PUBLIC RELEASE

Technical Report: February 2018 CKE 1

Human Resources Professionals Association

30 May 2018

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 165 candidates using computer-based testing at Prometric test centres from February 12 to 26, 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 best performing 150 items (while adhering to the prescribed distribution across functional areas). The mean score for first-time candidates (n=115[2]) was 106.3 (70.9%), and for all candidates it was 103.0 (68.7%). 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 October 2017 and June 2017 CKE 1 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 70.4% and a pass rate for all candidates of 64.2%.

This report, the analyses performed, and the processes followed are consistent with NCCA standards[3] and ISO 17024 standards.[4]

[1] Note: 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 a statistical outlier, 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 forms (using a combination of scored and experimental test items). Wickett selected 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. Absence of enemy items
4. Hitting the blueprint target value (+/− 3%) for each functional area
5. Maximizing spread across competencies
6. Reducing item exposure
7. Perceived psychometric effectiveness of the item, using statistics from previous administrations as available

The final form was proofed by Wickett for text errors and detection of potential enemy items. Items flagged as enemies were replaced. This form was reviewed by Ielean Tait and Jean Lazarus from the CHRP EVC on two subject matter experts in a session held at the HRPA offices on December 11, 2017. As per their input, six items were replaced and the form finalized.

The final form composition for the primary February 2018 CKE 1 is shown in Table 1. All functional areas are within two items of their targets, and as such, the two forms reflect the 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 candidates scores.

Table 1: Test form as administered

	Functional Area	Actual Items	Target	Variance
10	Strategy	7	7	—
20	Professional Practice	21	21	—
30	Organizational Effectiveness	21	21	—
40	Workforce Planning & Talent Management	21	21	—
50	Labour & Employee Relations	21	21	—
60	Total Rewards	21	21	—
70	Learning & Development	21	21	—
80	Health, Wellness, & Safe Workplace	21	21	—
90	HR Metrics, Reporting & Financial Management	21	21	—
	TOTAL	**175**	**175**	**—**

Testing Window

The examination was administered via computer-based testing at Prometric test sites primarily in Ontario. The testing window was from February 12 to 26, 2018, inclusive, and 165 candidates wrote the exam.

Candidates had access to a basic-function calculator on screen. No other aids or sources 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, the 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, 35 minutes (2 minutes longer than in October 2017). The time limit on the CKE 1 was 3½ hours, suggesting that time was not a factor in scores across candidates.

Nine (5%) of candidates took the full 3½ hours suggesting that those candidates may have wanted more time, and 2 candidates (1%) left at least one item blank suggesting those candidates timed out of the exam before being able to complete it. These metrics will continue to be monitored, and at the 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 −.10, suggesting that time constraints were not generally an issue for candidate performance. (Note that one 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 three days and the last three days was a drop of 1.3 marks out of 175.

As a matter of interest, candidate volumes were also examined across the window and 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 window rather than the start (but to a lesser degree at this administration than in the past).

Figure 1: Examination time distribution for all candidates

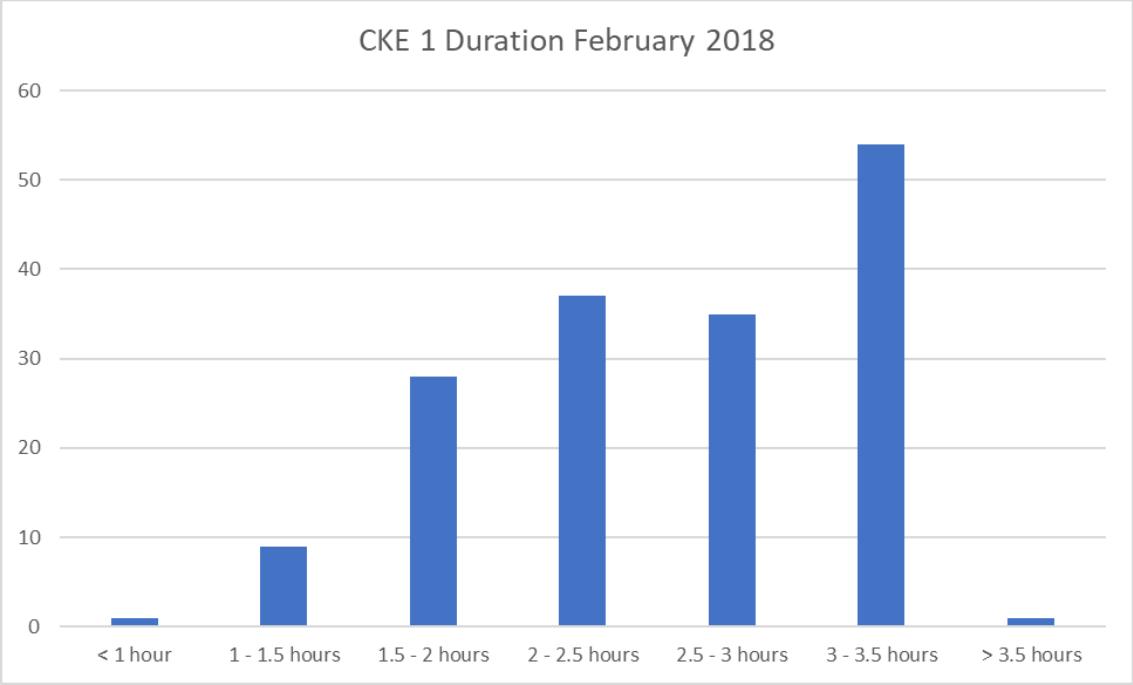
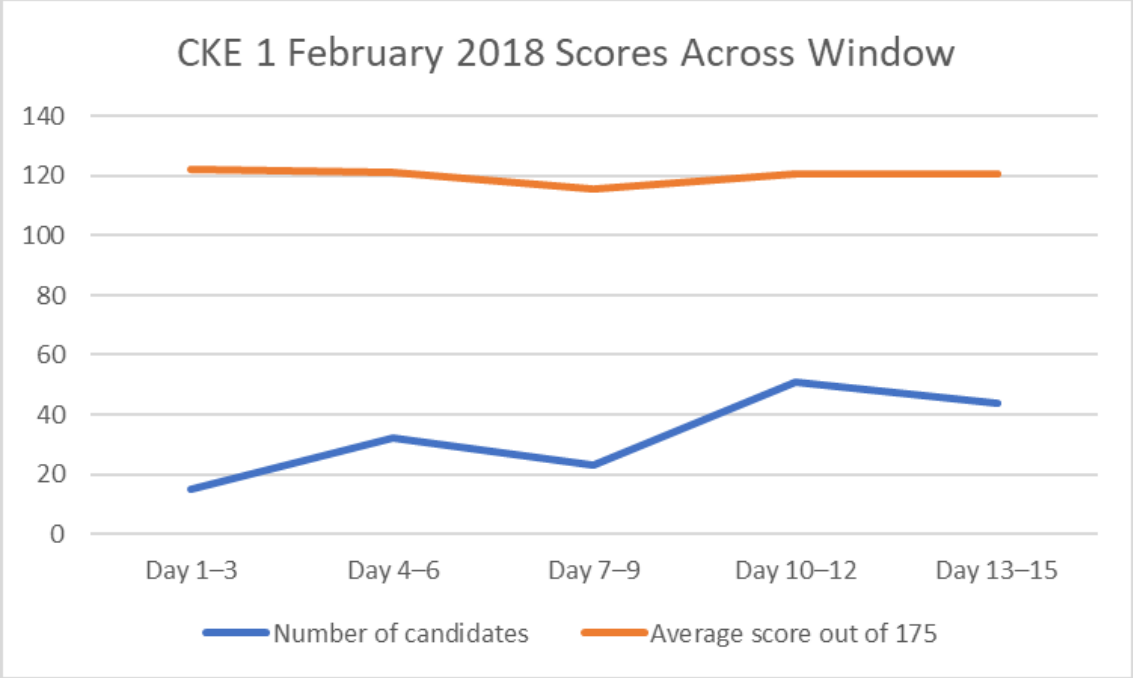


Figure 2: Candidate volume and score trends across testing window



After removing candidates who were administered a previously-used test form (who were scored using the same decisions employed at the time that form was originally used), scores were calculated for all remaining candidates based on the full set of 175 items. One candidate was flagged for an abnormally low or high score (*z* value outside +/− 3.0). Also, the 175 items

were arbitrarily broken into 7 blocks of 25 items for each candidate; the 7 resulting scores 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 and candidates with abnormal response patterns were also removed. To be conservative, candidates who had been granted a testing accommodation were also removed from the main analysis (simply because their testing conditions were not the same as the main group of candidates, and 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, five candidates were removed from scoring.

Candidates who had failed a previous HRPA examination (CKE, CKE 1, or CKE 2) scored lower than did those who had not (63.8% and 70.7%, respectively, on the full exam of 175 items). This difference was meaningful and significant ($t(110)=4.64$, $p<.001$); in keeping with standard procedures, these candidates were removed from subsequent analyses. The CKE 1 analysis proceeded with 115 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; 157 responses were obtained from candidates (response rate, 95%).

Table 2 shows the responses to the administration-related questions for CKE 1 candidates. Note that candidates were generally very positive about the administration experience. Table 3 shows the content-related questions for CKE 1 candidates. There was a tendency to more neutrality on these questions. The lower 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 essentially equally split between those who preferred computer and those who did not.

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.	72	47	12	18	8	4.1	76%
2.	I was well informed about what documents to bring to the exam location.	107	42	2	3	2	4.6	96%
3.	Proctors enforced the exam-day rules and the security procedures at the test centre were what I expected.	101	48	2	2	2	4.6	96%
4.	Proctors were professional and courteous.	104	44	6	1	0	4.7	95%
5.	The tutorial helped me understand how to complete the examination on the computer.	92	54	7	0	0	4.6	95%
6.	Navigation through the examination was easy and intuitive.	93	60	2	0	0	4.6	99%

*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.	85	53	7	9	0	4.4	90%
8.	Information available prior to exam day provided me with adequate details about the content and format of the exam.	46	52	33	16	5	4.0	64%
9.	I feel I was adequately prepared to write this examination.	15	64	52	20	2	3.8	52%
10.	The questions in the examination were clearly written.	30	61	41	21	1	3.9	59%
11.	The terminology used in the examination was accurate.	26	92	31	5	0	4.1	77%
12.	The situations presented in the examination were realistic.	27	102	21	4	0	4.1	84%
13.	The questions in the examination reflected the examination blueprint.	21	66	47	11	3	3.9	59%
14.	The examination was a fair assessment of my ability.	11	56	56	29	2	3.7	44%

*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.	32	21%
I feel that completing the examination on a computer decreased my performance.	25	16%
I feel that completing the examination on a computer had no effect on my performance.	97	63%

Initial Analysis

The full CKE 1 examination was 175 items, of which 150 were to be scored. The other 25 items were designated as experimental items. 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 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
Candidates	165
Candidates in analysis	115
Mean	123.9 (70.8%)
Range	73–155 (41.7–88.6%)
Cronbach's alpha	.91
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 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. Items were ranked from worst performing to best performing accordingly.

Key Validation

Key validation was conducted via web meeting on March 19, 2018, using the CHRP Examination Validation Committee (EVC). The EVC (Table 6) was first trained in 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

Member	Credential	Years of Relevant Experience	Year 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

*Unable to participate.

The group was informed that test reliability, as measured by Cronbach's alpha, was .909 based on the set of 175 potentially scored items and that this was above the generally accepted threshold of .80. They were also informed that part of the goal of the key validation review was to bring this value up if possible.

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 less direction on those with borderline statistics. Where available, any comments made by candidates to the items were also shown. The group made the decision based on content and the data through discussion; they removed the 25 items that they felt were least appropriate to retain for scoring. Because of the modest sample size for this administration, past item data were also used where available, and the group was directed not to place undue reliance on statistics exclusively from the February administration. Comments made by the panel members 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. However, most were moderate to

strong items. The final alpha for the set of 150 scored items was .915. The difficulties ranged from 33.0% to 93.9%, with a mean of 70.9%. The r_{pb}* values ranged from .01 to .53, with a mean of .25.

Table 7 presents the scored CKE 1's final fit to the examination blueprint. In all cases, the final number of scored items within a functional area fit within the established range.

The group endorsed the final set of items for use in scoring the February 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**	5	5	6	7	4% ± 1%
20	**Professional Practice**	17	14	18	22	12% ± 3%
30	**Organizational Effectiveness**	18	14	18	22	12% ± 3%
40	**Workforce Planning & Talent Management**	19	14	18	22	12% ± 3%
50	**Labour & Employee Relations**	20	14	18	22	12% ± 3%
60	**Total Rewards**	17	14	18	22	12% ± 3%
70	**Learning & Development**	19	14	18	22	12% ± 3%
80	**Health, Wellness & Safe Workplace**	17	14	18	22	12% ± 3%
90	**HR Metrics, Reporting & Financial Management**	18	14	18	22	12% ± 3%
Total		**150**				

Establishing the Pass Mark: Equating

Equating, as per Kolen and Brennan (2014),[5] was used to establish the pass mark for the February 2018 CKE 1. The goal of this process was to set a pass mark for the February 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 the Appendix. Specific information on the standard-setting session is provided in the Technical Report issued for the November 2015 administration.

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

Two equating procedures were conducted back to different administrations (June 2017 and October 2017). The intention following these two equating runs is to average them to arrive at a final pass mark for the February 2018 CKE 1. These administrations were chosen as the most recent administration and the administration corresponding roughly to the same administration month the previous year. The February 2017 administration was not used in this procedure because of the very low number of candidates sitting that administration.

Equating back to the October 2017 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 three 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 .28 (for February candidates).

Table 8 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 October 2017

Area	Actual	Target
10	3%	4%
20	14%	12%
30	8%	12%
40	8%	12%
50	14%	12%
60	14%	12%
70	14%	12%
80	14%	12%
90	14%	12%

Mean, Tucker, Levine observed-score, and Braun-Holland methods were computed to ascertain concordance of solutions. Given the sample sizes and similarities of test parameters, Tucker was considered the optimal method.

Table 9 shows some of the parameters used to derive the equating estimates, along with other parameters describing the test forms. Of note is that on the anchor items, the population taking the February 2018 CKE 1 scored marginally higher than the population taking the October 2017 CKE 1 (70.6% vs. 68.5%, respectively; $t(295)=1.22$, *ns*). Though not statistically significant, because the February 2018 CKE 1 candidates tended towards higher ability (based on the anchors), they would likely have a higher pass rate than seen in October.

The equating analysis bears this out (Table 10). The Tucker, Levine observed, and Braun-Holland methods indicate an integer pass mark of 98, and the mean method estimates it at 99. The Tucker value of 97.57 was extracted from this analysis for use in setting the final pass mark.

Table 9: Equating parameter table – to October 2017

		2017	2018
		October	February
	n	182	115
	Scored items	150	150
Mean score	Total	69.5%	70.9%
Mean score	Anchors	68.5%	70.6%

Table 10: Equating outcome table – to October 2017

	Pass Mark		Pass Rate	
Method	**Precise**	**Integer**	**All**	**First Time**
Combo October 2017	**98.39**	**99**	**59.2%**	**66.5%**
Tucker	97.57	98	*64.2%*	*70.4%*
Levine observed	97.13	98	*64.2%*	*70.4%*
Mean	98.15	99	*61.8%*	*69.6%*
Braun-Holland	97.73	98	*64.2%*	*70.4%*

Equating back to the June 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 three 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 .27 (for February candidates).

Table 11 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 11: Anchor item fit to blueprint – to June 2017

Area	Actual	Target
10	5%	4%
20	9%	12%
30	14%	12%
40	11%	12%
50	14%	12%
60	11%	12%
70	9%	12%
80	14%	12%
90	14%	12%

Mean, Tucker, Levine observed-score, and Braun-Holland methods were computed to ascertain concordance of solutions. Given the sample sizes and similarities of 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 2018 CKE 1 scored marginally lower than the population taking the June 2017 CKE 1 (70.8% vs. 72.0%, respectively; $t(260)=0.75$, *ns*). Because the February 2018 CKE 1 candidates were marginally lower in ability (based on the anchors), it is expected to have a slightly lower pass rate as compared June.

The equating analysis bears this out (Table 13). The Levine observed, Mean, and Braun-Holland methods all indicate an integer pass mark of 99, and the Tucker method indicates an integer pass mark of 98. The Tucker value of 97.849 was extracted from this analysis for use in setting the final pass mark.

Table 12: Equating parameter table – to June 2017

		2017	2018
		June	February
	n	147	115
	Scored items	150	150
Mean score	Total	69.2%	70.9%
Mean score	Anchors	72.0%	70.8%

Table 13: Equating outcome table – to June 2017

	Pass Mark		Pass Rate	
Method	**Precise**	**Integer**	**All**	**First Time**
Combo June 2017	**94.50**	**95**	**67.8%**	**75.5%**
Tucker	97.85	98	*64.2%*	*70.4%*
Levine observed	98.25	99	*61.8%*	*69.6%*
Mean	98.44	99	*61.8%*	*69.6%*
Braun-Holland	98.34	99	*61.8%*	*69.6%*

Combined Results

Table 14 shows the pass mark values across the two equating runs. The value in green highlight is the one that would be selected based on population parameters at each equating run. Though different weighting and averaging methods were considered, all resulted in a value between 97 and 98, and so the simple arithmetic mean (97.709924) of the two identified values was the recommended pass mark for the February 2018 CKE 1.

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 70.4% for first-time candidates is higher than that seen in October 2017 and lower than that seen in June 2017, which was predicted by the differences in performance on the anchors described above. The pass rate for all candidates was 64.2%. See Table 15 for historic pass rates.

The final pass mark value, and the process used to derive it, was presented to the same panel used for key validation (Table 6) via teleconference on March 21, 2018. The panel formally approved the pass mark (which was presented along with the consequent pass rate data) for recommendation to HRPA. HRPA subsequently accepted the recommendation from the panel and the pass mark was formally established.

Table 14: Equating outcome table – Combined results

	Oct17	Jun17
Tucker	97.6	97.8
Levine observed	97.1	98.3
Mean	98.2	98.4
Braun-Holland	97.7	98.3

Table 15: Historic pass rates

	Pass rate	
	All	1st time
Jun16	65.4%	69.1%
Nov16	58.8%	62.1%
Feb17	50.5%	62.5%
Jun17	67.8%	75.5%
Oct17	59.2%	66.5%
Feb18	**64.2%**	**70.4%**

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 nine functional areas were also computed for each candidate. An Excel file with the final candidate results was provided to HRPA.

Table 16 provides the means and standard deviations for the functional areas and for the total score, using all candidates who took the new February 2018 CKE 1 form. Table 17 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 16: Total and functional area scores for all candidates

Functional Area		Percentage	Mean	SD*
10	**Strategy**	60%	3.0	1.1
20	**Professional Practice**	71%	12.0	2.5
30	**Organizational Effectiveness**	67%	12.1	2.7
40	**Workforce Planning & Talent Management**	70%	13.4	2.5
50	**Labour & Employee Relations**	66%	13.3	2.9
60	**Total Rewards**	67%	11.3	2.9
70	**Learning & Development**	68%	12.9	3.0
80	**Health, Wellness & Safe Workplace**	74%	12.6	2.4
90	**HR Metrics, Reporting & Financial Management**	69%	12.5	2.7
Total score		68.7%	103.0	17.0

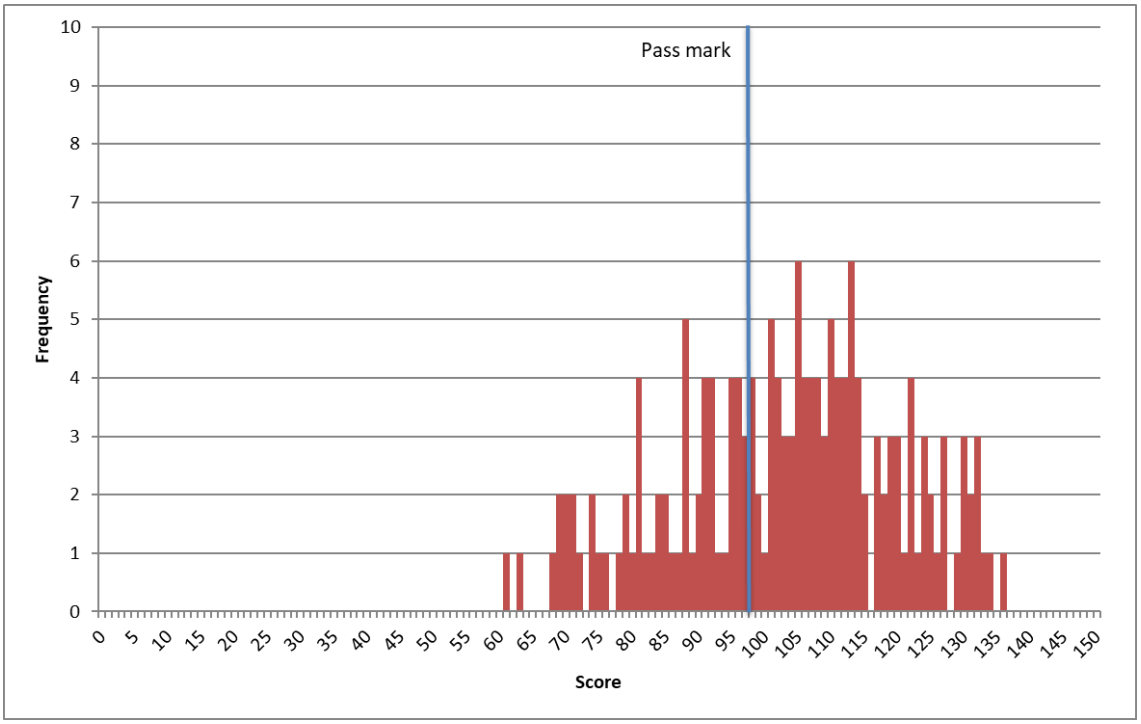
*SD = Standard deviation.

Table 17: Correlations between functional area scores for all candidates

Area*	10	20	30	40	50	60	70	80	90
10		.40	.46	.38	.30	.51	.38	.34	.42
20			.50	.58	.52	.58	.44	.52	.56
30				.52	.50	.57	.59	.44	.37
40					.54	.61	.57	.44	.51
50						.59	.48	.50	.52
60							.54	.61	.46
70								.45	.51
80									.45
90									

*See Table 16 for the full name of each functional area.

Figure 3: Score distribution for all candidates



Key Examination Metrics

Table 18 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.

Note that February 2017 was the first computer-based testing delivery of the CKE 1, but this was not considered material to data analysis or interpretation.

Table 18: Key examination metrics – Candidates included in analysis only

Index	February 2018	October 2017	June 2017	February 2017	November 2016
Scored items	150	150	150	150	150
Candidates	115	182	147	48	322
Mean	106.3 (70.9%)	104.3 (69.5%)	103.8 (69.2%)	100.5 (67.0%)	99.9 (66.6%)
Median	109 (72.7%)	106 (70.7%)	104 (69.3%)	101 (67.3%)	100 (66.7%)
Skewness	−0.483	−0.489	−0.230	−0.335	−0.141
Kurtosis[i]	−0.525	0.024	−0.578	−0.054	−0.656
Range	61–136 (40.7–90.7%)	57–134 (38.0–89.3%)	66–136 (44.0–90.7%)	56–131 (37.3–87.3%)	58–134 (38.7–89.3%)
Standard deviation	17.60	15.81	15.96	15.93	16.51
Cronbach's alpha	.92	.90	.89	.89	.90
Mean r_{pb}*	.25	.22	.22	.22	.23
SEM[ii]	5.13	5.11	5.19	5.24	5.22
SEM at the pass mark	5.56	5.43	5.58	5.51	5.50
Decision consistency (uncorrected)[iii]	.92	.87	.86	.87	.86
Perceived fairness[iv]	43%	49%	36%	42%	n/a
Pass mark	97.710	98.387	94.496	95.236	94.179
Effective pass mark	98	99	95	96	95
Pass rate	70.4%	66.5%	75.5%	62.5%	62.1%

[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 October 2017, the following exam development activities have taken place.

Item Writing

To fill gaps in the bank and renew content, item writing was conducted in October–December 2017. Item writers (see Table 19) were identified by HRPA and trained in a remote session by Wickett on October 30, 2017.

Table 19: Item writers

Writer	Credentials	Years of Relevant Experience	Industry
Cal Barber	MBA and CHRP	15+	Education – Professor at Seneca College
Carolyn Capretta	MBA and CHRL	8+	Education – Instructor at McMaster University
Dorothy Milardovic	Master's Degree in HR Management and CHRL	8+	Education – Coordinator of the Post Diploma HR Management Program at Conestoga College

The item writers were provided with training via teleconference, and received printable files covering the main elements of the training. The general guidance for writing quality multiple choice items was drawn primarily from Haladyna & Rodriguez (2013).[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 to 45 items to write, for a total of 90 items (one additional item was written in this process, and so 91 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 October–December 2017 there was need for group review of those items before moving them to formal validation and use on the CKE 1. The group had 99 items for consideration (taken predominately from the newly written items, supplemented with other unreviewed items in the bank required to fill gaps in the bank).

The review session was held January 15–17, 2018 at HRPA offices. The panel members who participated are shown in Table 20. This session involved the review of ELE items as well.

Table 20: Panel for item review session

Panelist	Credentials	Years of Relevant Experience	Industry
Laurence Frederick	CHRL	10–15	Consultant/Education
Annette Lawrence	CHRP	1–5	Legal services
Surbhi Sud	CHRL	10–14	Health and social services
Brittany Seguin	CHRP	1–5	Utilities
Sanjeev Pruthi	CHRP	10–14	Mid level management

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 83 items as suitable for CKE 1, moved 7 items to the CKE 2 bank, moved 1 item to the ELE bank, and rejected 8 items. Of the approved items, 28 saw text changes to the stem and/or options before approval.

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

Appendix

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.