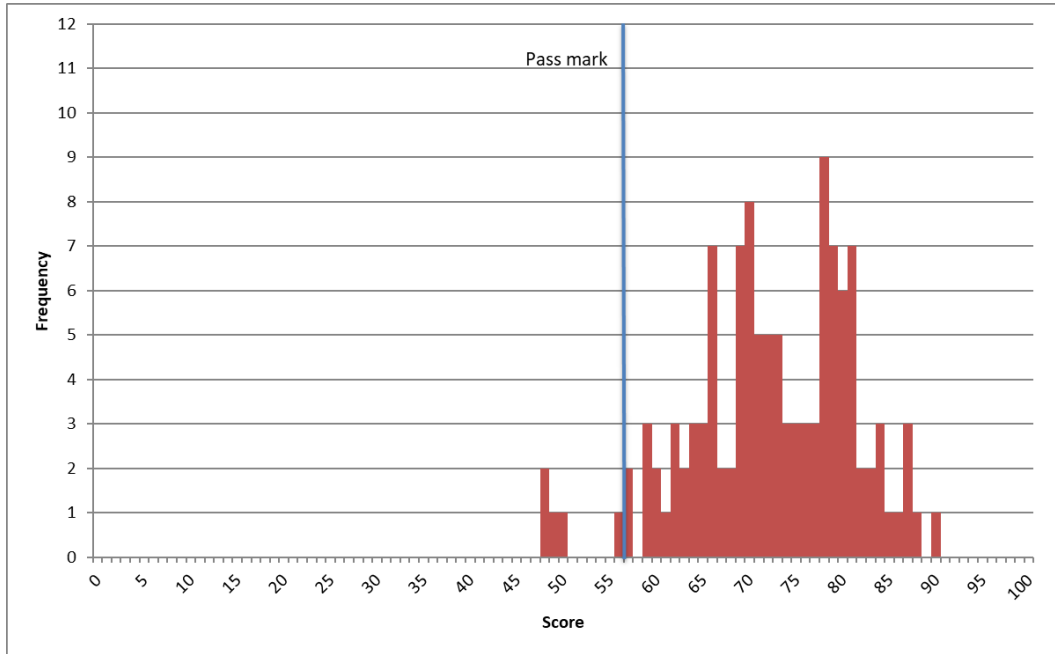
*SD = Standard deviation.

Table 19: Correlations between functional area scores for all candidates

Domain*	A	B	C
A		.58	.47
B			.43
C			

*See Table 18 for the full name of each functional area.

Figure 3: Score distribution for all candidates



Key Examination Metrics

Table 20 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.

Note that January 2017 was the first computer-based testing delivery of the ELE 1. The September 2016 pilot was administered on paper and candidates did not have access to legislation. In 2018 the time limit for candidates was increased from 3 hours to 3½ hours.

Table 20: Key examination metrics – Candidates included in analysis only

Index	January 2018	September 2017	May 2017	January 2017	September 2016
Scored items	100	100	98	97	95
Candidates	111	123	124	146	243
Mean	73.03 (73.0%)	71.54 (71.5%)	69.68 (71.1%)	70.98 (73.2%)	66.84 (70.4%)
Median	73 (73.0%)	72 (72.0%)	71 (72.4%)	71.5 (73.7%)	67 (70.5%)
Skewness	-0.432	-0.530	-0.664	-0.280	-0.276
Kurtosis	-0.023	0.073	-0.201	-0.155	-0.501
Range	49–90 (49.0– 90.0%)	44–90 (44.0– 90.0%)	44–86 (44.9– 87.8%)	47–90 (48.5– 92.8%)	44–82 (46.3– 86.3%)
Standard deviation	8.22	9.33	8.61	8.45	7.57
Cronbach's alpha	.76	.81	.79	.80	.73
Mean r_{pb}^*	.15	.18	.17	.18	.14
SEM ⁱ	4.04	4.08	3.99	3.80	3.94
SEM at the pass mark	4.59	4.61	4.47	4.30	4.33
Decision consistency (uncorrected) ⁱⁱ	.95	.96	.94	.93	.91
Perceived fairness ⁱⁱⁱ	59%	54%	51%	61%	n/a
Pass mark	56.817	53.363	53.248	57.462	54.376
Effective pass mark	57	54	54	58	55
Pass rate	97.3%	95.9%	95.2%	95.2%	93.4%

ⁱSEM = standard error of measurement.

ⁱⁱSubkoviak method.

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

Related Development Activities

Since the last administration of the ELE 1 in September 2017, the following exam development activities have taken place.

Item Revision

In light of the impending Bill 148 changes to the Employment Standards Act, 2000 and *Labour Relations Act*, a session was held to review and revise test items to ensure they were consistent with the planned changes. Note that candidates for the January 2018 administration of the ELE 1 were informed that the new legislation would not be reflected on that form of the exam.

The session was held November 20–22, 2018 at HRPAs offices. The panel members who participated are shown in Table 21.

Table 21: Panel for item revision session

Panelist	Credentials	Years of Relevant Experience	Industry
Graham Craig	CHRL	20–29	Consultant
Charles Creppy	CHRL	20–29	Community Services
Lorri Ann Kennis	CHRL	20–29	Provincially regulated
Bonnie O’Keefe	CHRL	20–29	Consultant
Karen Pantaleo	CHRL	20–29	Consultant

The panel members worked primarily individually, after receiving training on the process and discussion of the Bill 148 changes. The five panel members worked on separate items, with feedback from the facilitator and group discussion as necessary. Note that this session was held just before the bill was formally enacted into law.

This occasion was also used to edit items that had not performed well in the past, and panel members were provided with item and distractor statistics to assist in this process. Any comments that had been written in by candidates in previous administrations were also provided.

The panel members reviewed 237 items and cases texts and made changes to 159 of them.

Following the session, all the changed items were sent for editorial review before being integrated into the bank. These items were considered the same as newly written items, still requiring group review and EVC validation before use on future examinations.

Item Review

Following the item revision session in November 2017 there was need for group review of those items before moving them to formal validation and use on the ELE 1.

The review session was held January 15–17, 2018 at HRPAs offices. The panel members who participated are shown in Table 22. This session involved the review of CKE 1 items as well.

Table 22: Panel for item review session

Panelist	Credentials	Years of Relevant Experience	Industry
Laurence Frederick	CHRL	10–15	Consultant/Education
Annette Lawrence	CHRP	1–5	Legal services
Surbhi Sud	CHRL	10–14	Health and social services
Brittany Seguin	CHRP	1–5	Utilities
Sanjeev Pruthi	CHRP	10–14	Mid level management

The panel members received training on the review activity, and then worked primarily individually reviewing items to make sure they reflected current legislation (including changes stemming from Bill 148) and current practice. Where panel members proposed changes, these were discussed by the group before implementation.

The panel members reviewed and approved 92 items as suitable for both ELE 1 and ELE 2, 10 items as suitable for ELE 2 only, and rejected 4 items. Of the approved items, 14 saw text changes before approval.

The items were updated in the bank, and those that were approved were deemed ready for validation before use on future examinations.