

Technical Report: February 2019 CKE 1

Human Resources Professionals Association

12 March 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	20
Key Examination Metrics	23
Related Development Activities	24
Item Validation.....	24
CKE 1 Blueprint Revision.....	25
Appendix A	26
Appendix B	29

List of Tables

Table 1: Test form 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	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 October 2018	15
Table 9: Equating parameter table – To October 2018.....	15
Table 10: Equating outcome table – To October 2018	16
Table 11: Anchor item fit to blueprint – To February 2018.....	17
Table 12: Equating parameter table – To February 2018	17
Table 13: Equating outcome table – To February 2018.....	18
Table 14: Equating outcome table – Combined results	19
Table 15: Historical pass rates	19
Table 16: CHRP Examination Validation Committee members – Pass mark approval	20
Table 17: Total and functional area scores for all candidates	21
Table 18: Correlations between functional area scores for all candidates	21
Table 19: Key examination metrics – Candidates included in analysis only.....	23
Table 20: CHRP Examination Validation Committee – Item validation	24
Table 21: CKE 1 Blueprint structural variables	26
Table 22: Functional area weights on the CKE 1.....	27
Table 23: 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.....	22

Executive Summary¹

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

The Comprehensive Knowledge Exam 1 (CKE 1) was administered to 194 candidates using computer-based testing at Prometric test centres February 11–25, 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 candidates ($n=142^2$) was 112.1 (72.3%), and for all candidates it was 107.7 (69.5%), out of 155 scored items. Reliability was strong at .92. The final set of scored items adhered to the blueprint parameters.

The pass mark was set using equating back to the February 2018 and October 2018 administrations, yielding an integer pass mark of 103. 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 72.5% and a pass rate for all candidates of 61.9%.

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 one 175-item test form (using a combination of scored and experimental test items). Wickett constructed the final test form 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 form for text errors and detection of potential enemy items. Items flagged as enemies were replaced.

The final form composition for the primary February 2019 CKE 1 form is shown in Table 1. All functional areas are within the limits of their targets, and as such, the form reflects the blueprint (see Appendix A for the CKE 1 blueprint).

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

Table 1: Test form as administered

	Functional Area	Actual Items	Target	Variance
10	Strategy	7	7	—
20	Professional Practice	19	19	—
30	Organizational Effectiveness	23	23	—
40	Workforce Planning & Talent Management	23	23	—
50	Labour & Employee Relations	19	19	—
60	Total Rewards	23	23	—
70	Learning & Development	23	23	—
80	Health, Wellness & Safe Workplace	19	19	—
90	HR Metrics, Reporting & Financial Management	19	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 February 11–25, 2019, inclusive, and 194 candidates wrote the exam. Due to an issue at administration, one candidate wrote on February 28.

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, 39 minutes (5 minutes less than in October 2018). The time limit on the CKE 1 was 3½ hours, suggesting that time was not a factor in scores across candidates.

Eleven candidates (6%) took the full 3½ hours, suggesting that those candidates may have wanted more time, and 2 candidates (1%) 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 negligible at a value of $-.02$, suggesting that time constraints were not generally an issue for candidate performance. (Note that 1 candidate exceeded the time limit; this candidate was granted additional time in advance of the administration as an accommodation.)

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, and the difference between the first 3 days and the last 3 days was 0.0 (zero) 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 at the end of the window rather than the start.

Figure 1: Examination time distribution for all candidates

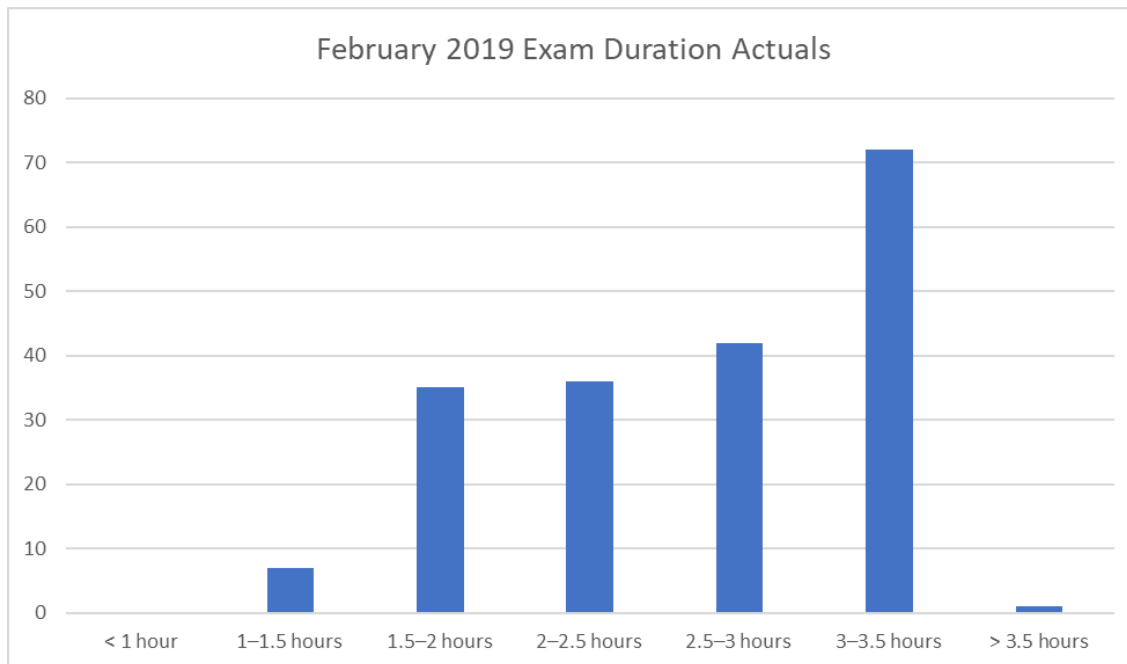
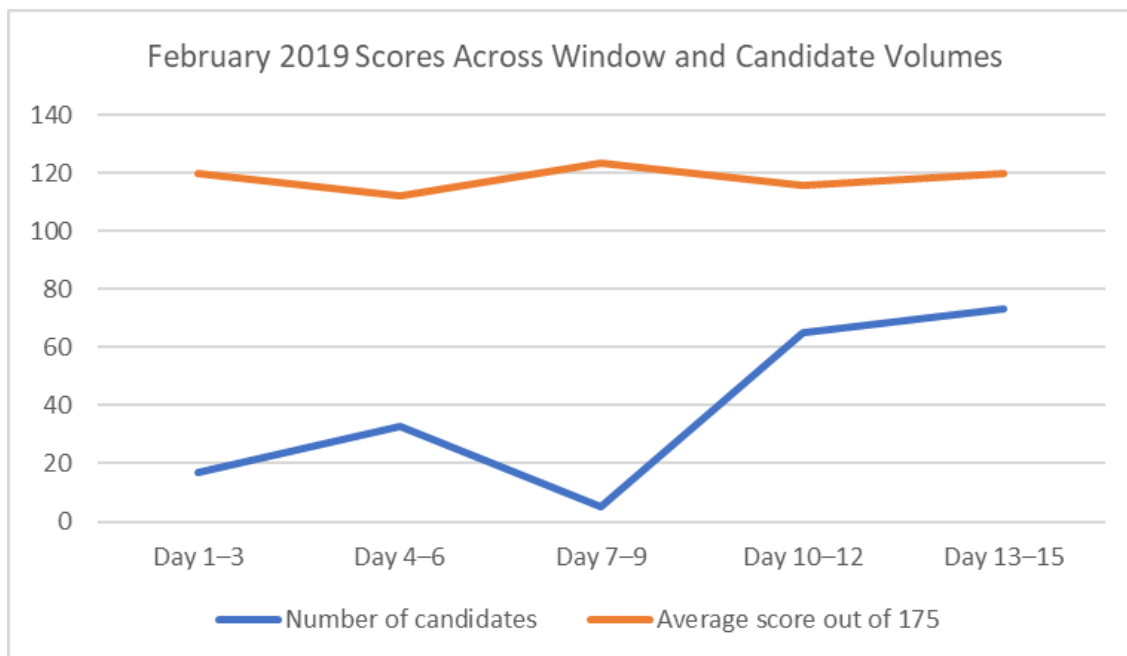


Figure 2: Candidate volume and score trends across testing window



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. 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. 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, 5 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 69.2%, respectively, on the full exam of 175 items). This difference was meaningful and significant ($t(137)=6.33$, $p<.001$). In keeping with standard procedures, these candidates were removed from subsequent analyses. The CKE 1 analysis proceeded with 142 candidates.

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

Post-Examination Survey

Candidates were provided access to the post-examination survey immediately after submitting their responses to the CKE 1; 187 responses were obtained from candidates (response rate, 96%).

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 about half of candidates felt it made no difference, and that where a preference was expressed it was modestly in favour of using a computer.

An open-ended question was also posed to candidates asking for any additional comments. Those comments were provided to HRP A for information and consideration. Nothing 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.	89	68	11	16	3	4.3	84%
2.	I was well informed about what documents to bring to the exam location.	125	56	1	3	1	4.6	97%
3.	Proctors enforced the exam-day rules and the security procedures at the test centre were what I expected.	113	58	8	3	0	4.6	94%
4.	Proctors were professional and courteous.	131	44	5	1	1	4.7	96%
5.	The tutorial helped me understand how to complete the examination on the computer.	115	57	4	2	0	4.6	97%
6.	Navigation through the examination was easy and intuitive.	120	54	4	1	1	4.6	97%

*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.	101	60	13	4	2	4.5	89%
8.	Information available prior to exam day provided me with adequate details about the content and format of the exam.	47	67	32	21	12	3.8	64%
9.	I feel I was adequately prepared to write this examination.	23	72	54	21	8	3.8	53%
10.	The questions in the examination were clearly written.	27	79	43	26	4	3.8	59%
11.	The terminology used in the examination was accurate.	25	101	35	16	2	3.9	70%
12.	The situations presented in the examination were realistic.	39	104	30	5	0	4.2	80%
13.	The questions in the examination reflected the examination blueprint.	25	69	56	21	5	3.8	53%
14.	The examination was a fair assessment of my ability.	19	65	54	30	10	3.6	47%

*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.	46	26%
I feel that completing the examination on a computer decreased my performance.	40	22%
I feel that completing the examination on a computer had no effect on my performance.	92	52%

Initial Analysis

The full CKE 1 examination was 175 items, of which approximately 150 were to be scored. The other 20–30 items were designated as experimental. However, because only one new form was administered, all items were potentially available for scoring and the focus of subsequent item analysis and key validation was on determining the best set of approximately 150 items that still reflected the examination blueprint.

The initial analysis summary statistics are presented in Table 5.

Table 5: Initial examination statistics

Index	CKE 1
Items	175
Total candidates	194
Candidates in analysis	142
Mean	121.9 (69.7%)
Range	65–155 (37.1–88.6%)
Cronbach's alpha	.92
Mean r_{pb}^*	.24

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

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

Wickett compiled these statistics, along with any comments made by candidates concerning flagged items, to identify items that may have been keyed incorrectly or that were performing poorly. Most emphasis was placed on the corrected point-biserials as evidence of item quality,

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 March 1, 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	Years on EVC	Industry
Sunday Ajao	CHRL	15–20	2	Banking/finance
Roxanne Chartrand	CHRL	20–29	1	Insurance
✓ Claire Chester	CHRL	10–14	2	Regulation/CPA
✓ Tanya Gopaul	CHRL	10–15	2	Banking
✓ Jean Lazarus	CHRL	15–19	2	Health services
Suman Seth	CHRL	15–20	1	Municipal government
Kriss Stone	CHRP	10–15	2	Real estate
Ielean Tait	CHRL	15–20	2	Environmental
Karen Weiler	CHRL	20–29	2	Software/ communications
Alyssa Young	CHRL	5–9	2	Non-profit

✓ Participated in the session.

The group was informed that test reliability, as measured by Cronbach's alpha, was .92 based on the set of 175 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 they were given less direction on those with borderline statistics. Where available, candidates' comments about the items were also shown. Adherence to the blueprint was also a factor, and the committee members were directed to include whether the exam was heavy or light in a functional area in their decision making. 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 February administration.

The group made decisions based on content and the data through discussion; they removed the 20 items that they felt were least appropriate to retain for scoring. Panel members' comments about specific items were recorded for future item revision activities.

Not all remaining items were strong-performing, and several items were retained that were very easy or very hard or that had a low corrected point-biserial. Most were moderate to strong items, however. The final alpha for the set of 155 scored items was .92. The difficulties ranged from 26.1% to 94.4%, with a mean of 72.3%. The r_{pb}^* values ranged from .00 to .55, with a mean of .26.

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 February 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	7	5	6	7	4% ± 1%
20 Professional Practice	17	14	17	20	11% ± 2%
30 Organizational Effectiveness	21	18	20	23	13% ± 2%
40 Workforce Planning & Talent Management	20	18	20	23	13% ± 2%
50 Labour & Employee Relations	16	14	17	20	11% ± 2%
60 Total Rewards	20	18	20	23	13% ± 2%
70 Learning & Development	22	18	20	23	13% ± 2%
80 Health, Wellness & Safe Workplace	16	14	17	20	11% ± 2%
90 HR Metrics, Reporting & Financial Management	16	14	17	20	11% ± 2%
Total	155				

*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 February 2019 CKE 1. The goal of this process was to set a pass mark for the February 2019 CKE 1 that would be equivalent to that set for previous CKE 1 administrations; that is, to set a

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

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.

Two equating procedures were conducted back to different administrations (February 2018 and October 2018). The intention following these 2 equating runs was to average them to arrive at a final pass mark for the February 2019 CKE 1. These administrations were chosen as the most recent administration and the administration corresponding to the same administration month the previous year.

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 arc equating is most prudent.

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

The selected set of anchor items had a mean difficulty of 0.71 and a mean corrected point-biserial of .29 (for February 2019 candidates).

Table 8 shows the fit of the set of anchor items to the blueprint, as percentages. The actual counts are aligned with targets and reflect the scope and approximate weighting across the full exam.

Table 8: Anchor item fit to blueprint – To October 2018

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

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 was considered the optimal method, though the Levine observed value was also viable.

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 February 2019 CKE 1 scored the same as the population taking the October 2018 CKE 1 (70.8% vs. 70.9%, respectively; $t(318)=0.05$, *ns*). Because the February 2019 CKE 1 candidates showed the same ability on average (based on the anchors), they would likely have approximately the same pass rate as seen in October. Based on the higher total score despite the similarity in anchor item performance, the February 2019 form is easier and the pass mark should increase to reflect that.

The equating analysis bears this out (Table 10). All methods indicate a pass mark of 102 or 103 and a pass rate close to the 75.8% observed in October. The Tucker value of 101.52 was extracted from this analysis for use in setting the final pass mark.

Table 9: Equating parameter table – To October 2018

		Oct. 2018	Feb. 2019
N		178	142
Scored items		155	155
Mean score	Total	69.1%	72.3%
	Anchors	70.9%	70.8%

Table 10: Equating outcome table – To October 2018

Method	Pass Mark		Pass Rate	
	Precise	Integer	All	First-time
Equating Oct. 2018	97.39	98	67.0%	75.8%
Tucker	101.52	102	65.8%	76.1%
Levine observed	101.93	102	65.8%	76.1%
Mean	102.45	103	61.7%	72.5%
Circle arc 1	102.70	103	61.7%	72.5%
Circle arc 2	102.62	103	61.7%	72.5%

Equating Back to the February 2018 Administration

Linear equating was the chosen method for setting the pass mark, given the sample sizes involved.

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

The selected set of anchor items had a mean difficulty of 0.73 and a mean corrected point-biserial of .28 (for February 2019 candidates).

Table 11 shows the fit of the set of anchor items to the blueprint, as percentages. The actual counts are aligned with targets and reflect the scope and approximate weighting across the full exam.

Table 11: Anchor item fit to blueprint – To February 2018

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

The mean, Tucker, Levine observed-score, and circle arc methods were computed to ascertain concordance of solutions. Given the sample sizes and deviations in test parameters, Tucker was considered the optimal 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 February 2019 CKE 1 scored the same as population taking the February 2018 CKE 1 (73.1% vs. 73.2%, respectively; $t(255)=0.05$, *ns*). Because the February 2019 candidates were of the same ability (based on the anchors), that exam is expected to have approximately the same pass rate as seen on the February 2018 exam. Based on the higher total score despite the similarity in anchor item performance, the February 2019 form is easier and the pass mark should increase to reflect that; further, there were 5 additional items included in the scoring of the February 2019 form.

The equating analysis bears out the comparable pass rate and increase in pass mark (Table 13). The various methods indicate an integer pass mark of 103 or 104. The Tucker observed-score value of 102.58 was extracted from this analysis for use in setting the final pass mark.

Table 12: Equating parameter table – To February 2018

		Feb. 2018	Feb. 2019
N		115	142
Scored items		150	155
Mean score	Total	70.9%	72.3%
	Anchors	73.1%	73.2%

Table 13: Equating outcome table – To February 2018

Method	Pass Mark		Pass Rate	
	Precise	Integer	All	First-time
Equating Feb. 2018	97.71	98	64.2%	70.4%
Tucker	102.58	103	61.7%	72.5%
Levine observed	102.39	103	61.7%	72.5%
Mean	103.44	104	57.5%	69.0%
Circle arc 1	103.24	104	57.5%	69.0%
Circle arc 2	103.23	104	57.5%	69.0%

Combined Results

Table 14 shows the pass mark values across the 2 equating runs. The value highlighted in green is the one that would be selected based on sample parameters at each equating run. Barring a sound reason to choose otherwise, the simple arithmetical mean (102.054) of the 2 identified values was the recommended pass mark for the February 2019 CKE 1. Note that every equating method shows an averaged pass mark between 102 and 103.

Using the established convention for this testing program, the mean combined value was rounded up to a cut score of 103. The resulting pass rate of 72.5% for first-time candidates is about midway between the pass rates for first-time candidates from October 2018 and February 2018, which a sensible result given that the February 2019 candidates scored at about the same level when compared to those samples. The pass rate for all candidates in February 2019 was 61.9%. See Table 15 for historical pass rates.

The final pass mark value, and the process used to derive it, was presented to the CHRP EVC (Table 16) via teleconference on March 4, 2019. No concerns were raised regarding the pass mark or pass rate, and the group reported feeling comfortable with the additional analyses as backing up the final value, and that the final pass rates are not out of line with past results. The panel formally approved the pass mark (which was presented along with the consequent pass rate data) for recommendation to HRPA. The HRPA Registrar participated on the call and accepted the panel's recommendation; the pass mark was formally established.

Table 14: Equating outcome table – Combined results

	Feb. 18	Oct. 18
Tucker	102.6	101.5
Levine observed	102.4	101.9
Mean	103.4	102.4
Circle arc 1	103.2	102.7
Circle arc 2	103.2	102.6

Table 15: Historical pass rates

	Pass rate	
	All	First-time
June 16	65.4%	69.1%
Nov. 16	58.8%	62.1%
Feb. 17	50.5%	62.5%
June 17	67.8%	75.5%
Oct. 17	59.2%	66.5%
Feb. 18	64.2%	70.4%
June 18	58.6%	66.2%
Oct. 18	67.0%	75.8%
Feb. 19	61.9%	72.5%

Table 16: CHRP Examination Validation Committee members – Pass mark approval

Member	Credential	Years of Relevant Experience	Years on EVC	Industry
Sunday Ajao	CHRL	15–20	2	Banking/finance
✓ Roxanne Chartrand	CHRL	20–29	1	Insurance
✓ Claire Chester	CHRL	10–14	2	Regulation/CPA
✓ Tanya Gopaul	CHRL	10–15	2	Banking
Jean Lazarus	CHRL	15–19	2	Health services
✓ Suman Seth	CHRL	15–20	1	Municipal government
Kriss Stone	CHRP	10–15	2	Real estate
Ielean Tait	CHRL	15–20	2	Environmental
Karen Weiler	CHRL	20–29	2	Software/ communications
Alyssa Young	CHRL	5–9	2	Non-profit

✓ Participated in the session.

Scoring

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

Table 17 provides the means and standard deviations for the functional areas and for the total score, using all candidates who took the new February 2019 CKE 1 form. Table 18 provides the correlations between each functional area. 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 17: Total and functional area scores for all candidates

Functional Area	Percentage	Mean	SD*
10 Strategy	71%	5.0	1.4
20 Professional Practice	69%	11.7	2.8
30 Organizational Effectiveness	71%	14.8	3.3
40 Workforce Planning & Talent Management	70%	14.0	3.0
50 Labour & Employee Relations	71%	11.3	2.5
60 Total Rewards	68%	13.5	3.1
70 Learning & Development	69%	15.1	3.7
80 Health, Wellness & Safe Workplace	70%	11.2	2.4
90 HR Metrics, Reporting & Financial Management	69%	11.1	2.3
Total score	69.5%	107.7	19.1

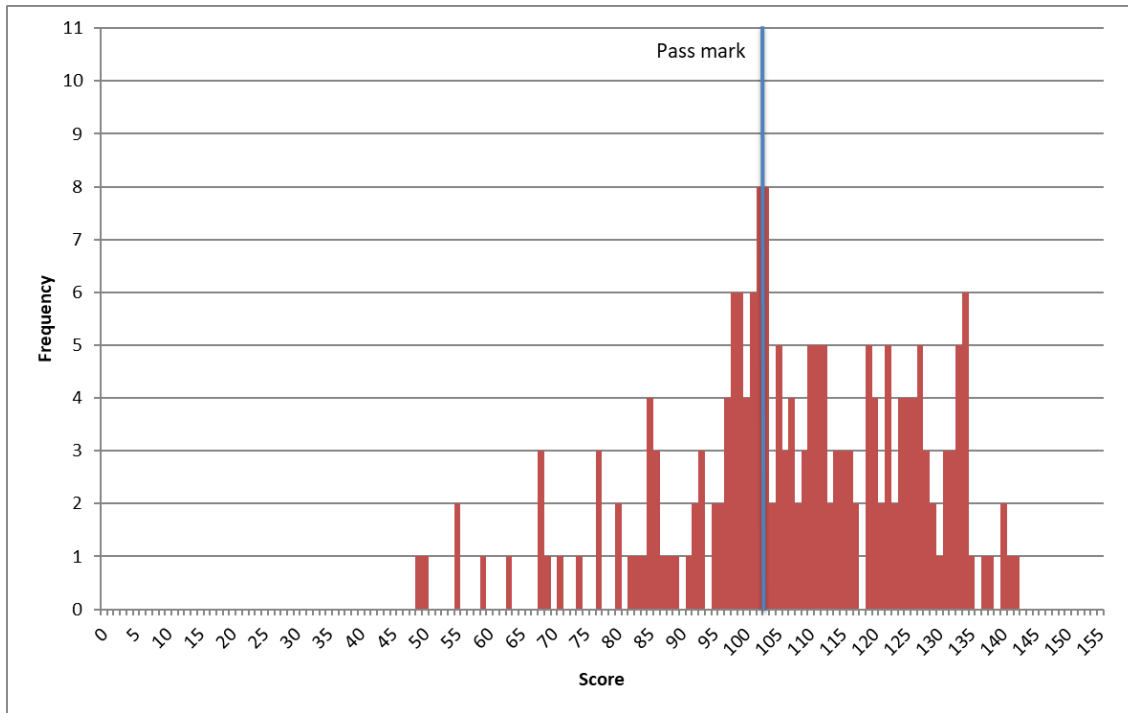
*SD = Standard deviation.

Table 18: Correlations between functional area scores for all candidates

Area*	10	20	30	40	50	60	70	80	90
10		.49	.51	.54	.45	.43	.52	.38	.35
20			.68	.64	.58	.63	.56	.64	.43
30				.63	.56	.63	.66	.56	.51
40					.60	.56	.57	.53	.50
50						.54	.57	.53	.45
60							.60	.56	.50
70								.57	.56
80									.49
90									

*See Table 17 for the full name of each functional area.

Figure 3: Score distribution for all candidates



Key Examination Metrics

Table 19 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 19: Key examination metrics – Candidates included in analysis only

Index	February 2019	October 2018	June 2018	February 2018	October 2017
Scored items	155	155	150	150	150
Candidates	142	178	157	115	182
Mean	112.1 (72.3%)	107.2 (69.1%)	102.0 (68.0%)	106.3 (70.9%)	104.3 (69.5%)
Median	114.5 (73.9%)	109 (70.3%)	103 (68.7%)	109 (72.7%)	106 (70.7%)
Skewness	-0.876	-0.462	-0.403	-0.483	-0.489
Kurtosis ⁱ	0.673	0.097	0.208	-0.525	0.024
Range	55–142 (35.5– 91.6%)	60–141 (38.7– 91.0%)	53–136 (35.3– 90.7%)	61–136 (40.7– 90.7%)	57–134 (38.0– 89.3%)
Standard deviation	18.45	15.35	16.68	17.60	15.81
Cronbach's alpha	.92	.88	.90	.92	.90
Mean r_{pb} [*]	.26	.20	.23	.25	.22
SEM ⁱⁱ	5.10	5.25	5.24	5.13	5.11
SEM at the pass mark	5.61	5.63	5.53	5.56	5.43
Decision consistency (uncorrected) ⁱⁱⁱ	.88	.89	.86	.92	.87
Perceived fairness ^{iv}	47%	46%	49%	43%	49%
Pass mark	102.054	97.387	96.622	97.710	98.387
Effective pass mark	103	98	97	98	99
Pass rate	72.5%	75.8%	66.2%	70.4%	66.5%

ⁱExcess

ⁱⁱSEM = standard error of measurement.

ⁱⁱⁱSubkoviak method.

^{iv}Based on responses to the post-examination survey.

Related Development Activities

Since the last administration of the CKE 1 in October 2018, the following exam development activities have taken place.

Item Validation

To provide sufficient items for upcoming administrations, a validation session was held November 19–20, 2018, at HRPAs offices. The CHRP EVC members who participated are listed in Table 20. This session involved the review of CHRP ELE items as well.

Table 20: CHRP Examination Validation Committee – Item validation

Member	Credential	Years of Relevant Experience	Years on EVC	Industry
Sunday Ajao	CHRL	15–20	1	Banking/finance
✓ Claire Chester	CHRL	10–14	1	Regulation/CPA
✓ Tanya Gopaul	CHRL	10–15	1	Banking
✓ Jean Lazarus	CHRL	15–19	1	Health services
Kriss Stone	CHRP	10–15	1	Real estate
Ilelean Tait	CHRL	15–20	1	Environmental
✓ Karen Weiler	CHRL	20–29	1	Software/ communications
✓ Alyssa Young	CHRL	5–9	1	Non-profit

✓ Participated in the session.

The EVC members received advance materials outlining the following:

- Purpose of the session
- Description of the CHRP credential
- CKE 1 and CHRP ELE blueprints
- Criteria for good test items
- Validation process
- Relevant legislation

At the session, abbreviated training was provided as these committee members were already informed about the credentials and general process from previous sessions. This training included a primer on psychometrics to orient the committee to the item statistics they would be working with.

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 an HR professional in Ontario (at the CHRP level),
- Move the item to the CKE 2 or CHRP ELE bank,
- Revise the item to make it suitable for use, or
- Declare the item unsound and send it back for revision or removal from the bank.

The bulk of the session saw the committee members reviewing items independently and submitting their assessments in blocks of approximately 15–25 items. Those assessments were tabulated and any items that were not validated as is by the full committee were discussed until there was agreement on changes and the future use of the item.

The panel members reviewed and validated 202 items as suitable for the CKE 1, designated 1 item to move to the CKE 2 bank, changed 1 item's competency code to a code only suitable for use on the ELE, designated 0 items as suitable for one of the ELEs, and rejected 3 items as unfixable. Very few items were edited as the items had gone through considerable review before getting to the committee for validation.

CKE 1 Blueprint Revision

At the validation session held November 19–20, 2018 (see section above), the CHRP EVC was asked to consider a minor revision to the purpose statement of the CKE 1 blueprint. The approved wording is as follows:

The CKE 1 assesses whether a candidate has the level of discipline-specific knowledge necessary to practice 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.

The changes were made in part for consistency with the CKE 2 blueprint and had already been reviewed as acceptable by the HRP A Registrar. These relatively minor changes to better reflect intent of the CKE 1 were deemed to be immediately in effect.

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 practice 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 21: 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 22: Functional area weights on the CKE 1

Functional Area		CKE 1	
		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 23: 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		C125		C165		C196
	C036	60	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 by 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.