PUBLIC RELEASE

Technical Report: February 2020 CKE 1

Human Resources Professionals Association

10 March 2020

ICKETT
MEASUREMENT SYSTEMS

Contents

List of Tables

List of Figures

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 199 candidates using computer-based testing at Prometric test centres February 10–25, 2020, inclusive. The examination comprised 175 four-option multiple choice items and had a 3½-hour time limit.

As per the CKE 1 blueprint, the exam was scored using the 145–155 best-performing items (while adhering to the prescribed distribution across functional areas). The mean score for first-time candidates[2] (*n*=144) was 109.1 (71.3%), and for all candidates it was 105.1 (68.7%), out of 153 scored items. Reliability was strong at .90. The final set of scored items adhered to the blueprint parameters.

The pass mark was set using equating back to the February 2019 and October 2019 administrations, yielding an integer pass mark of 99. 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 76.4% and a pass rate for all candidates of 65.3%.

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 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 February 2020 CKE 1 forms is shown in Table 1. All functional areas are within the limits of their targets, and therefore the forms reflect the blueprint (see Appendix A for the CKE 1 blueprint).

Note that at any administration, HRPA 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	19–20	19
30	Organizational Effectiveness	23–24	23
40	Workforce Planning & Talent Management	22–23	23
50	Labour & Employee Relations	19–20	19
60	Total Rewards	22–23	23
70	Learning & Development	23–24	23
80	Health, Wellness & Safe Workplace	19–20	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 10–25, 2020, inclusive, and 199 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, 37 minutes (7 minutes less than in October 2019; on average, form A candidates took 2 hours, 37 minutes, form B candidates took 2 hours, 42 minutes, and form C candidates took 2 hours, 32 minutes). The time limit on the CKE 1 was 3½ hours, suggesting that time was not a factor in scores across candidates. Two candidates who were granted additional time as a testing accommodation spent more than 3½ hours.

Eight candidates (4%) took the full 3½ hours, suggesting that those candidates may have wanted more time, and 2 candidates (1.0%) left at least 1 item blank, suggesting that those candidates timed out of the exam before being able to complete it. These metrics will continue to be monitored, but at present do not appear problematically high.

The correlation between scores on the 175 items and time spent writing the examination was small at a value of −.13 for form A, small at a value of −.11 for form B, and small at a value of −.11 for form C, suggesting that time constraints did not generally have an impact on candidate performance.

Candidate scores across the window were computed to look for any evidence of item exposure. As shown in Figure 2, there was little variation across the window. The difference between scores for candidates writing in the first 3 days and those writing in the last 3 days was an increase of 1.4 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

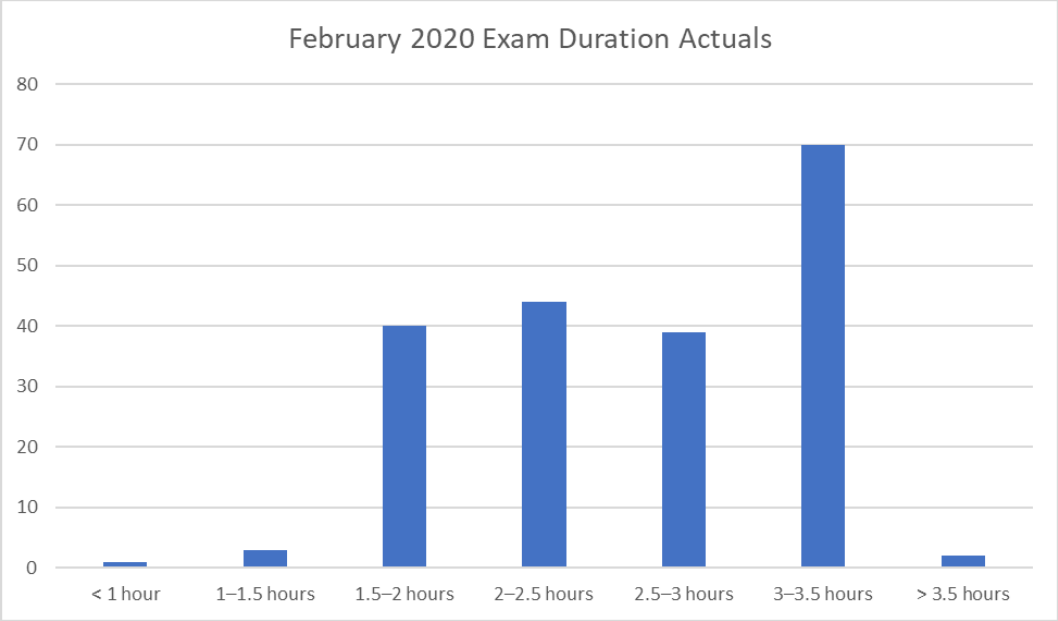
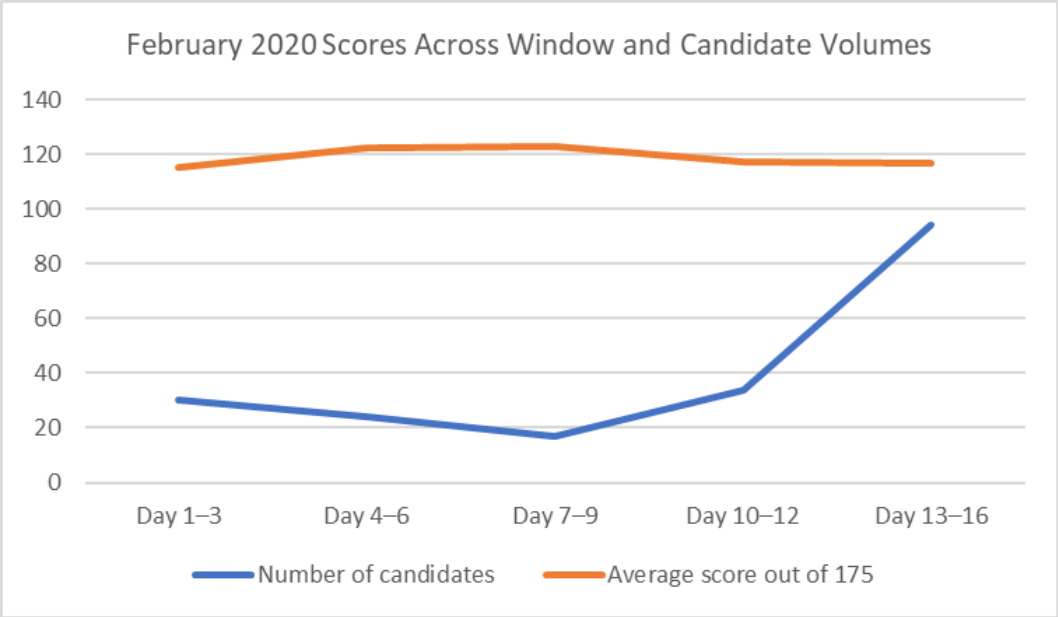


Figure 2: Candidate volume and score trends across testing window



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. No 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 (60.2% and 69.6%, respectively, on the full exam of 175 items). This difference was meaningful and significant ($t(131)=7.26$, $p<.001$). In keeping with standard procedures, these candidates were removed from subsequent analyses. The CKE 1 analysis proceeded with 144 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; 194 responses were obtained from candidates (response rate, 97%).

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.	97	72	9	14	2	4.3	87%
2.	I was well informed about what documents to bring to the exam location.	139	52	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.	141	50	3	0	0	4.7	98%
4.	Proctors were professional and courteous.	142	43	6	2	1	4.7	95%
5.	The tutorial helped me understand how to complete the examination on the computer.	123	65	5	1	0	4.6	97%
6.	Navigation through the examination was easy and intuitive.	133	57	2	2	0	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.	110	66	9	6	3	4.5	91%
8.	Information available prior to exam day provided me with adequate details about the content and format of the exam.	47	84	33	26	4	3.9	68%
9.	I feel I was adequately prepared to write this examination.	17	94	61	20	2	3.9	57%
10.	The questions in the examination were clearly written.	23	95	43	32	1	3.8	61%
11.	The terminology used in the examination was accurate.	26	121	32	14	1	4.0	76%
12.	The situations presented in the examination were realistic.	38	125	23	6	2	4.1	84%
13.	The questions in the examination reflected the examination blueprint.	19	90	65	19	1	3.9	56%
14.	The examination was a fair assessment of my ability.	22	73	55	36	8	3.6	49%

*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.	58	30%
I feel that completing the examination on a computer decreased my performance.	27	14%
I feel that completing the examination on a computer had no effect on my performance.	108	56%

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	199
Candidates in analysis	144
Mean score	109.9 (71.4%)
Score range	67–137 (43.5–89.0%)
Cronbach's alpha	.90
Mean r_{pb}*	.23

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

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

Wickett compiled these statistics, along with any comments made by candidates concerning specific items, to identify items that may have been keyed incorrectly or that were performing poorly. Most emphasis was placed on the corrected point-biserials as evidence of item quality and on difficulty through removal of ineffective very easy or very hard items. Items were ranked from worst performing to best performing accordingly.

Key Validation

Key validation was conducted via web meeting on February 27, 2020, 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 .90 based on the set of 154 potentially scored items and that this was well above the generally accepted threshold of .80.

The group was walked through the flagged items one at a time, with the recommendation that the worst-performing items be removed from scoring, but the group was given less direction on items with borderline statistics. Where available, candidates' comments about the items were also shown. Because of the modest sample size for this administration, past item data were also used where available, and the group was directed not to rely unduly on statistics exclusively from the February 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 .90. The difficulties ranged from 35.4% to 94.4%, with a mean of 71.3%. The r_{pb}* values ranged from −.06 to .46, with a mean of .23.

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 2020 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**	19	17	20	22	13% ± 2%
70	**Learning & Development**	20	17	20	22	13% ± 2%
80	**Health, Wellness & Safe Workplace**	17	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 February 2020 CKE 1. The goal of this process was to set a pass mark for the February 2020 CKE 1 that would be equivalent to that set for previous CKE 1 administrations; that is, to set a pass mark that would give each candidate the same probability of passing regardless of which form they took.

The passing standard for the CKE 1 was originally set after the November 2015 offering of the CKE 1 using the Modified Angoff method. General details on that method can be found in

[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.

Two equating procedures were conducted back to different administrations (February 2019 and October 2019). These two 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 October 2019 Administration

Linear equating was the chosen method for setting the pass mark. Linear equating is preferred with more than 100 candidates, and equipercentile equating is preferred with more than 1,000 candidates. With candidate samples of fewer than 100, mean or circle arc[6] equating is most prudent.

All candidates in the analysis (i.e., no repeat candidates or outliers) were used in the equating process. Delta plot analysis was used to identify anchor items showing substantial deviations (generally, although not exclusively, greater than 3 SD units) from expected difficulty values, with an emphasis on establishing an anchor set with difficulty equivalent to that of the full form (and equivalent within each functional area) that adhered to the blueprint. Items with an increase or decrease of 10% in terms of difficulty were also removed as anchors. Further, items with very high or low difficulty values and those with low corrected point-biserials were also flagged for potential removal from the anchor set. The goal was a strong midi-test (i.e., moderate range of difficulty, moderate to high discrimination, fit to blueprint) of sufficient length to estimate candidate ability.

The selected set of anchor items had a mean difficulty of 0.71 and a mean corrected point-biserial of .26 (for February 2020 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 October 2019

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%

*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 candidates taking the February 2020 CKE 1 scored modestly higher than the candidates taking the October 2019 CKE 1 (71.0% vs. 69.0%; $t(317)=1.40$, *ns*). Because the February 2020 CKE 1 candidates scored modestly higher (based on the anchors, non-significance notwithstanding), they would likely have a modestly higher pass rate as compared to October candidates.

The equating analysis bears this out (Table 10). All methods indicate a pass mark of 98 or 99. The first-time candidate pass rate based on this equating run is higher, as expected, than what was seen in October 2019. The Tucker equating value of 98.19 was extracted from this analysis for use in setting the final pass mark.

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

		Oct. 2019	Feb. 2020
	N	175	144
	Scored items	153	153
Mean score	Total	69.5%	71.3%
	Anchors	69.0%	71.0%

Table 10: Equating outcome table – Total pass mark, to October 2019

	Pass Mark		Pass Rate	
Method	**Precise**	**Integer**	**All**	**First-time**
Equating Oct. 2019	**97.50**	**98**	**66.2%**	**74.3%**
Tucker	98.19	99	*65.3%*	*76.4%*
Levine observed	98.07	99	*65.3%*	*76.4%*
Mean	97.91	98	*66.8%*	*78.5%*
Circle Arc 1	97.74	98	*66.8%*	*78.5%*
Circle Arc 2	97.74	98	*66.8%*	*78.5%*

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 arc[7] 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 0.72 and a mean corrected point-biserial of .25 (for February 2020 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. There was an unexpected dip in the proportion of suitable anchor items from functional area 80, which is suggestive of shifts in candidate competence in that area.

[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 February 2019

Area*	Actual	Target
10	5%	4%
20	13%	11%
30	16%	13%
40	11%	13%
50	11%	11%
60	16%	13%
70	13%	13%
80	5%	11%
90	11%	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 candidates taking the February 2020 CKE 1 scored marginally higher than the candidates taking the February 2019 CKE 1 (72.4% vs. 71.8%; $t(284)=0.37$, *ns*). Because the February 2020 CKE 1 candidates scored marginally higher (non-significance notwithstanding), they would likely have a marginally higher pass rate as compared to February 2019 candidates.

The equating analysis bears this out (Table 13). All methods indicate a pass mark of 99. The pass rate based on this equating run is higher, as expected, than what was seen in February 2019. The Tucker equating value of 98.78 was extracted from this analysis for use in setting the final pass mark.

Table 12: Equating parameter table – Total pass mark, to February 2019

		Feb. 2019	Feb. 2020
	N	142	144
	Scored items	155	153
Mean score	Total	72.3%	71.3%
	Anchors	71.8%	72.4%

Table 13: Equating outcome table – Total pass mark, to February 2019

	Pass Mark		Pass Rate	
Method	**Precise**	**Integer**	**All**	**First-time**
Equating Feb. 2019	**102.05**	**103**	**61.9%**	**72.5%**
Tucker	98.78	99	*65.3%*	*76.4%*
Levine observed	98.32	99	*65.3%*	*76.4%*
Mean	98.34	99	*65.3%*	*76.4%*
Circle Arc 1	98.24	99	*65.3%*	*76.4%*
Circle Arc 2	98.23	99	*65.3%*	*76.4%*

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. The weighted mean (by number of anchor items and number of candidates) of the 2 identified values was the recommended pass mark for the February 2020 CKE 1 (98.458).

Using the established convention for this testing program, the mean combined value was rounded up to a cut score of 99. The resulting pass rate of 76.4% for first-time candidates is modesty higher than what was seen in October 2019 and February 2019, as expected based on the February 2020 candidates being a somewhat better performing group on average. The pass rate for all candidates in February 2020 was 65.3%. 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 10, 2020. 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 HRPA. The HRPA Registrar joined the panel at the end of the call and accepted the panel's recommendation; the pass mark was formally established.

Table 14: Equating outcome table – Combined results

	Oct. 19	Feb. 19
Tucker	98.8	98.2
Levine observed	98.3	98.1
Mean	98.3	97.9
Circle arc 1	98.2	97.7
Circle arc 2	98.2	97.7

Table 15: Historical pass rates

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

Table 16: 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 HRPA.

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 2020 CKE 1 forms. Table 18 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 17: Total and functional area scores for all candidates

	Functional Area	Percentage	Mean	SD*
10	**Strategy**	73%	4.4	1.2
20	**Professional Practice**	68%	11.6	2.6
30	**Organizational Effectiveness**	69%	13.9	2.8
40	**Workforce Planning & Talent Management**	69%	13.8	2.7
50	**Labour & Employee Relations**	69%	11.7	2.6
60	**Total Rewards**	67%	12.8	2.9
70	**Learning & Development**	68%	13.5	3.1
80	**Health, Wellness & Safe Workplace**	70%	11.9	2.2
90	**HR Metrics, Reporting & Financial Management**	68%	11.5	2.4
Total score		68.7%	105.1	16.8

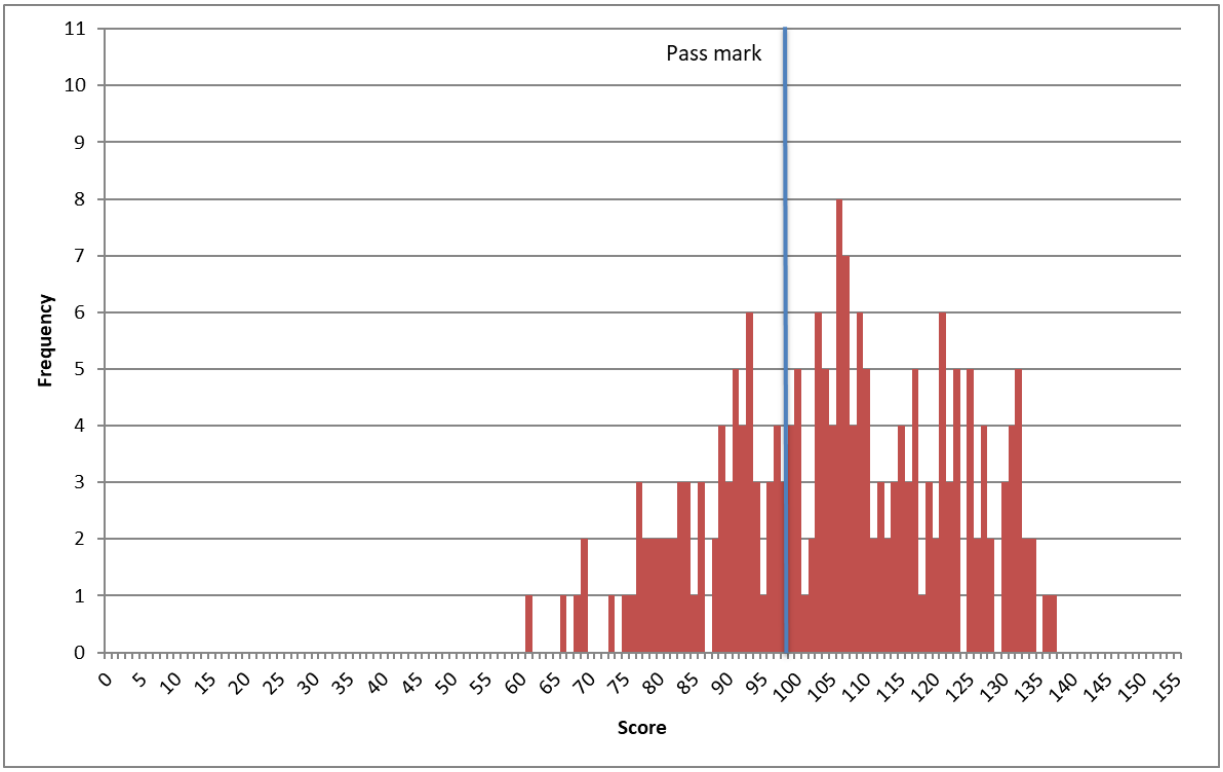
*SD = standard deviation.

Table 18: Correlations between functional area scores for all candidates

Area*	10	20	30	40	50	60	70	80	90
10		.47	.46	.49	.41	.49	.44	.40	.47
20			.62	.58	.53	.61	.51	.43	.45
30				.60	.53	.60	.54	.41	.46
40					.51	.59	.61	.51	.47
50						.52	.51	.40	.40
60							.59	.43	.45
70								.45	.38
80									.40
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 2020	October 2019	June 2019	February 2019	October 2018
Scored items	153	153	150	155	155
Candidates	144	175	128	142	178
Mean	109.1 (71.3%)	106.4 (69.5%)	101.1 (67.4%)	112.1 (72.3%)	107.2 (69.1%)
Median	110 (71.9%)	107 (69.9%)	100.5 (67.0%)	114.5 (73.9%)	109 (70.3%)
Skewness	−0.525	−0.466	0.002	−0.876	−0.462
Kurtosis[i]	−0.379	−0.057	−0.446	0.673	0.097
Range	66–137 (43.1–89.5%)	61–138 (39.9–90.2%)	60–138 (40.0–92.0%)	55–142 (35.5–91.6%)	60–141 (38.7–91.0%)
Standard deviation	16.59	15.76	16.43	18.45	15.35
Cronbach's alpha	.90	.89	.90	.92	.88
Mean r_{pb}*	.23	.22	.22	.26	.20
SEM[ii]	5.17	5.21	5.23	5.10	5.25
SEM at the pass mark	5.64	5.59	5.57	5.61	5.63
Decision consistency (uncorrected)[iii]	.90	.88	.86	.88	.89
Perceived fairness[iv]	49%	53%	42%	47%	46%
Pass mark	98.458	97.499	93.247	102.054	97.387
Effective pass mark	99	98	94	103	98
Pass rate	76.4%	74.3%	65.6%	72.5%	75.8%

[i]Excess

[ii]SEM = standard error of measurement.

[iii]Subkoviac 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 October 2019, the following exam development activities have taken place.

Validation

To provide sufficient items for upcoming administrations, a validation session was held with the EVC (see Table 20) at the HRPA offices on November 13–14, 2019. During these 2 days, CHRP ELE items were also validated, and that portion of the validation activity is reported in the technical report for that examination.

Table 20: CHRP Examination Validation Committee members – Validation

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

✓Participated in the session.

The EVC members received advance materials outlining:

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

The committee members received refresh training on the validation activity, and then worked primarily individually reviewing items to make sure they reflected current practice and were suitable to make decisions about who should receive the CHRP credential. Where committee members proposed changes, these were discussed by the group before implementation.

For each item, the committee was asked to either:

- Validate the item for use in the next 2 years to make decisions about who would be certified as a CHRP
- Move the item to the CKE 2 or CHRP ELE bank
- Revise the item to make it suitable for use
- Declare the item unsound and send it back for revision or removal from the bank

The bulk of the session saw the committee members reviewing items independently and submitting their assessments in blocks of approximately 10–20 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 committee validated 234 items as suitable for the CKE 1 and rejected 5 items. Thirty-eight items were revised prior to validation as part of this exercise. The committee also verified the functional area and competency for all items, and added rationales and references where missing, incomplete, or not current.

Appendix A

Blueprint

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 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		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
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

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.