

Technical Report: November 2022 CHRL-KE

HR | Human Resources
PA | Professionals Association

December 14, 2022



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	15
Key Validation.....	16
Establishing the Pass Mark: Equating.....	18
Scoring	27
Key Examination Metrics.....	30
Related Development Activities	31
Validation.....	31
Item Writing	32
Validation.....	33
Appendix A.....	36
Appendix B.....	38
Appendix C.....	39

List of Tables

Table 1: Test forms as administered	7
Table 2: Content-related post-examination survey questions*	11
Table 3: Administration-related post-examination survey questions*	12
Table 4: Testing location preference	12
Table 5: Actual testing location	12
Table 6: Testing location preference by actual testing location.....	13
Table 7: Reason for choosing own location (live remove proctoring candidates).....	13
Table 8: Evaluation of testing experience (live remove proctoring candidates).....	13
Table 9: Value in future candidates being able to test from their own location (live remote proctoring candidates).....	13
Table 10: Able to write at a convenient location (test centre candidates).....	14
Table 11: Evaluation of testing experience (test centre candidates)	14
Table 12: Value in future candidates being able to test from their own location (test centre candidates)	14
Table 13: Initial examination statistics	15
Table 14: Section item statistics	15
Table 15: CHRL Examination Validation Committee members – Key validation.....	17
Table 16: Final scored examination fit to blueprint	18
Table 17: Anchor item fit to blueprint – To May 2022	20
Table 18: Equating parameter table – Total pass mark, to May 2022.....	20
Table 19: Equating outcome table – Total pass mark, to May 2022	21
Table 20: Anchor item fit to blueprint – To November 2021	22
Table 21: Equating parameter table – Total pass mark, to November 2021	22
Table 22: Equating outcome table – Total pass mark, to November 2021	23
Table 23: Equating outcome table – Combined results, total pass mark.....	23
Table 24: Historical pass rates – Total pass mark	24
Table 25: Alignment between difficulty of anchors and full exam.....	25
Table 26: Equating outcome table – Combined results, functional area thresholds	25
Table 27: Equating summary table – Functional area thresholds	26
Table 28: Passing decisions – Total pass mark and functional areas.....	26

Table 29: CHRL Examination Validation Committee members – Pass mark approval.....	27
Table 30: Total and functional area scores for all candidates	28
Table 31: Correlations between functional area scores for all candidates	28
Table 32: Key examination metrics – Candidates included in analysis only.....	30
Table 33: CHRL Examination Validation Committee members – Validation	31
Table 34: Item writers.....	33
Table 35: CHRL Examination Validation Committee members – Validation	34
Table 36: CHRL-KE Blueprint structural variables.....	36
Table 37: Functional area weights on the CHRL-KE	37
Table 38: Competencies not eligible on the CHRL-KE	37

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.....	29

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 CHRL Knowledge Exam (CHRL-KE²) was administered to 258 candidates using computer-based testing via live remote proctoring and at Prometric test centres November 15–29, 2022, inclusive.³ The examination comprised 250 four-option multiple choice items and had a 5-hour time limit.

As per the CHRL-KE blueprint, the exam was scored using the 220–230 best-performing items (while adhering to the prescribed distribution across functional areas). The mean score for first-time candidates ($n=222^4$) was 155.6 (67.9%), and for all candidates it was 152.1 (66.4%), out of 229 scored items. Reliability was strong at .92. The final set of scored items adhered to the blueprint parameters.

The pass mark was set using equating back to the May 2022 and November 2021 administrations, yielding an integer pass mark of 144. 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 CHRL-KE. This pass mark resulted in a pass rate for first-time candidates of 69.8% and a pass rate for all candidates of 63.6% based on the total score pass mark requirement. Additionally, two candidates failed the exam based on not achieving the threshold required in at least one of the functional areas.

This report, the analyses performed, and the processes followed are consistent with NCCA standards⁵ and ISO 17024 standards.⁶

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 The CHRL-KE was titled the CKE 2 up until the Fall of 2020. Any reference in this report to past administrations of the CHRL-KE will use the new title.

³ Several candidates wrote just after the close of this window due to technical difficulties.

⁴ Excludes those who had failed an HRP A examination in the past, who were identified as being statistical outliers, or who had written an alternative test form.

⁵ National Commission for Certifying Agencies (2014). *Standards for the accreditation of certification programs*. Washington, DC: Institute for Credentialing Excellence.

⁶ 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 three 250-item test forms (using a combination of scored and experimental test items). Wickett constructed the final test forms according to the following parameters:

1. Including only items validated by the validation panel in the past 3 years
2. Fitting the total item count of 250
3. Excluding enemy items
4. Matching the blueprint target value (+/- 2%) for each functional area
5. Maximizing spread across competencies
6. Reducing item exposure
7. Selecting items with perceived psychometric effectiveness, using statistics from previous administrations as available

Wickett proofed the final forms for text errors and detection of potential enemy items. Items flagged as enemies were replaced.

After selecting the 250 items for each form, Wickett split the forms in half to allow for the administration of the exam in two sections. Section 1 was allocated 125 items and Section 2 was allocated 125 items. With each form, the two sections were set to balance for:

- Number of words
- Item difficulty
- Item discrimination (adjusted point-biserial)
- Number of experimental items
- Adherence to blueprint
- Number of anchor items

The final form composition for the primary November 2022 CHRL-KE forms is shown in Table 1. All functional areas are within 2 items of their targets, and therefore, the forms reflect the blueprint (see Appendix A for the CHRL-KE blueprint). Differences between targets and actuals reflects differential allocation of experimental items rather than a deviation from scored item targets.

Note that at any administration, HRPAs also makes 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: Test forms as administered

	Functional Area	Actual Items	Target
10	Strategy	27–28	27–28
20	Professional Practice	28–29	27–28
30	Organizational Effectiveness	36–37	35
40	Workforce Planning & Talent Management	32–33	35
50	Labour & Employee Relations	22–23	22–23
60	Total Rewards	23–24	25
70	Learning & Development	27–28	27–28
80	Health, Wellness & Safe Workplace	21	20
90	HR Metrics, Reporting & Financial Management	30–31	30
	TOTAL	250	250

Testing Window

The examination was administered via computer-based testing using live remote proctoring and at Prometric test sites primarily in Ontario. The testing window was November 15–29, 2022, inclusive, and 258 candidates wrote the exam⁷.

Candidates were able to select either a test centre (assuming one was available reasonably close to them and was open) or live remote proctoring from a location of their choosing. Standard security methods (as per Prometric protocols⁸) were employed for both methods. Candidates were allowed one 15-minute break after submitting section 1 and before beginning section 2. This break did not count against total time for the candidate.

Candidates had access to a basic-function calculator on screen. No other aids or resources were allowed.

⁷ Due to technical difficulties requiring the rescheduling of some candidates, testing continued through to November 30, 2022.

⁸ Information on procedures and security can be found at www.prometric.com/ProProctor and www.prometric.com/proproctorcandidate.

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 rosters provided by HRP and 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 250 items to verify key accuracy. Comments on items were also reviewed to identify any specific item-level issues. No problems were encountered.

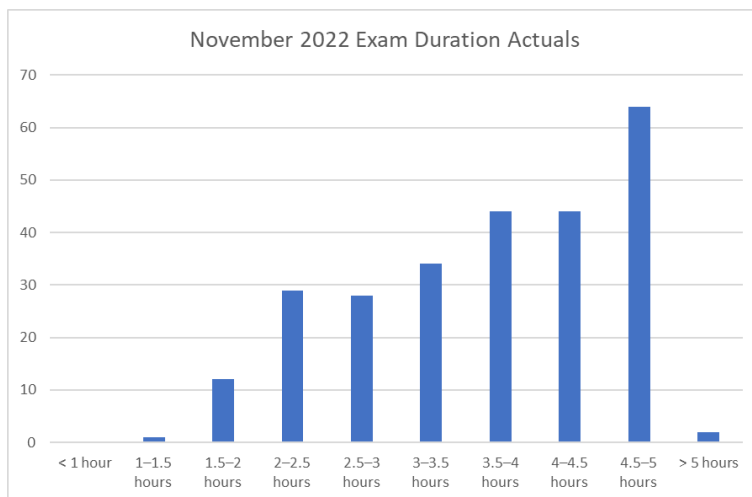
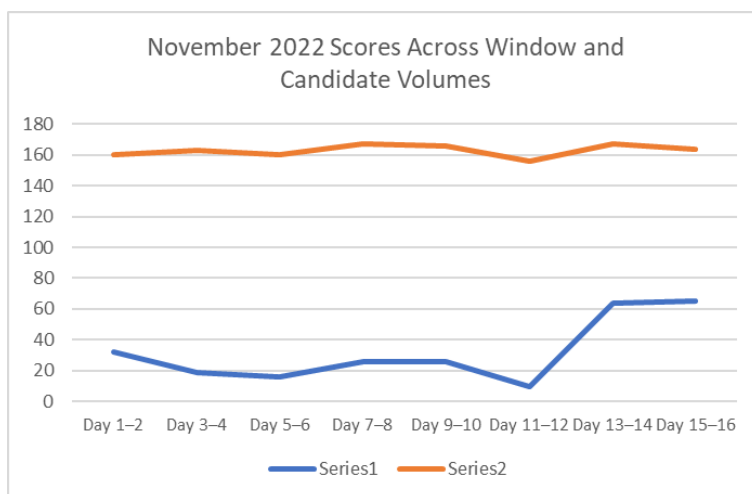
The average time taken by all candidates was assessed to detect potential examination timing concerns. The distribution is shown in Figure 1. The mean was 3 hours, 40 minutes (4 minutes less than in May 2022; on average, form A candidates took 3 hours, 46 minutes, form B candidates took 3 hours, 41 minutes, and form C candidates took 3 hours, 34 minutes). The time limit on the CHRL-KE was 5 hours, suggesting that time was not a factor in scores across candidates.

Sixteen candidates (6%) took the full 5 hours, suggesting that those candidates may have wanted more time, and 4 candidates (2%) left at least 1 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 at present do not appear problematically high.

The correlations between scores and time spent writing the examination were negligible at values of .12, .12, and .04 across the three forms, respectively, suggesting little overall relation between time spent on items and performance.

Candidate scores were computed across the window to look for any evidence of item exposure. As shown in Figure 2, the overall effect was for a decrease in scores across the window, though the magnitude is small. The difference between mean scores for candidates writing in the first 2 days and those writing in the last 2 days was an increase of 3.7 marks out of 250 (1.5%).

As a matter of interest, candidate volumes were also examined across the window; these are also shown in Figure 2. As is typical, candidates tended to book appointments towards the end of the testing window.

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 250 items. Three candidates were flagged for an abnormally low or high score (z value outside ± 3.0). Also, the 250 items were arbitrarily broken into 10 blocks of 25 items for each candidate; the 10 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. All outliers were removed from initial analyses; candidates with abnormal response patterns were also removed. Candidates who left more than 5 blanks were also removed from analysis. As a result of all of these factors, 4 candidates were removed from analysis.

Candidates who had failed a previous HRP examination (CKE, CHRP-KE, or CHRL-KE) scored lower than did those who had not (59.0% and 66.5%, respectively, on the full exam of

250 items). This difference was meaningful and significant ($t(104)=4.81$, $p<.001$). In keeping with standard procedures, these candidates were removed from subsequent analyses. The CHRL-KE analysis proceeded with 222 candidates.

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

Post-Examination Survey

Candidates were provided with access to the post-examination survey immediately after submitting their responses to the CHRL-KE; 253 responses were obtained from candidates (response rate, 98%). The survey was revised for the September 2020 administration to collect information specifically applicable to the use of live remote proctoring, and further revised for the May 2022 administration to refine that information.

Table 2 shows the responses to the content-related questions; there was a tendency to neutrality on these questions. The rating for perceived fairness (Question 8) warrants monitoring as it continues to be low. Table 3 shows the administration-related questions; candidates were generally very positive about the administration experience.

Table 2: Content-related post-examination survey questions*

	Question	SA	A	N	D	SD	Score	Agreement	Agreement last 5 [^]
1.	The time allotted for this examination was sufficient.	147	90	11	3	2	4.5	94%	90%
2.	Information available prior to exam day provided me with adequate details about the content and format of the exam.	64	100	43	34	11	3.9	65%	70%
3.	I feel I was adequately prepared to write this examination.	12	78	95	54	12	3.5	36%	41%
4.	The questions in the examination were clearly written.	18	117	64	44	9	3.6	54%	57%
5.	The terminology used in the examination was accurate.	19	143	63	24	3	3.8	64%	65%
6.	The situations presented in the examination were realistic.	35	155	46	15	2	4.0	75%	73%
7.	The questions in the examination reflected the examination blueprint.	18	113	84	30	5	3.8	52%	49%
8.	The examination was a fair assessment of my ability.	12	78	94	53	15	3.4	36%	38%

*Response categories: SA = strongly agree; A = agree; N = neutral; D = disagree; SD = strongly disagree.

[^]Mean value of candidate agreement across the previous 5 administrations.

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

	Question	SA	A	N	D	SD	Score	Agreement	Agreement last 5 [^]
9.	I was able to book to write the examination at a time that was convenient for me.	101	105	18	23	4	4.2	82%	81%
10.	I was well informed about the examination rules and regulations.	122	111	11	7	1	4.4	92%	95%
11.	Proctors enforced the exam-day rules.	150	94	6	1	0	4.6	97%	97%
12.	Proctors were professional and courteous.	148	85	10	9	0	4.5	92%	94%
13.	The tutorial helped me understand how to complete the examination on the computer.	128	100	15	0	1	4.5	93%	89%
14.	Navigation through the examination was easy and intuitive.	128	106	13	5	0	4.5	93%	94%

*Response categories: SA = strongly agree; A = agree; N = neutral; D = disagree; SD = strongly disagree.

[^]Mean value of candidate agreement across the previous 5 administrations.

Candidates were asked where they had preferred to write (Table 4) and where they actually wrote the examination (Table 5), and based on their response the questions that followed differed. Table 6 shows that candidates were generally able to write using the modality of their preference.

Table 4: Testing location preference

Response	Count	%
I preferred using my own location.	150	59%
I preferred going to a test centre.	86	34%
I had no preference.	19	7%

Table 5: Actual testing location

Response	Count	%
Test centre	93	36%
Own location	162	64%

Table 6: Testing location preference by actual testing location

Response	LRP*	TC^
I preferred using my own location.	146	4
I preferred going to a test centre.	3	83
I had no preference.	13	6

*Live remote proctoring (equivalent to 'own location').

^Test centre.

Candidates who indicated they tested in their own location (via live remote proctoring) responded to questions shown in Table 7 through Table 9. These candidates were generally positive about the experience and identified convenience as the main reason for choosing live remote proctoring. They were also very supportive of HRPAC continuing to offer the examination using live remote proctoring.

Table 7: Reason for choosing own location (live remote proctoring candidates)

Response	Count	%
No test centres were open in my area.	18	11%
I preferred to avoid being around other people.	13	8%
I liked the convenience of not having to travel to a test centre.	91	57%
I felt like I would perform better in my own environment.	28	17%
Other (please specify)	11	7%

Table 8: Evaluation of testing experience (live remote proctoring candidates)

	Count	%
Very positive	60	38%
Positive	59	37%
Neutral	35	22%
Negative	4	3%
Very negative	2	1%

Table 9: Value in future candidates being able to test from their own location (live remote proctoring candidates)

Response	Count	%
Yes	160	99%
No	1	1%

Candidates who indicated they tested in a test centre responded as shown in Table 10 through Table 12. These candidates were positive about being able to write at a convenient location and were also positive about their testing experience. They were also generally supportive of HRPAs continuing to offer the examination using live remote proctoring.

Table 10: Able to write at a convenient location (test centre candidates)

	Count	%
Strongly agree	42	46%
Agree	35	38%
Neither agree nor disagree	3	3%
Disagree	9	10%
Strongly disagree	3	3%

Table 11: Evaluation of testing experience (test centre candidates)

	Count	%
Very positive	35	38%
Positive	42	46%
Neutral	13	14%
Negative	2	2%
Very negative	0	0%

Table 12: Value in future candidates being able to test from their own location (test centre candidates)

Response	Count	%
Yes	83	90%
No	9	10%

Open-ended questions were also posed to candidates asking for any additional comments in general and regarding test delivery method. Those comments were provided to HRPAs for information and consideration. Nothing actionable with respect to scoring emerged in these comments.

Initial Analysis

The full CHRL-KE examination was 250 items, of which approximately 225 were to be scored. The other 20–30 items were designated as experimental. Across the 3 new forms, 230 items were available for scoring on each, after removing items designated as experimental.

The initial analysis summary statistics for the new form are presented in Table 13. The section statistics are presented in Table 14.

Table 13: Initial examination statistics

Index	CHRL-KE
Eligible items	230
Total candidates	258
Candidates in analysis	222
Mean score	156.1 (67.9%)
Score range	92–208 (40.0–90.4%)
Standard deviation	22.9
Cronbach's alpha	.92
Mean r_{pb}^*	.21

Table 14: Section item statistics

Index	Section 1	Section 2
Total items	125	125
Potentially scored items	115	115
Candidates in analysis	222	
Mean score	77.8 (67.7%)	78.2 (68.0%)
Standard deviation	11.7	12.1
Range	43–104	47–105
Mean time (minutes)*	114.3	106.6
Words	4802–4803	4803–4804

*Mean time is for all 125 items in each section.

Performance on the full exam was analyzed to identify potential differences between candidate writing in test centres and those writing using live remote proctoring. No meaningful difference was observed, with the test centre candidates performing essentially the same (65.7%) as the live remote proctoring candidates (65.5%; $t(256)=0.09$, *ns*).

Though not reported here, several additional analyses were added with administration to investigate potential candidate misconduct. These results were reported confidentially to HRP.

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 specific 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 excessively easy and excessively difficult items. Items were ranked from worst performing to best performing accordingly.

Key Validation

Key validation was conducted via web meeting on December 6, 2022, using members of the CHRL Examination Validation Committee (EVC). The EVC (Table 15) was reminded of basic item and test analysis methods and was oriented to the main statistics used to evaluate the quality of the CHRL-KE.

Table 15: CHRL Examination Validation Committee members – Key validation

Member	Credential	Years of Relevant Experience	Start on EVC	Industry
Nancy Richard, CHAIR	CHRL	15–19	2017	Canadian Nuclear Safety Commission
Jennifer King, VICE CHAIR	CHRL	20–29	2017	Banking Industry
✓ Sameera Akram	CHRL	15–19	2022	Global Science and Technology
✓ Liz Austin	CHRL	10–15	2021	Unionized
✓ Nadine Bellhouse	CHRL	15–19	2019	Printing
Jennifer Borges	CHRL	10–14	2017	Manufacturing
Tanya Dacres	CHRL	15–19	2021	Digital Business/ Transformation
Annette Dhanasar	CHRL	15–19	2017	Transportation and Technology
Maja Falarz	CHRL	5–9	2017	Stock Exchange
✓ Christine Kelsey	CHRL	5–9	2017	Media
Kelly McDonald	CHRL	15–19	2022	Port Authority
Cynthia Ogbarney-Tetteh	CHRL	15–19	2022	Municipal
✓ Karen Pantaleo	CHRL	20–29	2019	Healthcare / Consulting

✓ Participated in the session.

The group was informed that test reliability, as measured by Cronbach's alpha, was .919 based on the set of 230 potentially scored items and that this was above the generally accepted threshold of .80.

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 the group was given less direction on items with borderline statistics. Where available, candidates' comments about the items were also shown. The group made decisions based on content and the data through discussion; of the 7 items flagged with borderline statistics, 1 was removed from scoring. Panel members' comments about specific items were recorded for future item revision activities. The panel also evaluated experimental items for future use during this session.

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 in this sample of candidates. Most were moderate to strong items, however. The final alpha for the set of 229 scored items was .920. The difficulties ranged from 27.9% to 97.3%, with a mean of 67.9%. The r_{pb}^* values ranged from $-.08$ to $.48$ with a mean of $.21$.

Table 16 presents the scored CHRL-KE's final fit to the examination blueprint. In all cases, the final number of scored items in a functional area fit within the established range.

The group endorsed this final set of items for use in scoring the November 2022 CHRL-KE candidates who took this form.

Table 16: Final scored examination fit to blueprint

Functional Area		Actua	Min.	Target	Max.	Blueprint Range
10	Strategy	25	21	25	29	11% ± 2%
20	Professional Practice	25	21	25	29	11% ± 2%
30	Organizational Effectiveness	31	28	32	36	14% ± 2%
40	Workforce Planning & Talent Management	32	28	32	36	14% ± 2%
50	Labour & Employee Relations	21	17	21	25	9% ± 2%
60	Total Rewards	23	19	23	27	10% ± 2%
70	Learning & Development	25	21	25	29	11% ± 2%
80	Health, Wellness & Safe Workplace	19	14	18	22	8% ± 2%
90	HR Metrics, Reporting & Financial Management	28	23	27	32	12% ± 2%
Total		229				

Establishing the Pass Mark: Equating

Equating, as per Kolen and Brennan (2014),⁹ was used to establish the pass mark for the November 2022 CHRL-KE. The goal of this process was to set a pass mark for the November 2022 CHRL-KE that would be equivalent to that set for previous CHRL-KE 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 CHRL-KE was originally set after the November 2015 offering of the CHRL-KE using the Modified Angoff method. General details on that method can be found in Appendix B. Specific information on the standard setting session is provided in the technical report issued for the November 2015 administration.

To pass the CHRL-KE, a candidate must meet or surpass the overall test pass mark and meet or surpass the threshold set for each of the 9 functional areas. These thresholds are set independently and are described in turn.

Two equating procedures were conducted back to different administrations (May 2022 and November 2021). The intention following these equating runs was to average them to arrive at a final pass mark for the November 2022 CHRL-KE. These administrations were chosen because

⁹ Kolen, M.J., & Brennan, R.L. (2014). *Test equating, scaling, and linking*. New York, NY: Springer.

they were the most recent administration and the administration corresponding to approximately the same administration month the previous year.

Equating Back to the May 2022 Administration

Linear equating was the chosen method for setting the pass mark. Linear equating is preferred with more than 100 candidates, and equipercentile equating is preferred with more than 1,000 candidates. With candidate samples of fewer than 100, mean or circle arc¹⁰ 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 (and equivalent within each functional area) that adhered to the blueprint. Items with an increase or decrease of 10% in terms of difficulty were also removed as anchors. 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 0.67 and a mean corrected point-biserial of .23 (for November 2022 candidates).

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

¹⁰ Kim, S., & Livingston, S.A. (2010). Comparisons among small sample equating methods in a common-item design. *Journal of Educational Measurement*, 47, 286-298.

Table 17: Anchor item fit to blueprint – To May 2022

Area*	Actual	Target
10	10%	11%
20	11%	11%
30	14%	14%
40	14%	14%
50	8%	9%
60	11%	10%
70	12%	11%
80	9%	8%
90	11%	12%

*See Table 16 for the full name of each functional area.

The mean, Tucker, Levine observed-score, and circle arc methods were computed to ascertain concordance of solutions. Given the sample sizes and similarities of test parameters, Tucker equating was considered the preferred 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 November 2022 CHRL-KE scored marginally lower than the population taking the May 2022 CHRL-KE (67.2% vs. 68.1%, respectively; $t(441)=0.84$, *ns*). Because the November 2022 CHRL-KE candidates scored marginally lower (based on the anchors, non-significance notwithstanding), they would likely have a modestly lower or similar pass rate as compared to May 2022.

The equating analysis bears this out (Table 22). All methods indicate a pass mark of 143 or 144. The pass rate is, as expected, lower than what was seen in May 2022. The Tucker equating value of 143.681 was extracted from this analysis for use in establishing the final pass mark.

Table 18: Equating parameter table – Total pass mark, to May 2022

		May. 2022	Nov. 2022
Mean score	N	221	222
	Scored items	222	229
	Total	69.4%	67.9%
	Anchors	68.1%	67.2%

Table 19: Equating outcome table – Total pass mark, to May 2022

Method	Pass Mark		Pass Rate	
	Precise	Integer	All	First-time
Equating May. 2022	139.921	140	66.9%	73.3%
Tucker	143.681	144	63.6%	69.8%
Levine observed	143.773	144	63.6%	69.8%
Circle Arc 1	142.772	143	65.9%	71.2%
Circle Arc 2	142.769	143	65.9%	71.2%
Mean	143.293	144	63.6%	69.8%

Equating Back to the November 2021 Administration

Linear equating was the chosen method for setting the pass mark. Linear equating is preferred with more than 100 candidates, and equipercentile equating is preferred with more than 1,000 candidates. With candidate samples of fewer than 100, mean or circle arc¹¹ 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 (and equivalent within each functional area) that adhered to the blueprint. Items with an increase or decrease of 10% in terms of difficulty were also removed as anchors. 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 0.68 and a mean corrected point-biserial of .23 (for November 2022 candidates).

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

¹¹ Kim, S., & Livingston, S.A. (2010). Comparisons among small sample equating methods in a common-item design. *Journal of Educational Measurement*, 47, 286-298.

Table 20: Anchor item fit to blueprint – To November 2021

Area*	Actual	Target
10	11%	11%
20	11%	11%
30	13%	14%
40	15%	14%
50	9%	9%
60	10%	10%
70	10%	11%
80	9%	8%
90	11%	12%

*See Table 16 for the full name of each functional area.

The mean, Tucker, Levine observed-score, and circle arc methods were computed to ascertain concordance of solutions. Given the sample sizes and similarities of test parameters, Tucker equating was considered the preferred 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 November 2022 CHRL-KE scored marginally lower than the population taking the November 2021 CHRL-KE (67.5% vs. 68.5%, respectively; $t(458)=0.99$, *ns*). Because the November 2022 CHRL-KE candidates scored modestly lower (based on the anchors, non-significance notwithstanding), they would likely have a lower pass rate (or similar) than that observed in November 2021.

The equating analysis bears this out (Table 22). All methods indicate a pass mark of 143 or 144. The pass rate is, as expected, lower than what was seen in November 2021. The Tucker equating value of 143.589 was extracted from this analysis for use in establishing the final pass mark.

Table 21: Equating parameter table – Total pass mark, to November 2021

		Nov. 2021	Nov. 2022
Mean score	N	238	222
	Scored items	225	229
	Total	69.0%	67.9%
	Anchors	68.5%	67.5%

Table 22: Equating outcome table – Total pass mark, to November 2021

Method	Pass Mark		Pass Rate	
	Precise	Integer	All	First-time
Equating Nov. 2021	140.952	141	65.8%	74.4%
Tucker	143.589	144	63.6%	69.8%
Levine observed	143.744	144	63.6%	69.8%
Circle Arc 1	142.983	143	65.9%	71.2%
Circle Arc 2	142.982	143	65.9%	71.2%
Mean	143.213	144	63.6%	69.8%

Combined Results at Total Score Level

Table 23 shows the pass mark values across the two equating runs. The Tucker value is the one that would be selected based on sample parameters at each equating run. The weighted average (by number of anchor items and number of candidates) of the identified values, 143.635, was the recommended pass mark for the November 2022 CHRL-KE.

Using the established convention for this testing program, the weighted average value was rounded up to a cut score of 144. The resulting pass rate of 69.8% for first-time candidates is modestly lower than the values seen in May 2022 and November 2021 (see Table 24). The pass rate for all candidates was 63.6%.

Table 23: Equating outcome table – Combined results, total pass mark

	Nov. 21	May 22
Tucker	143.6	143.7
Levine observed	143.7	143.8
Circle Arc 1	143.0	142.8
Circle Arc 2	143.0	142.8
Mean	143.2	143.3

Table 24: Historical pass rates – Total pass mark

	All	1st time
Mar. 19	65.8%	76.5%
Jun. 19	64.0%	72.3%
Nov. 19	61.7%	70.3%
Mar. 20	64.9%	75.4%
Sep. 20	69.9%	76.0%
Mar. 21	66.5%	75.6%
Jun. 21	68.5%	76.6%
Nov. 21	65.8%	74.4%
May 22	66.9%	73.3%
Nov. 22	63.6%	69.8%

Functional Area Minimum Thresholds

The original functional area minimum thresholds were established in November 2015 to identify candidates who scored egregiously low on any individual functional area (see Appendix C for a conference presentation regarding this method). Since that time, equating has been employed to produce equivalent thresholds on subsequent administrations.

Tucker equating was employed for each functional area when equating back to May 2022 and November 2021 as this was the method selected for the total test score equating in those equating runs. The decisions outlined above to finalize anchor selection for the total test score equating were made so that they would also be appropriate to equating at the functional area level.

Table 25 shows alignment between anchor performance and full exam functional area score. The goal of close alignment was sufficiently achieved.

The resulting thresholds across each equating run are shown in Table 26.

Table 27 shows the outcomes and other relevant information related to equating of functional area thresholds. Note that 2 candidates failed the exam based solely on having missed the threshold for a functional area.

Table 28 shows the outcomes for each decision criterion. About 35% of the failing candidates failed at both the total score level and the functional area level; the remainder failed based only on the total score pass mark (excepting the 2 candidates who failed based on missing only the threshold on a functional area).

Table 25: Alignment between difficulty of anchors and full exam

Area	May 2022 Anchors	Nov. 2021 Anchors	Full Exam
10	64%	66%	67%
20	68%	67%	69%
30	69%	68%	64%
40	68%	69%	70%
50	64%	66%	68%
60	67%	65%	68%
70	69%	72%	69%
80	66%	70%	68%
90	68%	65%	70%

*See Table 16 for the full name of each functional area.

Table 26: Equating outcome table – Combined results, functional area thresholds

		10	20	30	40	50	60	70	80	90
To May 22	Tucker	11.75	9.83	11.07	13.32	9.03	8.49	11.58	6.36	12.09
To Nov. 21	Tucker	11.73	10.02	11.76	13.46	8.89	9.02	11.56	6.02	12.27
Weighted average		11.74	9.92	11.41	13.39	8.96	8.75	11.57	6.19	12.18
Integer		12	10	12	14	9	9	12	7	13

Table 27: Equating summary table – Functional area thresholds

Area*	Cut ⁱ	Integer ⁱⁱ	Items	Cut as %	Previous Cut % ⁱⁱⁱ	Alpha ^{iv}	Mean	Unique Fails ^v
10	11.74	12	25	47%	44%	.47	16.82	0
20	9.92	10	25	40%	37%	.53	17.19	0
30	11.41	12	31	37%	43%	.60	19.78	0
40	13.39	14	32	42%	37%	.59	22.32	0
50	8.96	9	21	43%	42%	.49	14.30	1
60	8.75	9	23	38%	38%	.63	15.55	0
70	11.57	12	25	46%	45%	.56	17.17	1
80	6.19	7	19	33%	36%	.60	12.92	0
90	12.18	13	28	44%	42%	.59	19.51	0

*See Table 16 for the full name of each functional area.

ⁱThreshold set through equating.

ⁱⁱRounded-up value of cut score as used for making candidate decisions.

ⁱⁱⁱThreshold set on previous administration.

^{iv}Cronbach's alpha for functional area.

^vNumber of candidates failing based on not meeting the functional area threshold who would have passed at the total score level.

Table 28: Passing decisions – Total pass mark and functional areas

Fails	Both measures	34	13.2%
	Total score only	60	23.3%
	Functional area score only	2	0.8%
Passes	Neither	162	62.8%

Pass Mark Approval

The total score pass mark, the thresholds for all functional areas, and the process used to derive them were presented to the CHRL EVC (Table 29) via teleconference on December 13, 2022. The committee approved the process and cut scores (which were presented along with the consequent pass rate) for recommendation to HRP.

The small reduction in pass rate was explained as likely attributable to the recent relaxing of education requirements to become eligible for the examination. It was also noted that functional area 50 was often the functional area missed when candidates failed only based on a functional area score. Wickett pointed out that very few candidates fail based on this criterion, and that all but one of the functional areas has been represented. The committee was satisfied with the discussions and explanations.

The HRP Exams Manager accepted the recommendation from the committee following the call, and the total and functional area cut scores were formally established.

Table 29: CHRL Examination Validation Committee members – Pass mark approval

Member	Credential	Years of Relevant Experience	Start on EVC	Industry
Nancy Richard, CHAIR	CHRL	15–19	2017	Canadian Nuclear Safety Commission
✓ Jennifer King, VICE CHAIR	CHRL	20–29	2017	Banking Industry
✓ Sameera Akram	CHRL	15–19	2022	Global Science and Technology
Liz Austin	CHRL	10–15	2021	Unionized
Nadine Bellhouse	CHRL	15–19	2019	Printing
✓ Jennifer Borges	CHRL	10–14	2017	Manufacturing
Tanya Dacres	CHRL	15–19	2021	Digital Business/Transformation
Annette Dhanasar	CHRL	15–19	2017	Real Estate
✓ Maja Falarz	CHRL	5–9	2017	Stock Exchange
✓ Christine Kelsey	CHRL	5–9	2017	Media
Kelly McDonald	CHRL	15–19	2022	Port Authority
Cynthia Ogbarmey-Tetteh	CHRL	15–19	2022	Municipal
Karen Pantaleo	CHRL	20–29	2019	Healthcare / Consulting

✓ Participated in the session.

Scoring

To finalize the scoring, repeat and outlier candidates who were not included in the item and form analysis were reinserted into the dataset. Scores for each of the 9 functional areas were also computed for each candidate. An Excel file with the final candidate results was provided to HRP.

Table 30 provides the means and standard deviations for the functional areas and for the total score, using all candidates who took the new November 2022 CHRL-KE forms. Table 31 provides the correlations between all functional areas. Caution should be exercised in interpreting differences between correlations. Variation can be explained largely by the number of items making up each functional area score. That is, functional areas with fewer items on the exam have lower correlations with the other functional areas. Figure 3 shows the distribution of scores for all candidates, along with the pass mark.

Table 30: Total and functional area scores for all candidates

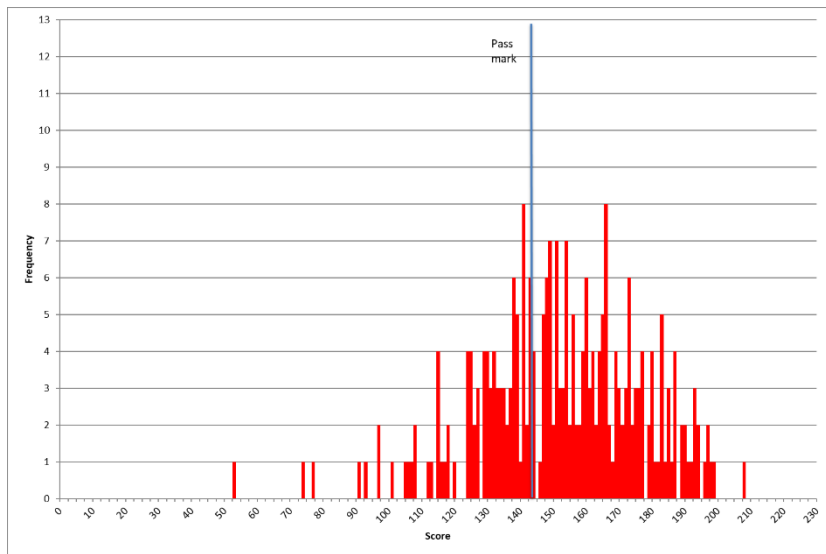
	Functional Area	Percentage	Mean	SD*
10	Strategy	66%	16.5	3.0
20	Professional Practice	67%	16.7	3.3
30	Organizational Effectiveness	62%	19.4	4.0
40	Workforce Planning & Talent Management	68%	21.8	4.0
50	Labour & Employee Relations	67%	14.0	2.9
60	Total Rewards	66%	15.2	3.5
70	Learning & Development	67%	16.8	3.4
80	Health, Wellness & Safe Workplace	67%	12.6	3.1
90	HR Metrics, Reporting & Financial Management	68%	19.0	3.8
Total score		66.4%	152.1	24.7

*SD = standard deviation.

Table 31: Correlations between functional area scores for all candidates

Area*	10	20	30	40	50	60	70	80	90
10		.52	.55	.60	.46	.50	.53	.52	.53
20			.60	.63	.59	.63	.60	.61	.63
30				.65	.49	.65	.60	.61	.63
40					.58	.68	.65	.60	.68
50						.52	.52	.56	.55
60							.60	.64	.61
70								.56	.60
80									.57
90									

*See Table 30 for the full name of each functional area.

Figure 3: Score distribution for all candidates

Key Examination Metrics

Table 32 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 32: Key examination metrics – Candidates included in analysis only

Index	November 2022	May 2022	November 2021	June 2021	March 2021
Scored items	229	222	225	224	226
Candidates	222	221	238	244	234
Mean	155.6 (67.9%)	154.0 (69.4%)	155.4 (69.0%)	155.6 (69.5%)	161.0 (71.2%)
Median	156 (68.1%)	157 (70.7%)	158 (70.2%)	157 (69.9%)	164 (72.6%)
Skewness	-0.250	-0.593	-0.421	-0.376	-0.335
Kurtosis ⁱ	-0.306	-0.061	-0.266	-0.150	-0.341
Range	91–208 (39.7– 90.8%)	82–197 (36.9– 88.7%)	93–208 (41.3– 92.4%)	84–198 (37.5– 88.4%)	94–209 (41.6– 92.5%)
Standard deviation	22.90	23.77	23.38	22.30	22.61
Cronbach's alpha	.92	.93	.93	.92	.92
Mean r_{pb}^*	.21	.23	.22	.21	.22
SEM ⁱⁱ	6.47	6.33	6.39	6.43	6.27
SEM at the pass mark	6.86	6.81	6.87	6.92	6.85
Decision consistency (uncorrected) ⁱⁱⁱ	.88	.91	.90	.89	.90
Perceived fairness ^{iv}	36%	42%	31%	34%	40%
Pass mark	143.635	139.921	140.952	140.137	143.181
Effective pass mark	144	140	141	141	144
Pass rate ^v	69.8%	73.3%	74.4%	76.6%	75.6%

ⁱExcess

ⁱⁱSEM = standard error of measurement.

ⁱⁱⁱSubkoviak method.

^{iv}Based on responses to the post-examination survey. Value here may differ from that presented in main body of report because this value includes only candidates in the analysis.

^vBased on total score criterion only.

Related Development Activities

Validation

To renew the validation of items expiring from usability, a validation session was held with the EVC (see Table 33) remotely on June 20 and 29, 2022.

Note that scheduling precluded all members from being available for both days; those marked as having participated attended the validation activity on at least 1 day.

Table 33: CHRL Examination Validation Committee members – Validation

Member	Credential	Years of Relevant Experience	Start on EVC	Industry
✓ Nancy Richard, CHAIR	CHRL	15–19	2017	Canadian Nuclear Safety Commission
✓ Jennifer King, VICE CHAIR	CHRL	20–29	2017	Banking Industry
✓ Sameera Akram	CHRL	15–19	2022	Global Science and Technology
✓ Liz Austin	CHRL	10–15	2021	Unionized
✓ Nadine Bellhouse	CHRL	15–19	2019	Printing
✓ Jennifer Borges	CHRL	10–14	2017	Manufacturing
Tanya Dacres	CHRL	15–19	2021	Digital Business/Transformation
✓ Annette Dhanasar	CHRL	15–19	2017	Real Estate
✓ Maja Falarz	CHRL	5–9	2017	Stock Exchange
Christine Kelsey	CHRL	5–9	2017	Media
✓ Kelly McDonald	CHRL	15–19	2022	Port Authority
✓ Cynthia Ogbarmey-Tetteh	CHRL	15–19	2022	Municipal
✓ Karen Pantaleo	CHRL	20–29	2019	Healthcare / Consulting Education –
✓ Laurie Torno	CHRL	20–29	2018	University focus on pensions/benefits/compensation

✓ Participated in the session.

The EVC members received advance materials outlining:

- Purpose of the session
- Description of the CHRL credential
- CHRL-KE blueprint

- Criteria for good test items
- Validation process

The committee members received refresh training on the validation activity on the first day of the session. For participants not able to join on the first day, they received individual training on the first day of their involvement. Each day, committee members were provided with 45 items via a secure file share site, and then worked individually reviewing items through the day, submitting their appraisal and any suggested revisions to Wickett through the day. They were directed to make sure the items reflected current practice and were suitable to make decisions about who should receive the CHRL credential.

At the end of each day, the committee convened online and were shown items flagged for revision. Where committee members proposed changes, these were discussed by the group before implementation.

For each item, the committee was asked to either:

- Validate the item for use in the next three years to make decisions about who would be certified as CHRL
- Move the item to the CHRP-KE or ELE bank
- Revise the item to make it suitable for use
- Declare the item unsound and send it back for revision or removal from the bank

The committee validated 87 items as suitable for the CHRL-KE, rejected 3 items, and shifted 0 items for eligibility consideration in the CHRP-KE or ELE bank. Twenty-nine items were revised prior to validation as part of this exercise. The committee also verified the functional area and competency for all items, and added rationales and references where missing, incomplete, or not current.

Item Writing

To fill gaps in the bank and renew content, item writing was conducted in July–September 2022. Item writers (see Table 34) were identified by HRP and trained in a remote session by Wickett on July 25, 27, and 28, 2022 (multiple sessions held to accommodate item writer schedules). Items were written for both the CHRP-KE and CHRL-KE during this activity.

Table 34: Item writers

Writer	Credentials	Location	Years of HR Experience	Industry
Cal Barber	B. Comm, MBA	Toronto	40	Education
Nicole Bonenfant	CHRP, BA, MIRHR	Hamilton, Toronto	12	Health care, Education
Trina McGarvey	MIR, CHRL	Napanee	20	Government
Steve Risavy	PhD, MASc, CHRL, CHRP	Waterloo	11	Education
Michelle Soulliere	BSc, MPA	Hamilton	13	Education

The item writers were provided with training via teleconference, and received additional materials covering the main elements of the training. The general guidance for writing quality multiple choice items was drawn primarily from Haladyna & Rodriguez (2013).¹²

Each item writer was selected based on expertise in identified functional areas, and they were assigned items within those functional areas. More specifically, each item writer was assigned competencies (drawn from the *HRPA Professional Competency Framework* [2014]) that were to be the focus of their items. Item writers were assigned 25 items to write each, but in the end the actual completion counts ranged from 15 to 40, for a total of 140 items. A 6th item writer dropped out before writing any items and the original total count target was 150 items.

The item writers had access to the style guide that governs language usage on the HRPAs exams and were provided with recent electronic textbooks as necessary. Item writers were required to include at least one authoritative source to back up each test item, and also provide rationales for the correct and incorrect answers.

Each item writer worked remotely, sending items to Wickett for review and comment via a secure file share site. Items were exchanged until such time as the item writer was comfortable with the content and Wickett was comfortable that the item would be successful at validation and upon use with candidates. This generally required several iterations per item.

Once all items were drafted and declared complete, they went through a reference verification to ensure the references were accurate and backed up each item. After that, they were sent to a certified professional editor for editorial. Items were adjusted based on this input and comments noted if future reviewers would need to attend to specific content concerns.

Validation

To renew the validation of items expiring from usability, a validation session was held with the EVC (see Table 33) remotely on November 8 and 10, 2022.

¹² Haladyna, T. M., & Rodriguez, M.C. (2013). *Developing and validating test items*. New York, NY: Routledge.

Note that scheduling precluded all members from being available for both days; those marked as having participated attended the validation activity on at least 1 day.

Table 35: CHRL Examination Validation Committee members – Validation

Member	Credential	Years of Relevant Experience	Start on EVC	Industry
✓ Nancy Richard, CHAIR	CHRL	15–19	2017	Canadian Nuclear Safety Commission
✓ Jennifer King, VICE CHAIR	CHRL	20–29	2017	Banking Industry
✓ Sameera Akram	CHRL	15–19	2022	Global Science and Technology
✓ Liz Austin	CHRL	10–15	2021	Unionized
✓ Nadine Bellhouse	CHRL	15–19	2019	Printing
✓ Jennifer Borges	CHRL	10–14	2017	Manufacturing
✓ Tanya Dacres	CHRL	15–19	2021	Digital Business/Transformation
✓ Annette Dhanasar	CHRL	15–19	2017	Real Estate
✓ Maja Falarz	CHRL	5–9	2017	Stock Exchange
✓ Christine Kelsey	CHRL	5–9	2017	Media
✓ Kelly McDonald	CHRL	15–19	2022	Port Authority
Cynthia Ogbarmey-Tetteh	CHRL	15–19	2022	Municipal
✓ Karen Pantaleo	CHRL	20–29	2019	Healthcare / Consulting

✓ Participated in the session.

The EVC members received advance materials outlining:

- Purpose of the session
- Description of the CHRL credential
- CHRL-KE blueprint
- Criteria for good test items
- Validation process

The committee members received refresh training on the validation activity on the first day of the session. For participants not able to join on the first day, they received individual training on the first day of their involvement. Each day, committee members were provided with 47 items via a secure file share site, and then worked individually reviewing items through the day, submitting their appraisal and any suggested revisions to Wickett through the day. They were directed to make sure the items reflected current practice and were suitable to make decisions about who should receive the CHRL credential.

At the end of each day, the committee convened online and were shown items flagged for revision. Where committee members proposed changes, these were discussed by the group before implementation.

For each item, the committee was asked to either:

- Validate the item for use in the next three years to make decisions about who would be certified as CHRL
- Move the item to the CHRP-KE or ELE bank
- Revise the item to make it suitable for use
- Declare the item unsound and send it back for revision or removal from the bank

The committee validated 89 items as suitable for the CHRL-KE, rejected 5 items, and shifted 0 items for eligibility consideration in the CHRP-KE or ELE bank. Ten items were revised prior to validation as part of this exercise. The committee also verified the functional area and competency for all items, and added rationales and references where missing, incomplete, or not current.

Appendix A

Blueprint

CHRL Knowledge Examination

Human Resources Professionals Association

Version 2.0

Approved by CHRL Exam Validation Committee March 13, 2018

Approved by HRPA Registrar March 14, 2018

Effective June 2018 administration

Credentials

Passing the CHRL Knowledge Examination (CHRL-KE) is a requirement for certification for CHRL candidates. The examination reflects the *HRPA Professional HR Competency Framework* (2014).

Purpose

The CHRL-KE assesses whether a candidate has the level of discipline-specific knowledge necessary to practise human resources management at the CHRL level in a manner that is consistent with the protection of the public interest. Knowledge related exclusively to employment and workplace legislation is assessed on the CHRL Employment Law Examination.

The CHRL credential requires candidates to demonstrate competence across all 9 functional areas, and the CHRL-KE operationalizes this by requiring demonstration of proficiency at both the total score level and on each functional area. Very low performance on any functional area (as defined through standard setting with a confidence threshold adjustment at the 95% level) is taken as evidence of not demonstrating the required level of competence to earn the CHRL.

Structure

The structural variables provide high-level guidance as to what the examination will be like.

Table 36: CHRL-KE Blueprint structural variables

Item types	Independent 4-option multiple choice
Length	250 items in total
	20–30 experimental items
Duration	Up to 5 hours
Delivery mode	Computer-based testing in proctored test centres
Frequency	3 windows per year

Content Weighting

The functional area weights were set in 2014 through a national survey and modified slightly in 2018 to remove weighting for competencies most appropriately tested on the CHRL

Employment Law Examination. Within each functional area, items are distributed roughly evenly across the related competencies.

Table 37: Functional area weights on the CHRL-KE

Functional Area		Weight	Range
10	Strategy	11%	+/- 2%
20	Professional Practice	11%	+/- 2%
30	Organizational Effectiveness	14%	+/- 2%
40	Workforce Planning & Talent Management	14%	+/- 2%
50	Labour & Employee Relations	9%	+/- 2%
60	Total Rewards	10%	+/- 2%
70	Learning & Development	11%	+/- 2%
80	Health, Wellness & Safe Workplace	8%	+/- 2%
90	Human Resources Metrics, Reporting & Financial Management	12%	+/- 2%

Table 38: Competencies not eligible on the CHRL-KE

FA	Comp
20	C035
	C036
	C037
50	C117
60	C139
80	C177
	C179
90	C204
	C205

Minor amendments made October 22, 2018, by CHRL EVC, with approval of the Registrar.

Appendix B

MODIFIED ANGOFF METHOD

WHAT IT IS → The Modified Angoff method of setting cut scores is the most popular method used with high-stakes examinations. With this method, experts evaluate each item on a test for difficulty and judge how likely it is that someone who is borderline in performance will get each item correct. Borderline candidates have, by definition, just enough competence to be considered competent (e.g., to pass the test). Any candidate showing the same or a higher level of performance as a borderline candidate is thus a “passing” candidate, and any candidate showing performance below the level of a borderline candidate is a “failing” candidate. The method has been successfully defended in court as being a fair method of setting cut scores that are used to make high-stakes decisions about candidates.

HOW IT'S DONE → The Modified Angoff method typically requires 5 to 15 experts in the field and is facilitated by a psychometrician. There are many variations of the Modified Angoff method used in practice, but generally the process begins with detailed training on how to apply ratings, followed by development of a description of the borderline candidate. Once training is complete (including a calibration exercise to make sure all raters have fully grasped the method), ratings are applied individually by each rater and compiled by the psychometrician. Discrepancies across raters are identified and flagged for discussion. Raters then have an opportunity to discuss their ratings and to rerate any items if the new information is considered cause to do so. In some cases, the psychometrician will introduce data from previous administrations of the item to further refine judgments. Once all items have been rated, an average Angoff rating for the exam is calculated by simply taking the average of all item ratings. The result is the cut score for the exam as a whole.

WHY IT'S USED → The benefit of the Modified Angoff method is that the resulting cut scores set an objective hurdle for candidates. Candidates who demonstrate performance above the borderline level (as systematically established by experts) are considered to have sufficient competence, and those below that level are considered to have insufficient competence. The proportion of candidates deemed below or above the cut score is not arbitrary and depends only on the actual ability of those candidates. For examinations resulting in pass/fail decisions, the implication of this is that all candidates would pass if they all showed better than the minimal accepted level of competence (i.e., above the borderline), or they would all fail if they all showed less than the minimal accepted level of competence. What is important is whether each candidate scores above or below the cut score, with that cut score being set based on the actual difficulty of the test and the expected performance of candidates showing the lowest level of acceptable performance. Because of this, the Modified Angoff method fairly assesses individual candidates on their own merits.

References

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



© 2015 Wickett Measurement Systems Inc.

Appendix C



Human Resources
Professionals
Association



Requiring candidates to pass all sub-tests on a certification exam (aka, non-compensatory scoring of certification exams)

Claude Balthazard, Vice-President, Regulatory Affairs and Registrar, Human Resources Professionals Association

John Wickett, Lead consultant and Principal, Wickett Measurement Systems Inc.

The challenge	Directive to require candidates to achieve thresholds in nine functional areas, in addition to an overall threshold, before a pass result will be granted. A candidate who passes overall, but who fails just one of nine functional areas, will fail and must retake the entire test.
The facts	<ol style="list-style-type: none"> 1. Brand new high-stakes certification exam. 2. Exam with 225 scored four-option multiple-choice items. 3. Each functional area has 18 to 31 items, depending on blueprint weight.
The issues	<ol style="list-style-type: none"> 1. Pass/fail decisions will need to be made based on subscores with as few as 18 items. 2. Decisions need to be defensible and candidate appeal must be anticipated.
What we did	<ol style="list-style-type: none"> 1. Standard two-round Modified Angoff with eight judges conducted after initial administration. 2. Overall pass mark established using mean of all Angoffed values, with no adjustments. Pass mark was 138.5 out of 225, yielding a pass rate of 68.8%. 3. To calculate threshold for each functional area: <ol style="list-style-type: none"> a. Calculate the conditional standard error of measurement around the mean Angoff value for the functional area using the Lord method.¹ b. Multiply the CSEM by 2.417 to provide 95% one-tailed confidence across all nine comparisons.² This is equivalent to 99.22% confidence for each independent comparison. c. Subtract the resulting value from the mean Angoff value for the functional area. d. Use the rounded-up integer of this resulting value as the cut score for that functional area. 4. Based on only the functional area thresholds, nine additional candidates failed the exam. Thresholds ranged from 30% to 50% across functional areas, well below the mean performances (ranging from 57% to 73%).
What this accomplished	<ol style="list-style-type: none"> 1. Candidates cannot pass the examination if they are <i>substantially</i> unknowledgeable in any one area. The format forces candidates to be generalists to at least some extent and not rely on strengths in a few areas. 2. Candidates who know their stuff across the board, with no areas of extreme weakness, will pass . . . exactly in line with the goals of the program.
Considerations for others	<ol style="list-style-type: none"> 1. Consider explicitly how pass/fail decisions will be prioritized. <ol style="list-style-type: none"> a. In this case, for the overall score, a balance was struck where errors on either side of the pass mark were balanced. b. For the functional area thresholds, however, the priority was placed on <i>not</i> failing someone based on any one function area unless we were more than 95% sure. 2. The functional areas all had lower reliabilities (.44 to .71) than the overall score (.92), but this was accounted for by the CSEM adjustment. So while it is true that making decisions solely on subscores with so few items would be problematic, doing so in conjunction with an appropriate overall score pass mark may help achieve program goals.

¹ Feldt, L.S., Steffen, M. & Gupta, N.C. (1985). A comparison of five methods for estimating the standard error of measurement at specific score levels. *Applied Psychological Measurement*, 9, 251-361.

² Gupta, S.S. (1963). Probability integrals of multivariate normal and multivariate t. *The Annals of Mathematical Statistics*, 34, 792-828.