

Technical Report: August 2020 CKE 1

HR | Human Resources
PA | Professionals Association

8 November 2020



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	14
Key Validation.....	15
Establishing the Pass Mark: Equating.....	17
Scoring	25
Key Examination Metrics	28
Related Development Activities	29
Item Writing	29
Validation.....	30
Item Review.....	32
Appendix A	34
Appendix B	36

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	12
Table 5: Preferred location (live remove proctoring candidates)	13
Table 6: Reason for choosing own location (live remove proctoring candidates)	13
Table 7: Evaluation of testing experience (live remove proctoring candidates)	13
Table 8: Value in future candidates being able to test from their own location (live remote proctoring candidates)	13
Table 9: Able to write at a convenient location (test centre candidates)	14
Table 10: Value in future candidates being able to test from their own location (test centre candidates)	14
Table 11: Initial examination statistics – Combined across forms	15
Table 12: CHRP Examination Validation Committee members – Key validation	16
Table 13: Final scored examination fit to blueprint	17
Table 14: Anchor item fit to blueprint – To June 2019	18
Table 15: Equating parameter table – Total pass mark, to June 2019	19
Table 16: Equating outcome table – Total pass mark, to June 2019	19
Table 17: Anchor item fit to blueprint – To October 2019	20
Table 18: Equating parameter table – Total pass mark, to October 2019	21
Table 19: Equating outcome table – Total pass mark, to October 2019	21
Table 20: Anchor item fit to blueprint – To February 2020	22
Table 21: Equating parameter table – Total pass mark, to February 2020	23
Table 22: Equating outcome table – Total pass mark, to February 2020	23
Table 23: Equating outcome table – Combined results	24
Table 24: Historical pass rates	25
Table 25: CHRP Examination Validation Committee members – Pass mark approval	25
Table 26: Total and functional area scores for all candidates	26
Table 27: Correlations between functional area scores for all candidates	27
Table 28: Key examination metrics – Candidates included in analysis only	28
Table 29: Item writers	29

Table 30: CHRP Examination Validation Committee members – Validation.....30

Table 31: Panel for item review session.....32

Table 32: CKE 1 Blueprint structural variables34

Table 33: Functional area weights on the CKE 1.....35

Table 34: Competencies not eligible on the CKE 135

List of Figures

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

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

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

Executive Summary¹

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

Special note: This administration replaced the normally scheduled administrations in June and October 2020 because of COVID-19 closures.

The Comprehensive Knowledge Exam 1 (CKE 1) was administered to 310 candidates² using computer-based testing at Prometric test centres August 31–September 15, 2020, inclusive. The examination comprised 175 four-option multiple choice items and had a 3½-hour time limit.

As per the CKE 1 blueprint, the exam was scored using the 145–155 best-performing items (while adhering to the prescribed distribution across functional areas). The mean score for first-time candidates³ ($n=241$) was 106.6 (71.1%), and for all candidates it was 103.6 (69.1%), out of 150 scored items. Reliability was strong at .88. The final set of scored items adhered to the blueprint parameters.

The pass mark was set using equating back to the June 2019, October 2019 and February 2020 administrations, yielding an integer pass mark of 96. 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 CKE 1. This pass mark resulted in a pass rate for first-time candidates of 75.9% and a pass rate for all candidates of 70.1%.

This report, the analyses performed, and the processes followed are consistent with NCCA standards⁴ and ISO 17024 standards.⁵

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

² Excludes 15 candidates who wrote the examination late due to technical issues.

³ 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.

⁴ National Commission for Certifying Agencies (2014). *Standards for the accreditation of certification programs*. Washington, DC: Institute for Credentialing Excellence.

⁵ 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 3 new 175-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 year
2. Fitting the total item count of 175
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.

The final form composition for the August 2020 CKE 1 forms is shown in Table 1. All functional areas are within the limits of their targets, and therefore the forms reflect the blueprint (see Appendix A for the CKE 1 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	7	7
20	Professional Practice	20	19
30	Organizational Effectiveness	21–22	23
40	Workforce Planning & Talent Management	21–22	23
50	Labour & Employee Relations	19–20	19
60	Total Rewards	21–22	23
70	Learning & Development	21–22	23
80	Health, Wellness & Safe Workplace	22	19
90	HR Metrics, Reporting & Financial Management	19–20	19
	TOTAL	175	175

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 August 31–September 15, 2020, inclusive, and 310 candidates wrote the exam. Due to several technical issues, 15 additional candidates wrote the examination after the close of the testing window, and due to timing issues were not included in the item and form analyses that follow.

Candidates were able to select either a test centre (assuming one was available reasonably close to them) or live remote proctoring from a location of their choosing. Standard security methods (as per Prometric protocols⁶) were employed for both methods. Candidates were allowed up to one unscheduled 10-minute break during the examination (the examination timer was not stopped during this break).

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

⁶ 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 roster provided by Prometric to ensure that all .xml files had been received. Further, each candidate total score as computed by Prometric was reconciled with that computed by Wickett for the full set of 175 items to verify key accuracy. Comments on items were also reviewed to identify any specific item-level issues. No problems were encountered.

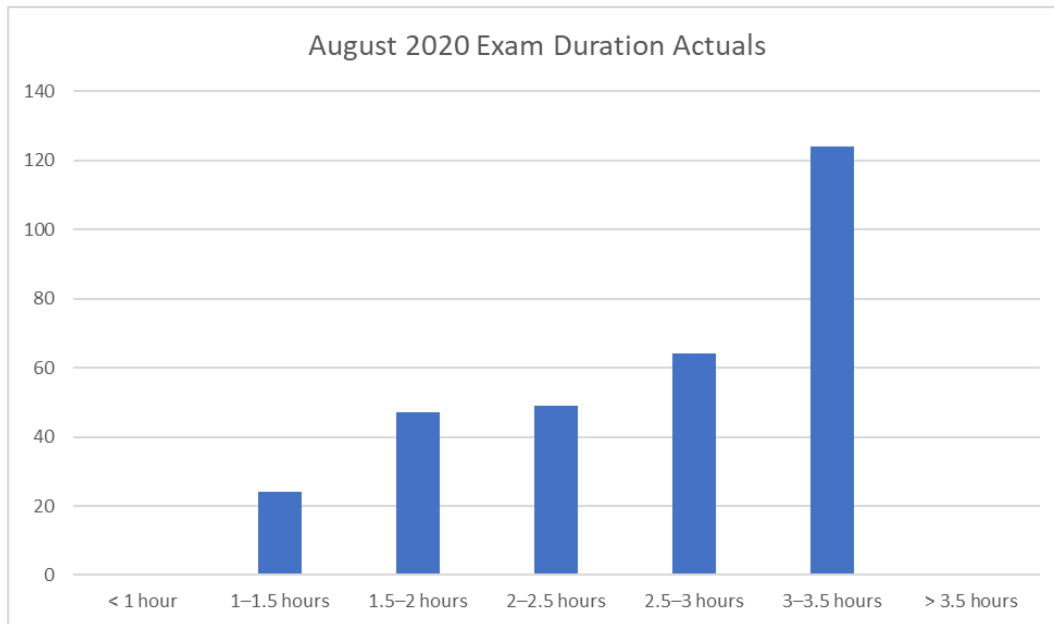
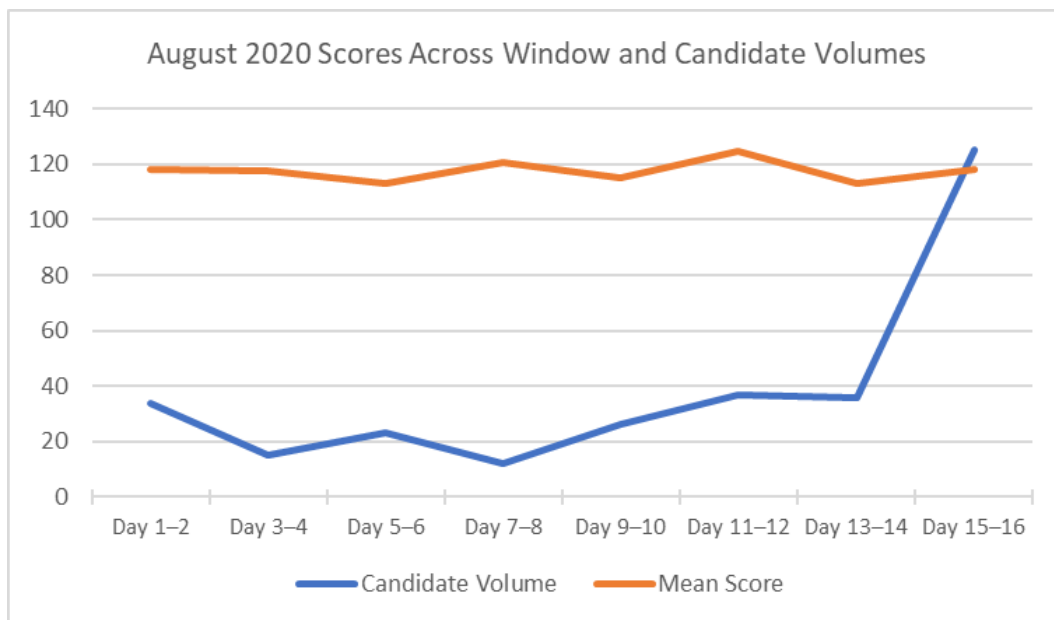
The average time taken by all candidates was assessed to detect potential examination timing concerns. The distribution is shown in Figure 1. The mean was 2 hours, 38 minutes (1 minute more than in February 2020; on average, form A candidates took 2 hours, 37 minutes, form B candidates took 2 hours, 42 minutes, and form C candidates took 2 hours, 35 minutes). The time limit on the CKE 1 was 3½ hours, suggesting that time was not a factor in scores across candidates. Though some candidates were granted additional time as a testing accommodation, none of these candidates exceeded the time limit of 3½ hours.

Sixteen candidates (5%) took the full 3½ hours, suggesting that those candidates may have wanted more time, and 5 candidates (1.6%) 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 correlation between scores on the 175 items and time spent writing the examination was small at a value of $-.21$ for form A, negligible at a value of $.05$ for form B, and small at a value of $.14$ for form C, suggesting that time constraints did not generally have an impact on candidate performance.

Candidate scores across the window were computed to look for any evidence of item exposure. As shown in Figure 2, there was little variation across the window. The difference between scores for candidates writing in the first 2 days and those writing in the last 2 days was an increase of 0.3 marks out of 175.

As a matter of interest, candidate volumes were also examined across the window; these are also shown in Figure 2. Though not psychometrically meaningful, there is a clear pattern for candidates to prefer to book towards the end of the window rather than the start (though some of this effect is due to required rescheduling of candidates from earlier in the window to the end).

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 175 items. Two candidates were flagged for an abnormally low or high score (z value outside ± 3.0). Also, the 175 items were arbitrarily broken into 7 blocks of 25 items for each candidate; the 7 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 5 or more blanks were also flagged for removal from analysis (no candidates were flagged on this criterion). To be conservative, candidates who had been granted a testing accommodation with changed administration conditions were also removed from the main analysis (simply because their testing conditions were not the same as the main group of candidates, even though each accommodation was granted on the premise that it would make the testing experience equivalent in terms of opportunity to demonstrate competence). As a result of all of these factors, 7 candidates were removed from analysis.

Candidates who had failed a previous HRP A examination (CKE, CKE 1, or CKE 2) scored lower than did those who had not (60.3% and 69.1%, respectively, on the full exam of 175 items). This difference was meaningful and significant ($t(125)=7.94$, $p<.001$). In keeping with standard procedures, these candidates were removed from subsequent analyses. The CKE 1 analysis proceeded with 241 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 CKE 1; 304 responses were obtained from candidates (response rate, 99%). The survey was revised for the August administration to collect information specifically applicable to the use of live remote proctoring.

Table 2 shows 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 responses to the administration-related questions. Note that 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.	179	104	8	10	3	4.5	93%	90%
2.	Information available prior to exam day provided me with adequate details about the content and format of the exam.	100	128	36	33	7	4.0	75%	66%
3.	I feel I was adequately prepared to write this examination.	22	123	107	42	10	3.7	48%	53%
4.	The questions in the examination were clearly written.	44	159	65	35	1	3.9	67%	61%
5.	The terminology used in the examination was accurate.	41	175	69	18	0	4.0	71%	72%
6.	The situations presented in the examination were realistic.	59	195	41	7	1	4.1	84%	81%
7.	The questions in the examination reflected the examination blueprint.	36	140	93	30	2	3.9	58%	57%
8.	The examination was a fair assessment of my ability.	33	108	102	49	8	3.7	47%	48%

*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.	152	122	12	11	5	4.4	91%	<i>Revised Aug. 2020</i>
10.	I was well informed about the examination rules and regulations.	171	119	7	5	0	4.5	96%	<i>Revised Aug. 2020</i>
11.	Proctors enforced the exam-day rules.	194	102	5	2	0	4.6	98%	<i>Revised Aug. 2020</i>
12.	Proctors were professional and courteous.	212	84	2	5	0	4.7	98%	96%
13.	The tutorial helped me understand how to complete the examination on the computer.	185	96	19	2	1	4.6	93%	95%
14.	Navigation through the examination was easy and intuitive.	182	110	9	1	1	4.6	96%	97%

*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 wrote the examination, and based on their response the questions that followed differed (see Table 4).

Table 4: Testing location

Response	Count	%
Test centre	54	18%
Own location	248	82%

Candidates who indicated they tested in the own location (via live remote proctoring) responded to questions shown in Table 5 through Table 8. These candidates favoured using their own location versus a test centre, were generally positive about the experience and felt that HRPAs should continue to offer the option in the future. As expected, COVID-19 related concerns were a motivating factor for many in choosing live remote proctoring (shown primarily in the 'other' write in responses), though a substantial portion identified the convenience of not having to travel as a main driver.

Table 5: Preferred location (live remove proctoring candidates)

Response	Count	%
I preferred using my own location.	171	69%
I preferred going to a test centre.	36	15%
I have no preference.	40	16%

Table 6: Reason for choosing own location (live remove proctoring candidates)

Response	Count	%
No test centres were open in my area.	39	16%
I preferred to avoid being around other people.	46	19%
I liked the convenience of not having to travel to a test centre.	91	37%
I felt like I would perform better in my own environment.	26	11%
Other (please specify)	45	18%

Table 7: Evaluation of testing experience (live remove proctoring candidates)

	Count	%
Very positive	90	36%
Positive	92	37%
Neutral	50	20%
Negative	9	4%
Very negative	6	2%

Table 8: Value in future candidates being able to test from their own location (live remote proctoring candidates)

Response	Count	%
Yes	236	96%
No	9	4%

Candidates who indicated they tested in a test centre responded as shown in Table 9 and Table 10. These candidates were positive about being able to write at a convenient location, and were also supportive of HRPAs continuing to offer the option of writing using live remote proctoring in the future.

Table 9: Able to write at a convenient location (test centre candidates)

	Count	%
Strongly agree	25	46%
Agree	20	37%
Neither agree nor disagree	2	4%
Disagree	4	7%
Strongly disagree	3	6%

Table 10: Value in future candidates being able to test from their own location (test centre candidates)

Response	Count	%
Yes	48	89%
No	6	11%

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 HRP A for information and consideration. Nothing actionable with respect to scoring emerged in these comments.

Initial Analysis

The full CKE 1 examination was 175 items, of which approximately 150 were to be scored. The other 20–30 items were not intended to be scored. Across the 3 new forms, 151 items were available for scoring on each, after removing items designated as experimental.

The initial analysis summary statistics are presented in Table 11.

Table 11: Initial examination statistics – Combined across forms

Index	Value
Items	151
Total candidates	308
Candidates in analysis	241
Mean score	107.1 (71.0%)
Score range	68–134 (45.0–88.7%)
Cronbach's alpha	.88
Mean r_{pb}^*	.20

A simple comparison between scores obtained by test centre candidates (mean score of 67.1%) and live remote proctoring candidates (mean score of 69.4%) was made to evaluate if there was any problematic difference in performance. The small number of candidates means this analysis is inconclusive, but a lack of significant effect ($t(306)=1.45$, *ns*) is at least supportive of there being no overall difference in candidate performance. There is a leaning towards higher scores from live remote proctoring candidates that will be monitored across future administrations.

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 and on difficulty through removal of ineffective very easy or very hard items. Items were ranked from worst performing to best performing accordingly.

Key Validation

Key validation was conducted via web meeting on September 18, 2020, using members of the CERP Examination Validation Committee (EVC). The EVC (Table 12) was reminded of basic item and test analysis methods and was oriented to the main statistics used to evaluate the quality of the CKE 1.

Table 12: CHRP Examination Validation Committee members – Key validation

Member	Credential	Years of Relevant Experience	Joined EVC	Industry
Sunday Ajao	CHRL	15–20	2017	Banking/finance
✓ Roxanne Chartrand	CHRL	20–29	2018	Insurance
Claire Chester	CHRL	10–15	2017	Health services
✓ Tanya Gopaul	CHRL	10–15	2017	Banking
Jean Lazarus	CHRL	15–19	2017	Health services
Suman Seth	CHRL	15–19	2018	Government
✓ Kriss Stone	CHRL	10–15	2017	Real estate
✓ Ielean Tait	CHRL	15–20	2017	Environmental
✓ Patricia Verkley	CHRL	10–15	2019	Not-for-profit
✓ Karen Weiler	CHRL	20–29	2017	Software/ communications

✓ Participated in the session.

The group was informed that test reliability, as measured by Cronbach's alpha, was .88 based on the set of 151 potentially scored items and that this was well 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. Because of the modest sample size for this administration, past item data were also used where available, and the group was directed not to rely unduly on statistics exclusively from the August administration.

The group made decisions based on content and the data through discussion; they removed the 1 item that they felt was inappropriate to retain for scoring. Panel members' comments about specific items were recorded for future item revision activities. The group also reviewed and made decisions about the future use of experimental items.

Not all remaining items were strong-performing, and several items were retained that were easy or 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 150 scored items was .88. The difficulties ranged from 33.2% to 96.3%, with a mean of 71.1%. The r_{pb}^* values ranged from $-.09$ to $.44$, with a mean of $.20$.

Table 13 presents the scored CKE 1'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 the final set of items for use in scoring the August 2020 CKE 1 candidates who took this form.

Table 13: Final scored examination fit to blueprint

Functional Area	Actual	Min.	Target*	Max.	Blueprint Range
10 Strategy	6	5	6	7	4% ± 1%
20 Professional Practice	16	14	17	19	11% ± 2%
30 Organizational Effectiveness	19	17	20	22	13% ± 2%
40 Workforce Planning & Talent Management	19	17	20	22	13% ± 2%
50 Labour & Employee Relations	17	14	17	19	11% ± 2%
60 Total Rewards	19	17	20	22	13% ± 2%
70 Learning & Development	19	17	20	22	13% ± 2%
80 Health, Wellness & Safe Workplace	18	14	17	19	11% ± 2%
90 HR Metrics, Reporting & Financial Management	17	14	17	19	11% ± 2%
Total	150				

*Adds to 154 due to rounding.

Establishing the Pass Mark: Equating

Equating, as per Kolen and Brennan (2014),⁷ was used to establish the pass mark for the August 2020 CKE 1. The goal of this process was to set a pass mark for the August 2020 CKE 1 that would be equivalent to that set for previous CKE 1 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 CKE 1 was originally set after the November 2015 offering of the CKE 1 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.

Three equating procedures were conducted back to different administrations (June 2019, October 2019, and February 2020). Three separate procedures were conducted to reduce the effects of sample variability and arrive at the most accurate equated pass mark.

⁷ Kolen, M.J., & Brennan, R.L. (2014). *Test equating, scaling, and linking*. New York, NY: Springer.

Equating Back to the June 2019 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 arc⁸ 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 0.71 and a mean corrected point-biserial of .26 (for August 2020 candidates).

Table 14 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. Note that though Functional Area 10 did not yield a usable anchor item, only 1 item is actually expected and so this is not greatly discordant.

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

Area*	Actual	Target
10	0%	4%
20	12%	11%
30	15%	13%
40	9%	13%
50	9%	11%
60	12%	13%
70	15%	13%
80	15%	11%
90	12%	11%

*See Table 13 for the full name of each functional area.

⁸ Kim, S., & Livingston, S.A. (2010). Comparisons among small sample equating methods in a common-item design. *Journal of Educational Measurement*, 47, 286-298.

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, though an equal case could have been made for selecting the Levine observed-score method.

Table 15 shows some of the parameters used to derive the equating estimates, along with other parameters describing the test forms. Of note is that on the anchor items, the candidates taking the August 2020 CKE 1 scored higher than the candidates taking the June 2019 CKE 1 (71.2% vs. 67.2%; $t(367)=2.54$, $p<.05$). Because the August 2020 CKE 1 candidates scored higher (based on the anchors), they would likely have a higher pass rate as compared to June 2019 candidates.

The equating analysis bears this out (Table 16). All methods indicate a pass mark of 94 to 96. The first-time candidate pass rate based on this equating run is higher, as expected, than what was seen in June 2019. The Tucker equating value of 95.38 was extracted from this analysis for use in setting the final pass mark, though it was noted that the Levine observed value was meaningfully lower.

Table 15: Equating parameter table – Total pass mark, to June 2019

		Jun. 2019	Aug. 2020
	N	128	241
	Scored items	150	150
Mean score	Total	67.4%	71.1%
	Anchors	67.2%	71.2%

Table 16: Equating outcome table – Total pass mark, to June 2019

Method	Pass Mark		Pass Rate	
	Precise	Integer	All	First-time
Equating Jun. 2019	93.25	94	56.6%	65.6%
Tucker	95.38	96	70.1%	75.9%
Levine observed	93.91	94	73.7%	79.3%
Mean	94.90	95	71.1%	76.8%
Circle Arc 1	94.53	95	71.1%	76.8%
Circle Arc 2	94.53	95	71.1%	76.8%

Equating Back to the October 2019 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 arc⁹ 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 0.71 and a mean corrected point-biserial of .24 (for August 2020 candidates).

Table 17 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.

Table 17: Anchor item fit to blueprint – To October 2019

Area*	Actual	Target
10	6%	4%
20	15%	11%
30	12%	13%
40	12%	13%
50	12%	11%
60	12%	13%
70	9%	13%
80	12%	11%
90	12%	11%

*See Table 13 for the full name of each functional area.

⁹ Kim, S., & Livingston, S.A. (2010). Comparisons among small sample equating methods in a common-item design. *Journal of Educational Measurement*, 47, 286-298.

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 18 shows some of the parameters used to derive the equating estimates, along with other parameters describing the test forms. Of note is that on the anchor items, the candidates taking the August 2020 CKE 1 scored modestly higher than the candidates taking the October 2019 CKE 1 (71.0% vs. 69.2%; $t(414)=1.39$, *ns*). Because the August 2020 CKE 1 candidates scored modestly higher (non-significance notwithstanding), they would likely have a modestly higher pass rate as compared to October 2019 candidates.

The equating analysis bears this out (Table 19). All methods indicate a pass mark of 96. The pass rate based on this equating run is higher, as expected, than what was seen in October 2019. The Tucker equating value of 95.95 was extracted from this analysis for use in setting the final pass mark.

Table 18: Equating parameter table – Total pass mark, to October 2019

		Oct. 2019	Aug. 2020
	N	175	241
	Scored items	153	150
Mean score	Total	69.5%	71.1%
	Anchors	69.2%	71.0%

Table 19: Equating outcome table – Total pass mark, to October 2019

Method	Pass Mark		Pass Rate	
	Precise	Integer	All	First-time
Equating Oct. 2019	97.50	98	66.2%	74.3%
Tucker	95.95	96	70.1%	75.9%
Levine observed	95.11	96	70.1%	75.9%
Mean	95.85	96	70.1%	75.9%
Circle Arc 1	95.84	96	70.1%	75.9%
Circle Arc 2	95.83	96	70.1%	75.9%

Equating Back to the February 2020 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 0.71 and a mean corrected point-biserial of .22 (for August 2020 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.

Table 20: Anchor item fit to blueprint – To February 2020

Area*	Actual	Target
10	5%	4%
20	12%	11%
30	12%	13%
40	10%	13%
50	12%	11%
60	12%	13%
70	14%	13%
80	10%	11%
90	14%	11%

*See Table 13 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.

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 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 candidates taking the August 2020 CKE 1 scored negligibly lower than the candidates taking the February 2020 CKE 1 (71.1% vs. 71.6%; $t(383)=0.38$, *ns*). Because the August 2020 CKE 1 candidates scored negligibly lower (non-significance notwithstanding), they would likely have a modestly lower pass rate as compared to February 2020 candidates.

The equating analysis bears this out (Table 22). All methods indicate a pass mark of 97. The pass rate based on this equating run is lower, as expected, than what was seen in February 2020. The Tucker equating value of 96.63 was extracted from this analysis for use in setting the final pass mark.

Table 21: Equating parameter table – Total pass mark, to February 2020

		Feb. 2020	Aug. 2020
	N	144	241
	Scored items	153	150
Mean score	Total	71.3%	71.1%
	Anchors	71.6%	71.1%

Table 22: Equating outcome table – Total pass mark, to February 2020

Method	Pass Mark		Pass Rate	
	Precise	Integer	All	First-time
Equating Feb. 2020	98.46	99	65.3%	76.4%
Tucker	96.63	97	67.2%	74.3%
Levine observed	96.17	97	67.2%	74.3%
Mean	96.53	97	67.2%	74.3%
Circle Arc 1	96.87	97	67.2%	74.3%
Circle Arc 2	96.87	97	67.2%	74.3%

Combined Results

Table 23 shows the pass mark values across the 3 equating runs. The value highlighted in green is the one that would be selected based on sample parameters at each equating run. The weighted mean (by number of anchor items and number of candidates) of the 3 identified values was the preliminary pass mark for the August 2020 CKE 1 (96.024).

Using the established convention for this testing program, the mean combined value would be rounded up to 97. However, two factors suggested that 96 would be a more equitable pass mark: 1) The June 2019 equating procedure could have yielded a lower pass mark if the Levine Observed method was used (which was tenable given the change in candidate performance); 2) the pattern of pass rates and expected changes from past administrations given the performance of August 2020 candidates was more aligned to a pass mark of 96 than 97. Further, the weighted or arithmetic mean of all methods except Tucker was a value between 95 and 96 (yielding an integer pass mark of 97). Given the margin of error around any equating analysis, 96 was considered a defensible alternative to 96.02 in this instance.

With a pass mark of 96, the pass rate for first-time August 2020 candidates was 75.9%, lower than in February 2020 (as expected based on anchor set performance), and higher than in June and October 2019 (as expected based on anchor set performance). With a pass mark of 97 (which is the result of the weighted average of 96.02 is adopted), the pass rate for first-time August candidates was 74.3%, lower than in February 2020 (as expected), the same as in October 2019 (contrary to expectation given anchor performance), and higher than in June 2019 (as expected).

The two final pass mark values, and the processes used to derive them, were presented to the CHRP EVC (Table 25) via teleconference on September 21, 2020. The EVC was given the choice between both defensible outcomes, and discussed the merits of both approaches from the perspective of providing the most equitable pass mark. After discussion, an initial vote was recorded as 4–1 in favour of 96.00. Though consensus is not necessary, the EVC was provided more time to discuss varying views before taking a final vote. After this discussion, the committee voted unanimously to support 96.00 as the pass mark for the August 2020 CKE 1. The HRP A Registrar joined the panel at the end of the call and accepted the committee's recommendation; the pass mark was formally established.

Table 23: Equating outcome table – Combined results

	Jun. 19	Oct. 19	Feb. 20
Tucker	95.4	96.0	96.6
Levine observed	93.9	95.1	96.2
Mean	94.9	95.8	96.5
Circle arc 1	94.5	95.8	96.9
Circle arc 2	94.5	95.8	96.9

Table 24: Historical pass rates

	All	1st time
Feb. 17	50.5%	62.5%
Jun. 17	67.8%	75.5%
Oct. 17	59.2%	66.5%
Feb. 18	64.2%	70.4%
Jun. 18	58.6%	66.2%
Oct. 18	67.0%	75.8%
Feb. 19	61.9%	72.5%
Jun. 19	56.6%	65.6%
Oct. 19	66.2%	74.3%
Feb. 20	65.3%	76.4%
Aug. 20	70.1%	75.9%

Table 25: CHRP Examination Validation Committee members – Pass mark approval

Member	Credential	Years of Relevant Experience	Joined EVC	Industry
Sunday Ajao	CHRL	15–20	2017	Banking/finance
Roxanne Chartrand	CHRL	20–29	2018	Insurance
✓ Claire Chester	CHRL	10–15	2017	Health services
✓ Tanya Gopaul	CHRL	10–15	2017	Banking
Jean Lazarus	CHRL	15–19	2017	Health services
Suman Seth	CHRL	15–19	2018	Government
✓ Kriss Stone	CHRL	10–15	2017	Real estate
✓ Ileean Tait	CHRL	15–20	2017	Environmental
✓ Patricia Verkley	CHRL	10–15	2019	Not-for-profit
Karen Weiler	CHRL	20–29	2017	Software/ communications

✓ 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 26 provides the means and standard deviations for the functional areas and for the total score, using all candidates who took the new August 2020 CKE 1 forms. Table 27 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 26: Total and functional area scores for all candidates

Functional Area	Percentage	Mean	SD*
10 Strategy	75%	4.5	1.1
20 Professional Practice	70%	11.2	2.4
30 Organizational Effectiveness	69%	13.1	2.8
40 Workforce Planning & Talent Management	72%	13.8	2.7
50 Labour & Employee Relations	69%	11.8	2.3
60 Total Rewards	69%	13.0	2.8
70 Learning & Development	65%	12.4	3.2
80 Health, Wellness & Safe Workplace	69%	12.5	2.3
90 HR Metrics, Reporting & Financial Management	67%	11.4	2.4
Total score	69.1%	103.6	15.6

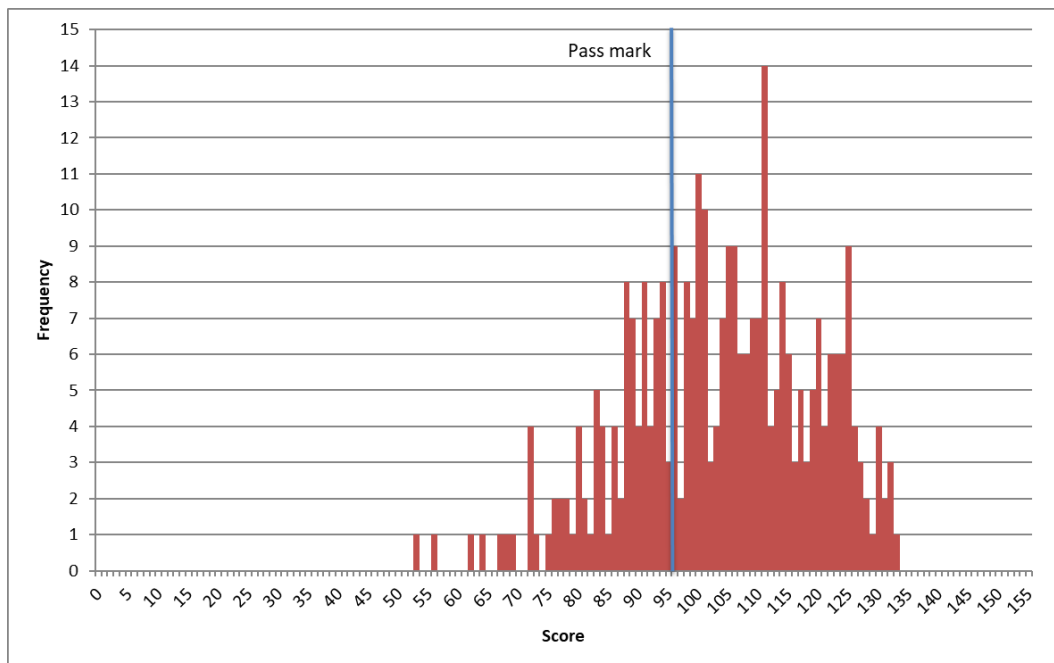
*SD = standard deviation.

Table 27: Correlations between functional area scores for all candidates

Area*	10	20	30	40	50	60	70	80	90
10		.23	.30	.32	.30	.38	.40	.22	.35
20			.44	.45	.36	.48	.45	.37	.35
30				.54	.42	.53	.60	.37	.46
40					.45	.56	.56	.37	.49
50						.41	.51	.37	.39
60							.55	.35	.48
70								.39	.55
80									.34
90									

*See Table 26 for the full name of each functional area.

Figure 3: Score distribution for all candidates



Key Examination Metrics

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

Index	August 2020	February 2020	October 2019	June 2019	February 2019
Scored items	150	153	153	150	155
Candidates	241	144	175	128	142
Mean	106.6 (71.1%)	109.1 (71.3%)	106.4 (69.5%)	101.1 (67.4%)	112.1 (72.3%)
Median	108 (72.0%)	110 (71.9%)	107 (69.9%)	100.5 (67.0%)	114.5 (73.9%)
Skewness	-0.402	-0.525	-0.466	0.002	-0.876
Kurtosis ⁱ	-0.498	-0.379	-0.057	-0.446	0.673
Range	67–133 (44.7– 88.7%)	66–137 (43.1– 89.5%)	61–138 (39.9– 90.2%)	60–138 (40.0– 92.0%)	55–142 (35.5– 91.6%)
Standard deviation	14.88	16.59	15.76	16.43	18.45
Cronbach's alpha	.88	.90	.89	.90	.92
Mean r_{pb}^*	.20	.23	.22	.22	.26
SEM ⁱⁱ	5.17	5.17	5.21	5.23	5.10
SEM at the pass mark	5.61	5.64	5.59	5.57	5.61
Decision consistency (uncorrected) ⁱⁱⁱ	.89	.90	.88	.86	.88
Perceived fairness ^{iv}	45%	49%	53%	42%	47%
Pass mark	96.000	98.458	97.499	93.247	102.054
Effective pass mark	96	99	98	94	103
Pass rate	75.9%	76.4%	74.3%	65.6%	72.5%

ⁱExcess

ⁱⁱSEM = standard error of measurement.

ⁱⁱⁱSubkoviac 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.

Related Development Activities

Since the last administration of the CKE 1 in February 2020, the following exam development activities have taken place.

Item Writing

To fill gaps in the bank and renew content, item writing was conducted in March–May 2020. Item writers (see Table 29) were identified by HRP A and trained in a remote session by Wickett on March 13, 2020 (repeated on March 16 for those unavailable).

Table 29: 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	11	Health care, Education
Enda Soostar	CHRL, LL.M.	Barrie	30	Education
Nicole Vincic	CHRL, BA, MA	Hamilton	22	Education

The item writers were provided with training via teleconference, and received printable files covering the main elements of the training. The general guidance for writing quality multiple choice items was drawn primarily from Haladyna & Rodriguez (2013).¹¹

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 each to write, for a total of 100 items.

The item writers had access to the style guide that governs language usage on the HRP A 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 review, validation and upon use with candidates. This generally required several iterations per item.

¹¹ Haladyna, T. M., & Rodriguez, M.C. (2013). *Developing and validating test items*. New York, NY: Routledge.

Once all items were drafted and declared complete, they were sent 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 provide sufficient items for upcoming administrations, a validation session was held with the EVC (see remotely on April 22, 23, 27 and 29, 2020. Though originally planned to be an in person session, it was converted to remote because of COVID-19. During these 4 sessions, CHRP ELE items were also validated, and that portion of the validation activity is reported in the technical report for that examination.

Note that scheduling precluded all identified members from being available for all 4 days; those marked as having participated attended the validation activity on at least 1 day.

For clarity, the items newly written in 2020 were not available for this validation session as they must first go through review.

Table 30) remotely on April 22, 23, 27 and 29, 2020. Though originally planned to be an in person session, it was converted to remote because of COVID-19. During these 4 sessions, CHRP ELE items were also validated, and that portion of the validation activity is reported in the technical report for that examination.

Note that scheduling precluded all identified members from being available for all 4 days; those marked as having participated attended the validation activity on at least 1 day.

For clarity, the items newly written in 2020 were not available for this validation session as they must first go through review.

Table 30: CHRP Examination Validation Committee members – Validation

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

✓ Participated in the session.

The EVC members received advance materials outlining:

- Purpose of the session
- Description of the CHRP credential
- CKE 1 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, the received individual training on the first day of their involvement. Each day, committee members were provided with approximately 50–70 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 CHRP 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 2 years to make decisions about who would be certified as a CHRP
- Move the item to the CKE 2 or CHRP 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

At the close of each day, committee members were walked through a process to verify deletion of all item files in use that day.

The committee validated 51 items as suitable for the CKE 1 and rejected 2 items. Three 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 Review

Following the item writing exercise in March–May 2020 there was need for group review of those items before moving them to formal validation and use on the CKE 1. The group had 103 items for consideration (taken from newly written items, supplemented with 3 other unreviewed items in the bank).

The session was held with a panel of HR professionals (see Table 31) remotely on June 16, 17, 24 and 25, 2020. Though originally planned to be an in person session, it was converted to remote because of COVID-19. Note that scheduling precluded all identified members from being available for all 4 days.

Table 31: Panel for item review session

Reviewer	Credentials	Years of Relevant Experience	Industry
Olivia Ariss	CHRL	8	Energy
Amuna Baraka-Clarke	CHRL	19	Not-for-profit
Brittany Browney	CHRL	10	Public sector
Ramon Calanza	CHRL	15	Transportation
Kelly Cotton	CHRL	15+	Construction
Susan Ehlert	CHRP, CHRL	7	Education
Sue Haywood	CHRL	20	Professional services
Lola Obomighie	CHRL	13	Education
Oshin Soni	CHRL	10	Mining
Michelle Sultan	CHRL	15	Education

The panel members received training on the review activity on the first day of the session. Each day, committee members were provided with approximately 25 items via a secure file share site, and then worked individually reviewing and revising 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 with an appropriate key and plausible distractors.

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

At the close of each day, panel members were walked through a process to verify deletion of all item files in use that day.

The committee approved 90 items as effective CKE 1 items, rejected 12 items as unfixable and moved 1 item to the Employment Law Examination bank. Fifty-seven of the approved items were revised before approval, and the remaining 33 were considered sound without revision. The committee also verified the functional area and competency for all items.

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

Appendix A

Blueprint

Comprehensive Knowledge Examination 1

Human Resources Professionals Association
Version 2.2

Approved by CHRP Exam Validation Committee April 9, 2018

Approved by HRPA Registrar April 11, 2018

Effective June 2018

Credential

Passing the Comprehensive Knowledge Examination 1 is a requirement for certification for CHRP candidates. The examination reflects the *HRPA Professional HR Competency Framework* (2014).

Purpose

The CKE 1 assesses whether a candidate has the level of discipline-specific knowledge necessary to practise human resources management at the CHRP 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 CHRP Employment Law Examination.

Structure

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

Table 32: CKE 1 Blueprint structural variables

Item types	Independent 4-option multiple choice
Length	175 items in total
	20–30 experimental items
Duration	Up to 3½ 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 to reflect an equal importance across the functional areas, except with a lower expectation for Strategy. The weights were modified slightly in 2018 to remove weighting for competencies most appropriately tested on the CHRP

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

Table 33: Functional area weights on the CKE 1

Functional Area		Weight	Range
10	Strategy	4%	+/- 1%
20	Professional Practice	11%	+/- 2%
30	Organizational Effectiveness	13%	+/- 2%
40	Workforce Planning & Talent Management	13%	+/- 2%
50	Labour & Employee Relations	11%	+/- 2%
60	Total Rewards	13%	+/- 2%
70	Learning & Development	13%	+/- 2%
80	Health, Wellness & Safe Workplace	11%	+/- 2%
90	Human Resources Metrics, Reporting & Financial Management	11%	+/- 2%

Table 34: Competencies not eligible on the CKE 1

FA	Comp	FA	Comp	FA	Comp	FA	Comp
10	C005	40	C084	70	C152	80	C177
	C007		C089		C155		C179
	C009	50	C113		C156		C187
	C011		C114		C158		C192
	C012		C117		C159	90	C194
	C017		C123		C163		C195
20	C035	60	C125		C165		C196
	C036		C139		C166		C204
	C037	C141	C171		C205		
	C041	C143	C172		C206		
30	C050	C146	C173	C210			
	C056		C175				
	C057						
	C065						

Minor amendments made November 20, 2018, by CHRP 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.