

Technical Report: October 2019 CKE 1

Human Resources Professionals Association

9 November 2019

Revised 17 December 2019



Contents

Executive Summary	4
Administration	5
Form Setting	5
Testing Window	6
Analysis	7
Data Cleaning and Integrity Checks.....	7
Post-Examination Survey.....	9
Initial Analysis	11
Key Validation.....	12
Establishing the Pass Mark: Equating.....	13
Scoring	22
Key Examination Metrics	24
Related Development Activities	25
Item Writing	25
Item Review.....	26
Appendix A	27
Appendix B	29

List of Tables

Table 1: Test forms as administered	6
Table 2: Administration-related post-examination survey questions*	10
Table 3: Content-related post-examination survey questions*	10
Table 4: Preference regarding computer-based testing versus pencil-and-paper	11
Table 5: Initial examination statistics – Combined across forms	11
Table 6: CHRP Examination Validation Committee members – Key validation	12
Table 7: Final scored examination fit to blueprint	13
Table 8: Anchor item fit to blueprint – To June 2019	15
Table 9: Equating parameter table – Total pass mark, to June 2019	15
Table 10: Equating outcome table – Total pass mark, to June 2019	16
Table 11: Anchor item fit to blueprint – To October 2018	17
Table 12: Equating parameter table – Total pass mark, to October 2018	17
Table 13: Equating outcome table – Total pass mark, to October 2018	18
Table 14: Anchor item fit to blueprint – To February 2019	19
Table 15: Equating parameter table – Total pass mark, to February 2019	19
Table 16: Equating outcome table – Total pass mark, to February 2019	20
Table 17: Equating outcome table – Combined results	21
Table 18: Historical pass rates	21
Table 19: CHRP Examination Validation Committee members – Pass mark approval	21
Table 20: Total and functional area scores for all candidates	22
Table 21: Correlations between functional area scores for all candidates	23
Table 22: Key examination metrics – Candidates included in analysis only	24
Table 23: Item writers	25
Table 24: Panel for item review session	26
Table 25: CKE 1 Blueprint structural variables	27
Table 26: Functional area weights on the CKE 1	28
Table 27: Competencies not eligible on the CKE 1	28

List of Figures

Figure 1: Examination time distribution for all candidates	8
Figure 2: Candidate volume and score trends across testing window	8
Figure 3: Score distribution for all candidates	23

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 Comprehensive Knowledge Exam 1 (CKE 1) was administered to 231 candidates using computer-based testing at Prometric test centres October 14–28, 2019, 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 candidates² ($n=175$) was 106.4 (69.5%), and for all candidates it was 103.2 (67.5%), out of 153 scored items. Reliability was strong at .89. The final set of scored items adhered to the blueprint parameters.

The pass mark was set using equating back to the October 2018, February 2019, and June 2019 administrations, yielding an integer pass mark of 98. 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 74.3% and a pass rate for all candidates of 66.2%.

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

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

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

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

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

Administration

Form Setting

Using only validated test items, Wickett Measurement Systems prepared 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 October 2019 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).

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	6–7	7
20	Professional Practice	17–18	19
30	Organizational Effectiveness	22–23	23
40	Workforce Planning & Talent Management	23–24	23
50	Labour & Employee Relations	19–20	19
60	Total Rewards	23–24	23
70	Learning & Development	22–23	23
80	Health, Wellness & Safe Workplace	20	19
90	HR Metrics, Reporting & Financial Management	19–20	19
	TOTAL	175	175

Testing Window

The examination was administered via computer-based testing at Prometric test sites primarily in Ontario. The testing window was October 14–28, 2019, inclusive, and 231 candidates wrote the exam.

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

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.

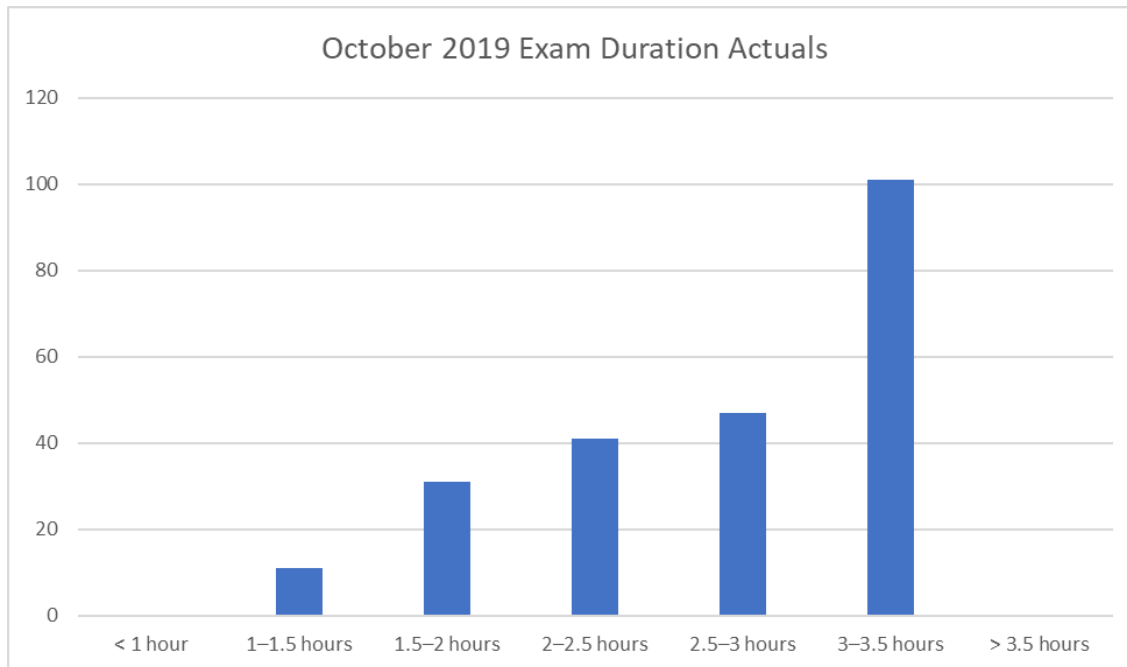
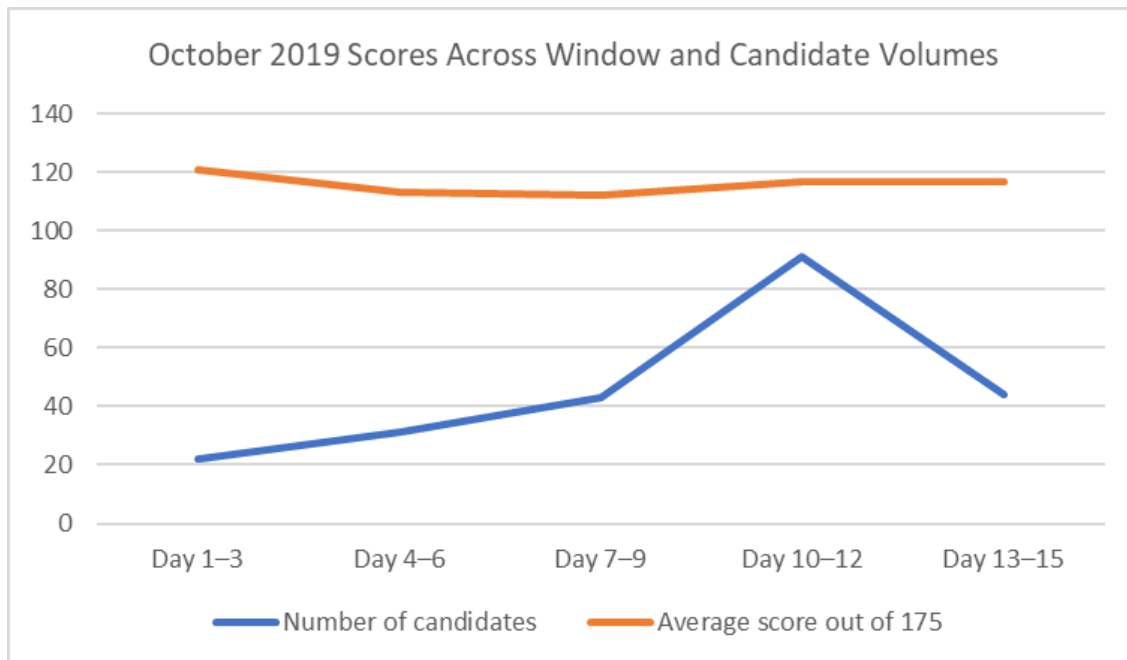
The average time taken by all candidates was assessed to detect potential examination timing concerns. The distribution is shown in Figure 1. The mean was 2 hours, 44 minutes (11 minutes more than in June 2019; on average, form A required 2 hours, 42 minutes, form B required 2 hours, 45 minutes, and form C required 2 hours, 43 minutes). The time limit on the CKE 1 was 3½ hours, suggesting that time was not a factor in scores across candidates.

Twenty-seven candidates (12%) took the full 3½ hours, suggesting that those candidates may have wanted more time (or that they opted to stay the entire sitting verifying answers), and 1 candidate (0.4%) left at least 1 item blank, suggesting that that candidate 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 modest at a value of .31 for form A, small at a value of -.14 for form B, and modest at a value of -.28 for form C, suggesting that time constraints were not generally an issue for candidate performance and not in a consistent direction.

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 3 days and those writing in the last 3 days was a decrease of 4.1 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.

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. One candidate was 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. 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, 3 candidates were removed from analysis.

Candidates who had failed a previous HRPA examination (CKE, CKE 1, or CKE 2) scored lower than did those who had not (60.4% and 67.9%, respectively, on the full exam of 175 items). This difference was meaningful and significant ($t(104)=5.25, p<.001$). In keeping with standard procedures, these candidates were removed from subsequent analyses. The CKE 1 analysis proceeded with 175 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; 228 responses were obtained from candidates (response rate, 99%).

Table 2 shows the responses to the administration-related questions. Note that candidates were generally very positive about the administration experience. Table 3 shows the content-related questions; there was a tendency to more neutrality on these questions. The rating for perceived fairness (Question 14) warrants monitoring as it continues to be low.

Candidates were asked to express their opinion regarding whether completing the examination on a computer affected their performance. Table 4 shows that over half of candidates felt it made no difference, and where a preference was expressed it was for using a computer.

An open-ended question was also posed to candidates asking for any additional comments. Those comments were provided to HRPA for information and consideration. Nothing in the comments or survey data raised concerns about item analysis or scoring.

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

	Question	SA	A	N	D	SD	Score	Agreement
1.	I was able to book a seat to write the examination at a time that was convenient for me.	108	91	8	19	2	4.3	87%
2.	I was well informed about what documents to bring to the exam location.	160	64	3	1	0	4.7	98%
3.	Proctors enforced the exam-day rules and the security procedures at the test centre were what I expected.	158	60	8	2	0	4.7	96%
4.	Proctors were professional and courteous.	168	52	4	4	0	4.7	96%
5.	The tutorial helped me understand how to complete the examination on the computer.	155	66	6	1	0	4.7	97%
6.	Navigation through the examination was easy and intuitive.	155	68	4	0	1	4.7	98%

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

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

	Question	SA	A	N	D	SD	Score	Agreement
7.	The time allotted for this examination was sufficient.	135	71	12	8	2	4.5	90%
8.	Information available prior to exam day provided me with adequate details about the content and format of the exam.	62	94	41	22	9	4.0	68%
9.	I feel I was adequately prepared to write this examination.	32	92	65	36	3	3.8	54%
10.	The questions in the examination were clearly written.	34	114	46	30	4	3.8	65%
11.	The terminology used in the examination was accurate.	26	135	49	17	0	4.0	71%
12.	The situations presented in the examination were realistic.	44	133	41	9	1	4.1	78%
13.	The questions in the examination reflected the examination blueprint.	26	110	72	19	0	3.9	60%
14.	The examination was a fair assessment of my ability.	24	94	60	43	7	3.6	52%

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

Table 4: Preference regarding computer-based testing versus pencil-and-paper

Question	Count	%
I feel that completing the examination on a computer improved my performance.	71	31%
I feel that completing the examination on a computer decreased my performance.	32	14%
I feel that completing the examination on a computer had no effect on my performance.	125	55%

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, 154 items were available for scoring on each, after removing items designated as experimental.

The initial analysis summary statistics are presented in Table 5.

Table 5: Initial examination statistics – Combined across forms

Index	Value
Items	154
Total candidates	231
Candidates in analysis	175
Mean score	107.3 (69.6%)
Score range	62–139 (40.3–90.3%)
Cronbach's alpha	.89
Mean r_{pb}^*	.21

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, though difficulty was also factored in to avoid 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 November 4, 2019, using members of the CHRP Examination Validation Committee (EVC). The EVC (Table 6) 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 6: 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
Alyssa Young	CHRL	5–9	2017	Not-for-profit

✓ Participated in the session.

The group was informed that test reliability, as measured by Cronbach's alpha, was .89 based on the set of 154 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. 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 October 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 153 scored items was .89. The difficulties ranged from 33.7% to 95.4%, with a mean of 69.5%. The r_{pb}^* values ranged from $-.06$ to $.50$, with a mean of $.22$.

Table 7 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 October 2019 CKE 1 candidates who took this form.

Table 7: 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	17	14	17	19	11% ± 2%
30 Organizational Effectiveness	20	17	20	22	13% ± 2%
40 Workforce Planning & Talent Management	20	17	20	22	13% ± 2%
50 Labour & Employee Relations	17	14	17	19	11% ± 2%
60 Total Rewards	20	17	20	22	13% ± 2%
70 Learning & Development	20	17	20	22	13% ± 2%
80 Health, Wellness & Safe Workplace	16	14	17	19	11% ± 2%
90 HR Metrics, Reporting & Financial Management	17	14	17	19	11% ± 2%
Total	153				

*Adds to 154 due to rounding.

Establishing the Pass Mark: Equating

Equating, as per Kolen and Brennan (2014),⁵ was used to establish the pass mark for the October 2019 CKE 1. The goal of this process was to set a pass mark for the October 2019 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

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

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 (October 2018, February 2019, and June 2019). The third procedure (to February 2019) was conducted because of discrepancies between the 2 planned procedures. The intention following these 3 equating runs was to average them to arrive at a final pass mark for the October 2019 CKE 1. The 2 planned administrations were chosen because they were the most recent administration and the administration corresponding to the same administration month the previous year.

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 arc⁶ 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 0.69 and a mean corrected point-biserial of .25 (for October 2019 candidates).

Table 8 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.

⁶ 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 8: Anchor item fit to blueprint – To June 2019

Area*	Actual	Target
10	2%	4%
20	11%	11%
30	13%	13%
40	15%	13%
50	11%	11%
60	13%	13%
70	13%	13%
80	11%	11%
90	13%	11%

*See Table 7 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 9 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 October 2019 CKE 1 scored the same as the population taking the June 2019 CKE 1 (68.9% vs. 68.8%). Because the October 2019 CKE 1 candidates scored the same (based on the anchors), they would likely have approximately the same pass rate as was seen in June.

The equating analysis bears this out, in part (Table 10). All methods indicate a pass mark of 99. The pass rate based on this equating run is higher, however, than what was seen in June 2019. This appears to be due simply to sampling fluctuation where the June sample had a small peak of candidates just below the pass mark for that test form. The Tucker equating value of 98.92 was extracted from this analysis for use in setting the final pass mark.

Table 9: Equating parameter table – Total pass mark, to June 2019

		June 2019	Oct. 2019
N		128	175
Scored items		150	153
Mean score	Total	67.4%	69.5%
	Anchors	68.8%	68.9%

Table 10: 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	98.92	99	64.5%	72.0%
Levine observed	98.97	99	64.5%	72.0%
Mean	98.42	99	64.5%	72.0%
Circle Arc 1	98.29	99	64.5%	72.0%
Circle Arc 2	98.26	99	64.5%	72.0%

Equating Back to the October 2018 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 arc⁷ 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 0.70 and a mean corrected point-biserial of .23 (for October 2019 candidates).

Table 11 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.

⁷ 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 11: Anchor item fit to blueprint – To October 2018

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

*See Table 7 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 12 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 October 2019 CKE 1 scored the same as the population taking the October 2018 CKE 1 (70.3% vs. 70.5%). Because the October 2019 CKE 1 candidates scored the same (based on the anchors), they would likely have approximately the same pass rate as was seen in October 2018.

The equating analysis bears this out (Table 13). All methods indicate a pass mark of 97 or 98. The pass rate based on this equating run is essentially the same, as expected, as what was seen in October 2018. The Tucker equating value of 96.83 was extracted from this analysis for use in setting the final pass mark.

Table 12: Equating parameter table – Total pass mark, to October 2018

		Oct. 2018	Oct. 2019
N		178	175
Scored items		155	153
Mean score	Total	69.1%	69.5%
	Anchors	70.5%	70.3%

Table 13: Equating outcome table – Total pass mark, to October 2018

Method	Pass Mark		Pass Rate	
	Precise	Integer	All	First-time
Equating Oct. 2018	97.39	98	67.0%	75.8%
Tucker	96.83	97	68.0%	76.0%
Levine observed	96.98	97	68.0%	76.0%
Mean	96.93	97	68.0%	76.0%
Circle Arc 1	97.08	98	66.2%	74.3%
Circle Arc 2	97.08	98	66.2%	74.3%

Equating Back to the February 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 .25 (for October 2019 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.

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.

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

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

*See Table 7 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 15 shows some of the parameters used to derive the equating estimates, along with other parameters describing the test forms. Of note is that on the anchor items, the population taking the October 2019 CKE 1 scored the same as the population taking the February 2019 CKE 1 (70.8% vs. 70.9%). Because the October 2019 CKE 1 candidates scored the same (based on the anchors), they would likely have approximately the same pass rate as was seen in February.

The equating analysis bears this out, in part (Table 16). All methods indicate a pass mark of 97. The pass rate based on this equating run is modestly higher, however, than what was seen in February 2019. This appears to be due simply to sampling fluctuation where the February sample had a cluster of candidates right around the pass mark that test form. The Tucker equating value of 96.95 was extracted from this analysis for use in setting the final pass mark.

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

		Feb. 2019	Oct. 2019
N		142	175
Scored items		155	153
Mean score	Total	72.3%	69.5%
	Anchors	70.9%	70.8%

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

Method	Pass Mark		Pass Rate	
	Precise	Integer	All	First-time
Equating Feb. 2019	102.05	103	61.9%	72.5%
Tucker	96.95	97	68.0%	76.0%
Levine observed	96.68	97	68.0%	76.0%
Mean	96.44	97	68.0%	76.0%
Circle Arc 1	96.33	97	68.0%	76.0%
Circle Arc 2	96.28	97	68.0%	76.0%

Combined Results

Table 17 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 recommended pass mark for the October 2019 CKE 1 (97.499).

Using the established convention for this testing program, the mean combined value was rounded up to a cut score of 98. The resulting pass rate of 74.3% for first-time candidates is close to what was seen in October 2018 and February 2019, and modestly higher than what was seen in June 2019. The pass rate for all candidates in October 2019 was 66.2%. See Table 18 for historical pass rates.

The final pass mark value, and the process used to derive it, was presented to the CHRP EVC (Table 19) via teleconference on November 8, 2019. No concerns were raised regarding the pass mark or pass rates. The panel formally approved the pass mark (which was presented along with the consequent pass rate data) for recommendation to HRP. The HRP Registrar participated on the call and accepted the panel's recommendation; the pass mark was formally established.

Table 17: Equating outcome table – Combined results

	Oct. 18	Feb. 18	Jun. 19
Tucker	96.8	96.9	98.9
Levine observed	97.0	96.7	99.0
Mean	96.9	96.4	98.4
Circle arc 1	97.1	96.3	98.3
Circle arc 2	97.1	96.3	98.3

Table 18: 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%

Table 19: 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
✓ 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	Not-for-profit

✓ 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 20 provides the means and standard deviations for the functional areas and for the total score, using all candidates who took the new October 2019 CKE 1 forms. Table 21 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 20: Total and functional area scores for all candidates

Functional Area	Percentage	Mean	SD*
10 Strategy	77%	4.6	1.1
20 Professional Practice	72%	12.2	2.4
30 Organizational Effectiveness	63%	12.6	2.9
40 Workforce Planning & Talent Management	68%	13.6	2.7
50 Labour & Employee Relations	67%	11.4	2.7
60 Total Rewards	65%	13.0	3.0
70 Learning & Development	68%	13.7	3.1
80 Health, Wellness & Safe Workplace	67%	10.8	2.2
90 HR Metrics, Reporting & Financial Management	66%	11.2	2.3
Total score	67.5%	103.2	16.7

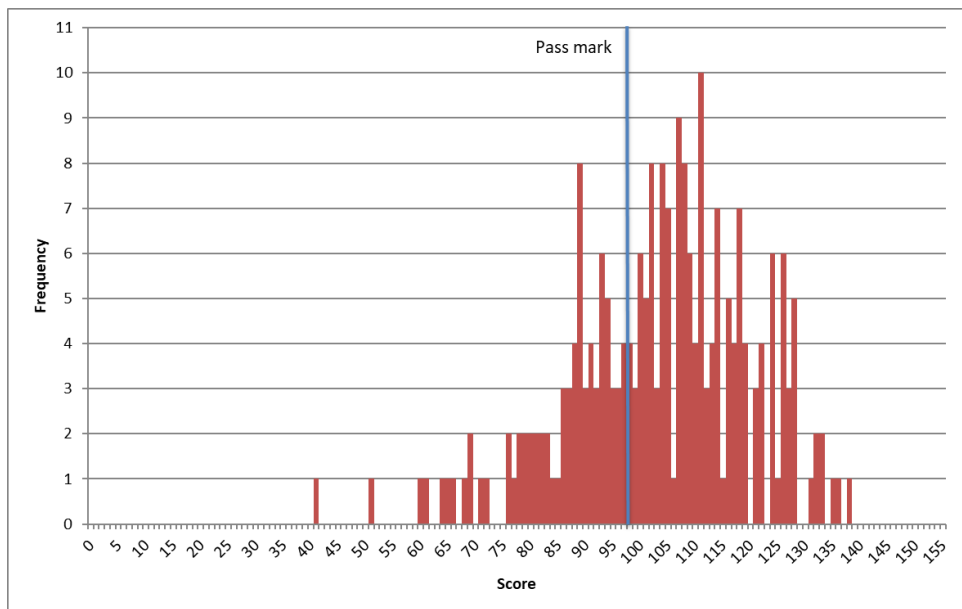
*SD = standard deviation.

Table 21: Correlations between functional area scores for all candidates

Area*	10	20	30	40	50	60	70	80	90
10		.27	.45	.43	.23	.44	.44	.40	.35
20			.50	.45	.52	.54	.44	.47	.45
30				.56	.47	.59	.66	.46	.50
40					.53	.59	.61	.49	.49
50						.55	.52	.49	.41
60							.62	.53	.50
70								.50	.44
80									.36
90									

*See Table 20 for the full name of each functional area.

Figure 3: Score distribution for all candidates



Key Examination Metrics

Table 22 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 22: Key examination metrics – Candidates included in analysis only

Index	October 2019	June 2019	February 2019	October 2018	June 2018
Scored items	153	150	155	155	150
Candidates	175	128	142	178	157
Mean	106.4 (69.5%)	101.1 (67.4%)	112.1 (72.3%)	107.2 (69.1%)	102.0 (68.0%)
Median	107 (69.9%)	100.5 (67.0%)	114.5 (73.9%)	109 (70.3%)	103 (68.7%)
Skewness	-0.466	0.002	-0.876	-0.462	-0.403
Kurtosis ⁱ	-0.057	-0.446	0.673	0.097	0.208
Range	61–138 (39.9– 90.2%)	60–138 (40.0– 92.0%)	55–142 (35.5– 91.6%)	60–141 (38.7– 91.0%)	53–136 (35.3– 90.7%)
Standard deviation	15.76	16.43	18.45	15.35	16.68
Cronbach's alpha	.89	.90	.92	.88	.90
Mean r_{pb}^*	.22	.22	.26	.20	.23
SEM ⁱⁱ	5.21	5.23	5.10	5.25	5.24
SEM at the pass mark	5.59	5.57	5.61	5.63	5.53
Decision consistency (uncorrected) ⁱⁱⁱ	.88	.86	.88	.89	.86
Perceived fairness ^{iv}	53%	42%	47%	46%	49%
Pass mark	97.499	93.247	102.054	97.387	96.622
Effective pass mark	98	94	103	98	97
Pass rate	74.3%	65.6%	72.5%	75.8%	66.2%

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.

Related Development Activities

Since the last administration of the CKE 1 in June 2019, the following exam development activities have taken place.

Item Writing

To fill gaps in the bank and renew content, item writing was conducted in June–August 2019. Item writers (see Table 23) were identified by HRP A and trained in a remote session by Wickett on June 28, 2019.

Table 23: Item writers

Writer	Credentials	Years of Relevant Experience	Industry
Nicole Bonenfant	CHRL	10 years working in HR	Healthcare, Education
David Dorward	MIRHR	40+ years working in HR	Utilities, Health, Finance
Carol Ann Samhaber	CHRL, EdD	20+ years teaching in HR	Professor at Algonquin College
Nicole Vincic	CHRL, M.A. Ed	20 years of working and teaching HR	Education; Professor at Mohawk College

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).⁹

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 15–30 items each to write, for a total of 105 items (one writer did not fully complete their items, and so 94 new items were obtained).

The item writers had access to the style guide that governs language usage on the HRP A exams and were provided with recent 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.

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.

⁹ Haladyna, T. M., & Rodriguez, M.C. (2013). *Developing and validating test items*. New York, NY: Routledge.

Item Review

Following the item writing exercise in June–August 2019 there was need for group review of those items before moving them to formal validation and use on the CKE 1. The group had 211 items for consideration (taken from newly written items, supplemented with other unreviewed items in the bank and items requiring revision required to fill gaps in the bank). A further set of 42 items with missing information was also available to the group for when time was available. The group was not expected to be able to review all items in the time available.

The 2½-day review session was held October 1–3, 2019 at HRPAs offices. The panel members who participated are shown in Table 24.

Table 24: Panel for item review session

Reviewer	Credentials	Years of Relevant Experience	Industry
Zaheer-Ud-Din Babar	CHRL	10	Consulting
Valentin Bachner	CHRL	21	Federal Government/ Consultant
Emily MacRobbie	CHRL	8	Consulting
Sarah Morgan	CHRL	7	Consulting
Anusha Neelakantan	CHRL	13	Manufacturing
Michelle Sultan	CHRL	10	Security
Lin Tian	CHRL, CPM	24	Consulting
Marisa Wallis	CHRL	27	Insurance

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

The panel members reviewed and approved 116 items as suitable for the CKE 1, moved 1 item to the CKE 2 bank, moved 3 items to the ELE bank, and rejected 12 items. Of the approved CKE 1 items, 63 saw text changes to the stem and/or options before approval. The group also verified the functional area, competency and references throughout the process.

The group reported great confidence in the final assessments made by the group.

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 25: 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 26: 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 27: 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.