PUBLIC RELEASE

Technical Report: October 2018 CKE 1

Human Resources Professionals Association

1 November 2018

ICKETT
MEASUREMENT SYSTEMS

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 **25**
Related Development Activities **26**
Item Writing 26
Item Review 27
Appendix A **28**
Appendix B **31**

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 June 2018 15
Table 9: Equating parameter table – To June 2018 15
Table 10: Equating outcome table – To June 2018 16
Table 11: Anchor item fit to blueprint – To October 2017 17
Table 12: Equating parameter table – To October 2017 17
Table 13: Equating outcome table – To October 2017 18
Table 14: Anchor item fit to blueprint – To February 2018 19
Table 15: Equating parameter table – To February 2018 19
Table 16: Equating outcome table – To February 2018 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 22
Table 20: Total and functional area scores for all candidates 23
Table 21: Correlations between functional area scores for all candidates 23
Table 22: Key examination metrics – Candidates included in analysis only 25
Table 23: Item writers 26
Table 24: Panel for item review session 27
Table 25: CKE 1 Blueprint structural variables 28
Table 26: Functional area weights on the CKE 1 29
Table 27: Competencies not eligible on the CKE 1 30

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 24

Executive Summary[1]

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 230 candidates using computer-based testing at Prometric test centres October 9–23, 2018, 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=178[2]) was 107.2 (69.1%), and for all candidates it was 103.4 (66.7%), out of 155 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 October 2017, February 2018 and June 2018 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 75.8% and a pass rate for all candidates of 67.0%.

This report, the analyses performed, and the processes followed are consistent with NCCA standards[3] 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 HRPA 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 October 2018 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, HRPA 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 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 October 9–23, 2018, inclusive, and 230 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 (12 minutes more than in June 2018). The time limit on the CKE 1 was 3½ hours, suggesting that time was not a factor in scores across candidates.

Eleven candidates (5%) 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 .03, 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 a drop of 2.1 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 at the end of the window rather than the start.

Figure 1: Examination time distribution for all candidates

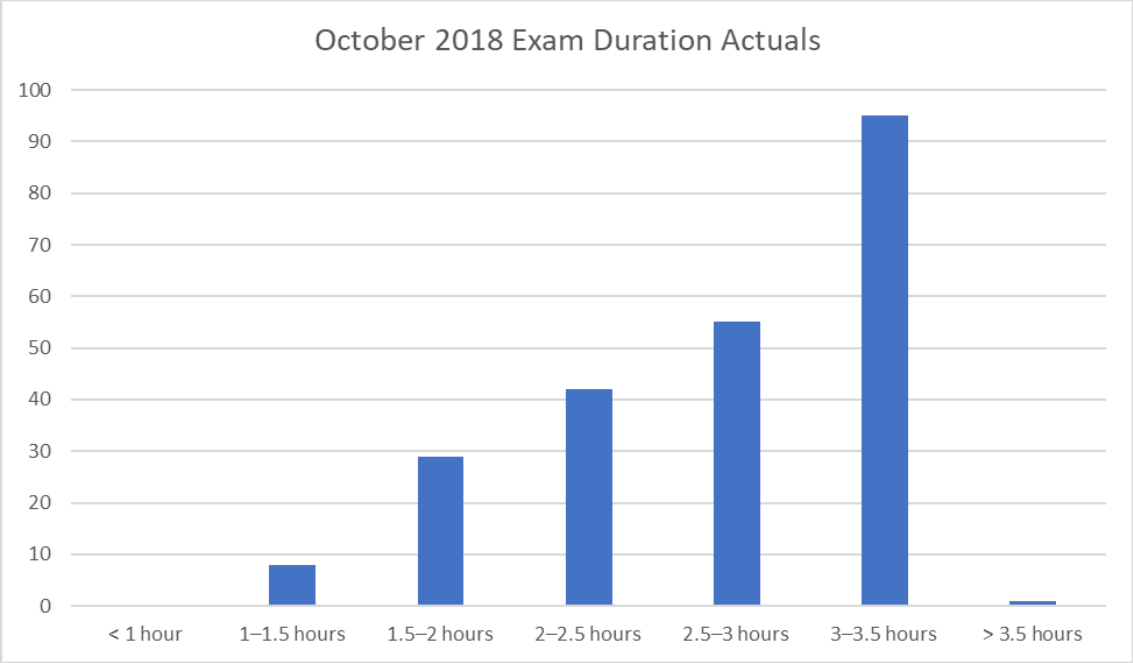
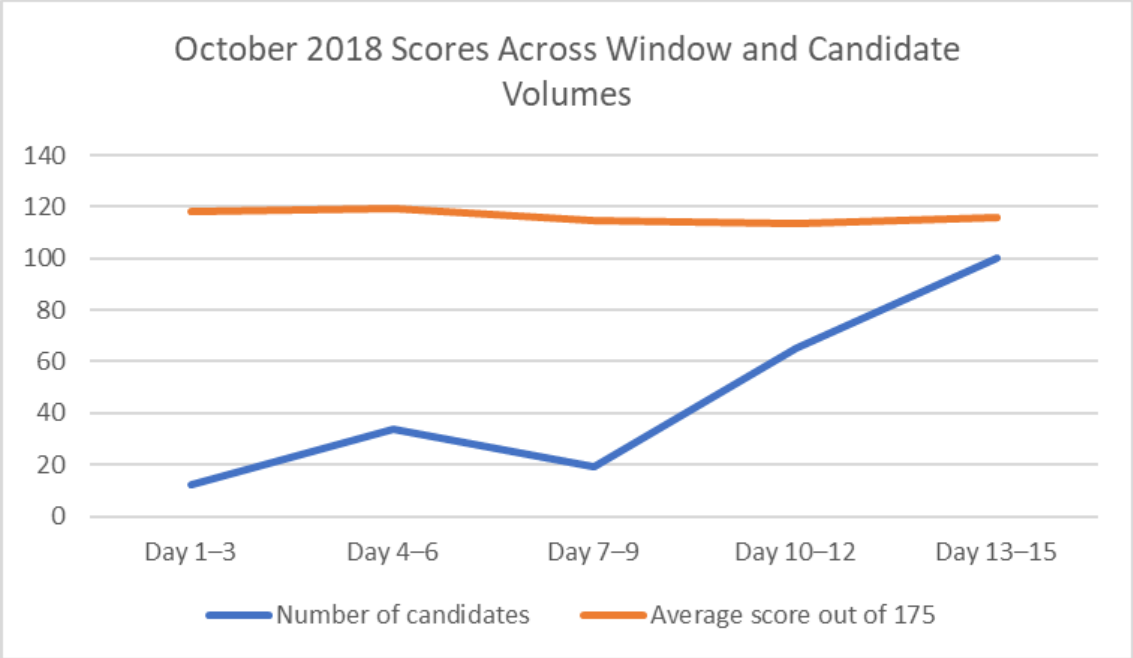


Figure 2: Candidate volume and score trends across testing window



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. Three 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, 6 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 (59.8% and 67.8%, respectively, on the full exam of 175 items). This difference was meaningful and significant ($t(77)=5.36$, $p<.001$). In keeping with standard procedures, these candidates were removed from subsequent analyses. The CKE 1 analysis proceeded with 178 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; 225 responses were obtained from candidates (response rate, 98%).

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 more than half of candidates felt it made no difference, and that where a preference was expressed it was 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 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.	125	71	10	11	8	4.4	87%
2.	I was well informed about what documents to bring to the exam location.	155	60	1	5	1	4.6	97%
3.	Proctors enforced the exam-day rules and the security procedures at the test centre were what I expected.	155	58	5	3	0	4.7	96%
4.	Proctors were professional and courteous.	154	57	5	5	0	4.7	95%
5.	The tutorial helped me understand how to complete the examination on the computer.	133	70	15	2	0	4.6	92%
6.	Navigation through the examination was easy and intuitive.	140	68	8	2	1	4.6	95%

*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.	118	75	9	13	4	4.4	88%
8.	Information available prior to exam day provided me with adequate details about the content and format of the exam.	56	87	49	17	8	4.0	66%
9.	I feel I was adequately prepared to write this examination.	22	94	64	32	6	3.7	53%
10.	The questions in the examination were clearly written.	24	101	54	35	4	3.7	57%
11.	The terminology used in the examination was accurate.	34	127	41	14	2	4.0	74%
12.	The situations presented in the examination were realistic.	35	142	33	6	2	4.1	81%
13.	The questions in the examination reflected the examination blueprint.	24	103	72	12	6	3.9	59%
14.	The examination was a fair assessment of my ability.	18	81	68	39	10	3.6	46%

*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.	56	26%
I feel that completing the examination on a computer decreased my performance.	33	15%
I feel that completing the examination on a computer had no effect on my performance.	129	59%

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	230
Candidates in analysis	178
Mean	119.8 (68.4%)
Range	75–155 (42.9–88.6%)
Cronbach's alpha	.87
Mean r_{pb}*	.18

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 October 29, 2018, 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	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
Ielean 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 group was informed that test reliability, as measured by Cronbach's alpha, was .87 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 October 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 .88. The difficulties ranged from 28.1% to 94.4%, with a mean of 69.1%. The r_{pb}* values ranged from .02 to .41, with a mean of .20.

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 2018 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	20	11% ± 2%
30	**Organizational Effectiveness**	21	18	20	23	13% ± 2%
40	**Workforce Planning & Talent Management**	19	18	20	23	13% ± 2%
50	**Labour & Employee Relations**	17	14	17	20	11% ± 2%
60	**Total Rewards**	21	18	20	23	13% ± 2%
70	**Learning & Development**	21	18	20	23	13% ± 2%
80	**Health, Wellness & Safe Workplace**	18	14	17	20	11% ± 2%
90	**HR Metrics, Reporting & Financial Management**	15	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 October 2018 CKE 1. The goal of this process was to set a pass mark for the October 2018 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 (June 2018, February 2018, and October 2017). The intention following these 3 equating runs was to average them to arrive at a final pass mark for the October 2018 CKE 1. These administrations were chosen because they were the most recent administrations and the administration corresponding to the same administration month the previous year. Though normally equating would not be conducted back to three administrations, but deviations between the first two planned equating runs led to adding a third for confirmation of the final pass mark.

Equating Back to the June 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 0.70 and a mean corrected point-biserial of .22 (for October 2018 candidates).

Table 14 shows the fit of the set of anchor items to the blueprint, as percentages. The actual counts are reasonably aligned with targets and reflect the scope and approximate weighting across the full exam.

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

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

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 2018 CKE 1 scored higher than the population taking the June 2018 CKE 1 (69.6% vs. 66.9%, respectively; $t(333)=1.86$, *ns*). Though not quite statistically significant, the difference is still substantial. Because the October 2018 CKE 1 candidates showed higher ability (based on the anchors), they would likely have a higher pass rate than seen in June.

The equating analysis bears this out (Table 16). All methods indicate the same integer pass mark of 99. The Tucker value of 98.88 was extracted from this analysis for use in setting the final pass mark.

Table 9: Equating parameter table – To June 2018

		June 2018	Oct. 2018
	n	157	178
	Scored items	150	155
Mean score	Total	68.0%	69.1%
Mean score	Anchors	66.9%	69.6%

Table 10: Equating outcome table – To June 2018

	Pass Mark		Pass Rate	
Method	**Precise**	**Integer**	**All**	**First-time**
Combo June 2018	**96.62**	**97**	**58.6%**	**66.2%**
Tucker	98.88	99	*66.1%*	*74.7%*
Levine observed	98.07	99	*66.1%*	*74.7%*
Mean	98.80	99	*66.1%*	*74.7%*
Circle arc 1	98.31	99	*66.1%*	*74.7%*
Circle arc 2	98.31	99	*66.1%*	*74.7%*

Equating Back to the October 2017 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 0.70 and a mean corrected point-biserial of .22 (for October 2018 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 October 2017

Area	Actual	Target
10	5%	4%
20	12%	11%
30	12%	13%
40	12%	13%
50	10%	11%
60	12%	13%
70	14%	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, Levine observed-score 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 October 2018 CKE 1 scored significantly higher than the population taking the October 2017 CKE 1 (70.0% vs. 66.5%, respectively; $t(358)=2.55$, $p<.05$). Because the October 2018 candidates were of greater ability (based on the anchors), that exam is expected to have a higher pass rate than the October 2017 exam.

The equating analysis bears out the higher pass rate, but with considerable variability across equating methods (Table 13). The various methods indicate an integer pass mark of 96 to 98, and all values are below the 99 estimated in the equating back to June 2018. The Levine observed-score value of 95.109 was extracted from this analysis for use in setting the final pass mark.

Table 12: Equating parameter table – To October 2017

		Oct. 2017	Oct. 2018
	n	182	178
	Scored items	150	155
Mean score	Total	69.5%	69.1%
	Anchors	66.5%	70.0%

Table 13: Equating outcome table – To October 2017

	Pass Mark		Pass Rate	
Method	**Precise**	**Integer**	**All**	**First-time**
Combo Oct. 2017	**98.39**	**99**	**59.2%**	**66.5%**
Tucker	96.53	97	70.9%	79.2%
Levine observed	95.11	96	*71.7%*	*79.8%*
Mean	97.37	98	*67.0%*	*75.8%*
Circle arc 1	96.40	97	*70.9%*	*79.2%*
Circle arc 2	96.35	97	*70.9%*	*79.2%*

Because of the variability observed within the October 2017 equating run and the divergence with the June 2018 equating run, it was decided to also equate to February 2018.

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 0.71 and a mean corrected point-biserial of .24 (for October 2018 candidates).

Table 14 shows the fit of the set of anchor items to the blueprint, as percentages. The actual counts are reasonably aligned with targets and reflect the scope and approximate weighting across the full exam.

Table 14: Anchor item fit to blueprint – To February 2018

Area	Actual	Target
10	3%	4%
20	13%	11%
30	8%	13%
40	13%	13%
50	10%	11%
60	18%	13%
70	15%	13%
80	10%	11%
90	13%	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.

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 2018 CKE 1 scored slightly higher than the population taking the February 2018 CKE 1 (70.5% vs. 69.7%, respectively; $t(291)=0.54$, *ns*). Because the October 2018 CKE 1 candidates showed slightly higher ability (based on the anchors, non-significance notwithstanding), they would likely have a higher pass rate than seen in February.

The equating analysis bears this out (Table 16). All methods indicate a pass mark of 98 or 99. The Tucker value of 98.17 was extracted from this analysis for use in setting the final pass mark.

Table 15: Equating parameter table – To February 2018

		Feb. 2018	Oct. 2018
	n	115	178
	Scored items	150	155
Mean score	Total	70.9%	69.1%
Mean score	Anchors	69.7%	70.5%

Table 16: Equating outcome table – To February 2018

	Pass Mark		Pass Rate	
Method	**Precise**	**Integer**	**All**	**First-time**
Combo February 2018	**97.71**	**98**	**64.2%**	**70.4%**
Tucker	98.17	99	*66.1%*	*74.7%*
Levine observed	97.69	98	*67.0%*	*75.8%*
Mean	97.61	98	*67.0%*	*75.8%*
Circle arc 1	97.12	98	*67.0%*	*75.8%*
Circle arc 2	97.09	98	*67.0%*	*75.8%*

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. Barring a sound reason to choose otherwise, the simple arithmetical mean (97.386980) of the 3 identified values was the recommended pass mark for the October 2018 CKE 1.

It is notable that the Tucker values for February 2018 and June 2018 equating were comparable, and that the equating to October 2017 stands out as more anomalous. However, this seems largely due to differences in anchor set variances across administrations, and so the mean equating values may be more interpretable (and they show closer concordance across equating runs). Note that most methods of averaging across the 3 equating runs will yield a value between 97 and 98, which will round up to 98.

Note that considerable error-checking and re-running of analyses in a second statistical package was conducted to verify these results (beyond the normal process owing to the spread of results). A full second set of equating analyses were also run where borderline outlier candidates were removed, and where anchor item selection was focused primarily on maximizing the correlation between anchor sets. This full analysis yielded a final averaged value of 97.48, which is essentially the same as what was computed in the primary analysis.

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 75.8% for first-time candidates is about 9½ percentage points higher than that seen in October 2017 and February 2018 and about 5½ percentage points higher than that seen in June 2018; these increases were anticipated based on the higher performance on the common anchor items. The pass rate for all candidates in October 2018 was 67.0%. 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 1, 2018. 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 17: Equating outcome table – Combined results

	Oct. 17	Feb. 18	June 18
Tucker	96.5	98.2	98.9
Levine observed	95.1	97.7	98.1
Mean	97.4	97.6	98.8
Circle arc 1	96.4	97.1	98.3
Circle arc 2	96.3	97.1	98.3

Table 18: 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%**

Table 19: CHRP Examination Validation Committee members – Pass mark approval

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
Ielean 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.

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 HRPA.

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 2018 CKE 1 form. Table 21 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 20: Total and functional area scores for all candidates

	Functional Area	Percentage	Mean	SD*
10	**Strategy**	75%	4.5	1.2
20	**Professional Practice**	73%	12.5	2.5
30	**Organizational Effectiveness**	61%	12.9	3.1
40	**Workforce Planning & Talent Management**	67%	12.7	2.6
50	**Labour & Employee Relations**	68%	11.5	2.4
60	**Total Rewards**	67%	14.0	3.2
70	**Learning & Development**	64%	13.5	3.5
80	**Health, Wellness & Safe Workplace**	69%	12.4	2.4
90	**HR Metrics, Reporting & Financial Management**	63%	9.4	2.2
Total score		66.7%	103.4	17.1

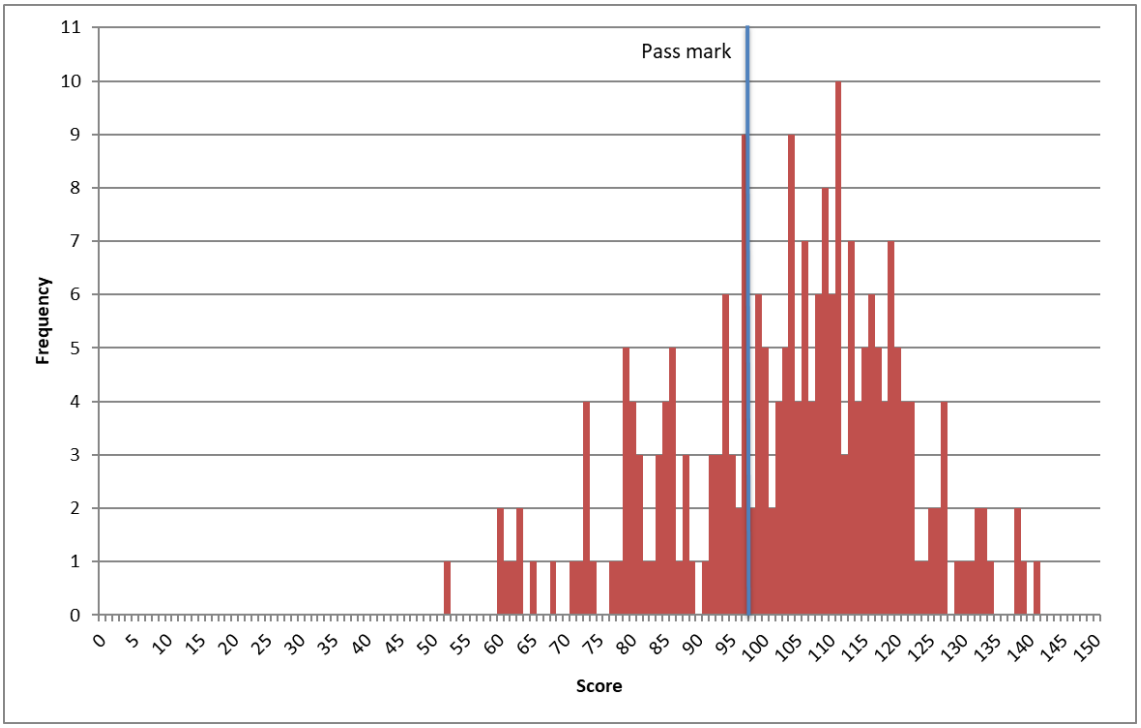
*SD = Standard deviation.

Table 21: Correlations between functional area scores for all candidates

Area*	10	20	30	40	50	60	70	80	90
10		.46	.46	.42	.39	.43	.45	.40	.38
20			.58	.55	.52	.59	.55	.44	.43
30				.48	.53	.54	.60	.46	.38
40					.42	.57	.55	.46	.41
50						.55	.46	.49	.37
60							.53	.49	.45
70								.44	.32
80									.39
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 2018	June 2018	February 2018	October 2017	June 2017
Scored items	155	150	150	150	150
Candidates	178	157	115	182	147
Mean	107.2 (69.1%)	102.0 (68.0%)	106.3 (70.9%)	104.3 (69.5%)	103.8 (69.2%)
Median	109 (70.3%)	103 (68.7%)	109 (72.7%)	106 (70.7%)	104 (69.3%)
Skewness	−0.462	−0.403	−0.483	−0.489	−0.230
Kurtosis[i]	0.097	0.208	−0.525	0.024	−0.578
Range	60–141 (38.7–91.0%)	53–136 (35.3–90.7%)	61–136 (40.7–90.7%)	57–134 (38.0–89.3%)	66–136 (44.0–90.7%)
Standard deviation	15.35	16.68	17.60	15.81	15.96
Cronbach's alpha	.88	.90	.92	.90	.89
Mean r_{pb}*	.20	.23	.25	.22	.22
SEM[ii]	5.25	5.24	5.13	5.11	5.19
SEM at the pass mark	5.63	5.53	5.56	5.43	5.58
Decision consistency (uncorrected)[iii]	.89	.86	.92	.87	.86
Perceived fairness[iv]	46%	49%	43%	49%	36%
Pass mark	97.387	96.622	97.710	98.387	94.496
Effective pass mark	98	97	98	99	95
Pass rate	75.8%	66.2%	70.4%	66.5%	75.5%

[i]Excess

[ii]SEM = standard error of measurement.

[iii]Subkoviac method.

[iv]Based on responses to the post-examination survey.

Related Development Activities

Since the last administration of the CKE 1 in June 2018, 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 2018. Item writers (see Table 23) were identified by HRPA and trained in a remote session by Wickett on June 25, 2018. Note that 5 writers were recruited, but 1 dropped out very early in the process.

Table 23: Item writers

Writer	Credentials	Years of Relevant Experience	Industry
Richard McFadden	MBA, CRSP, CHRL	29 years working in HR, 10 years teaching in HR	Education – Professor at Georgian College
Parbudyal Singh	PhD	5 years working in HR, 18 years teaching in HR	Education – Professor at York University
Carol Tuck-Riggs	PhD	20+ years working and teaching in HR	Education – Professor at Sheridan College
Gordon Wang	MHRM	6 years working in HR, 8 years teaching in HR	Education – Professor at York University; HR Consultant

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).[6]

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 102 new items were obtained).

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

[6] 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.

Item Review

Following the item writing exercise in June–August 2018 there was need for group review of those items before moving them to formal validation and use on the CKE 1. The group had 373 items for consideration (taken newly written items, supplemented with other unreviewed items in the bank and items requiring revision required to fill gaps in the bank). The group was not expected to be able to review all items in the time available.

The 2½-day review session was held September 24–26, 2018 at HRPA 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
Lisa Eisen	CHRL	14	Financial services
Maria Kapsamer	CHRL	21	Crown corporation
Mary Emirzian	CHRL	11	Construction
Paul Eves	CHRL	18	Education
Silvia Marabeti	CHRL	20	Not-for-profit homecare organization
Stephanie Izzard	CHRL	18 + CCHRA experience	Crown corporation (recently retired)
Suman Seth	CHRL	12	Municipal government

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 129 items as suitable for CKE 1, moved 7 items to the CKE 2 bank, and rejected 22 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 and competency for 20 items that had been reviewed in the past and revised or added rationales and references throughout the process.

The group reported great confidence in the final assessments made by the group. Except for not having as much time as preferred on the last day, the group reported satisfaction overall with the session.

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 2.1

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 knowledge required to be an effective human resources professional at the CHRP level, in Ontario. Knowledge related exclusively to employment-related legislation will be 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

		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
10	C005
	C007
	C009
	C011
	C012
	C017
20	C035
	C036
	C037
	C041
30	C050
	C056
	C057
	C065

FA	Comp
40	C084
	C089
50	C113
	C114
	C117
	C123
	C125
60	C139
	C141
	C143
	C146

FA	Comp
70	C152
	C155
	C156
	C158
	C159
	C163
	C165
	C166
	C171
	C172
	C173
	C175

FA	Comp
80	C177
	C179
	C187
	C192
90	C194
	C195
	C196
	C204
	C205
	C206
	C210

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.