

# How the National Exams are Developed and Scored

Claude Balthazard, C.H.R.P.  
Director HR Excellence and Registrar  
Human Resources Professionals Association  
June 2010



# Housekeeping

- Slides, archived webinar, and Q&A to be posted on HRPA website on the Office of the Registrar page within a week
- Time has been set aside for questions at the end of the webcast but you can submit your questions at any time
- All questions and their answers will be posted on the HRPA website

# Today's Webinar

- Who does what with respect to the exams
- Why does HRPAs delegate the development and scoring of the exams to CCHRA?
- Test development process
- Practice analysis
- HR Body of Knowledge and the RPCs®
- Test specifications
- Item writing, item review, and test assembly
- Standard setting and Angoff panels
- Fixed and floating cut-scores
- Post-exam review and final scoring of exams
- Manual rescores

# Two Audiences for Today's Webinar

- Those who wrote the exam in May 2010 and want to better understand how their exam was developed and scored
- Those who plan to write the NKE in October 2010 looking for some information that may help in passing the exam

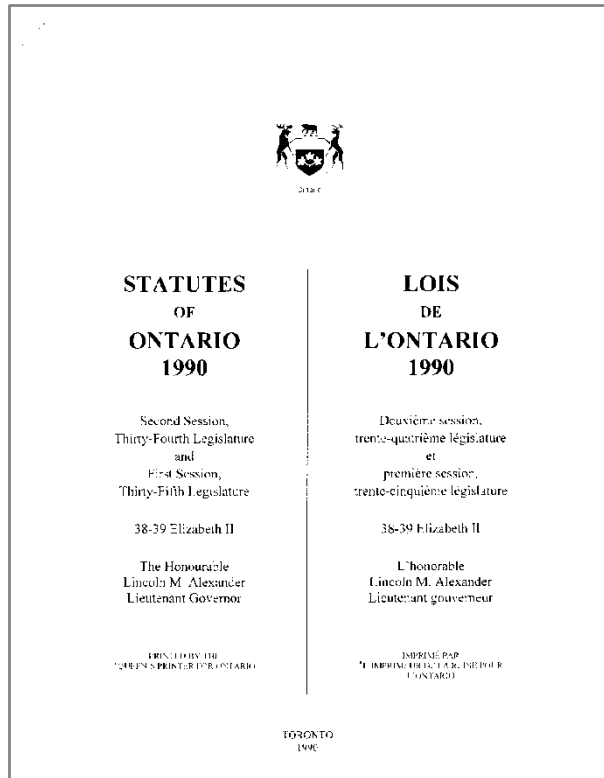
# Exam Results

- The results for the May 1 exams, for both the NKE and the NPPA, will be released at 12:00 noon on Friday, July 2, 2010
- Instructions as to how and where to find the results have been sent by email to all exam-writers; they are also posted on the HRPA web site
- A follow up email, with more detail will be sent within a week after that

# Brief History of Exam Requirement at HRPA

- HRPA introduced an exam requirement in 1994.
- The concern at the time was that there appeared to be significant differences in the quality of academic preparation across academic institutions despite the fact that that all courses met the same standards with respect to course content
- This exam was called the Comprehensive Provincial Exam (CPE).
- In 2003, the CPE was replaced by the National Knowledge Exam (NKE) and in 2004 the assessment of experience was replaced by the National Professional Practice Assessment (NPPA)
- Although both the NKE and the NPPA are exams, in Ontario, the NPPA was more aligned with the experience requirement (a virtual experience requirement)
- In May 2009, HRPA announced that it would return to an actual experience requirement that May 2010 would be the last administration of the NPPA in Ontario

# Statutory Authorities in Regards to the Exam Requirement



- Section 2(c) of the *Act* gives HRPAs the statutory authority to “*hold examinations and prescribe tests of competency deemed appropriate to qualify membership in and certification by the Association.*”
- Section 2(a) of the *Act* gives HRPAs the statutory authority to “*establish uniform province-wide standards of knowledge, experience and ethics for all persons engaged in the field of human resources management.*”

# Delegation to CCHRA

- In Ontario, the exams are prescribed pursuant to statutory authorities delegated to HRPAs
- HRPAs delegate to CCHRA the task of developing and scoring the exams
- Delegation is not a transfer of authority or an abdication of responsibility, HRPAs remain fully responsible and accountable for the exams that are developed and scored by CCHRA on its behalf

# Delegation to CCHRA

- Why does HRPA delegate the development and scoring of the exams to CCHRA?
  1. To manage possible conflicts of interest
  2. To benefit from economies of scale
  3. To facilitate mutual recognition of designation

# Reason #1: Conflict of Interest

- Professions are self-regulated
- The standard is supposed to be set at the appropriate level to protect the public interest but in order to increase membership a regulator may be tempted to lower standards; on the other hand, to increase the demand for certified professionals a regulator may be tempted to restrict supply by raising the standards
- To reduce the possibility of manipulation, and to increase public confidence in the process, standard setting is delegated to an arms-length body

# Reason #2: Economies of Scale

- Developing exams is an expensive proposition
- It makes sense to amortize the cost of developing and scoring exams by joining with other associations that have the same need

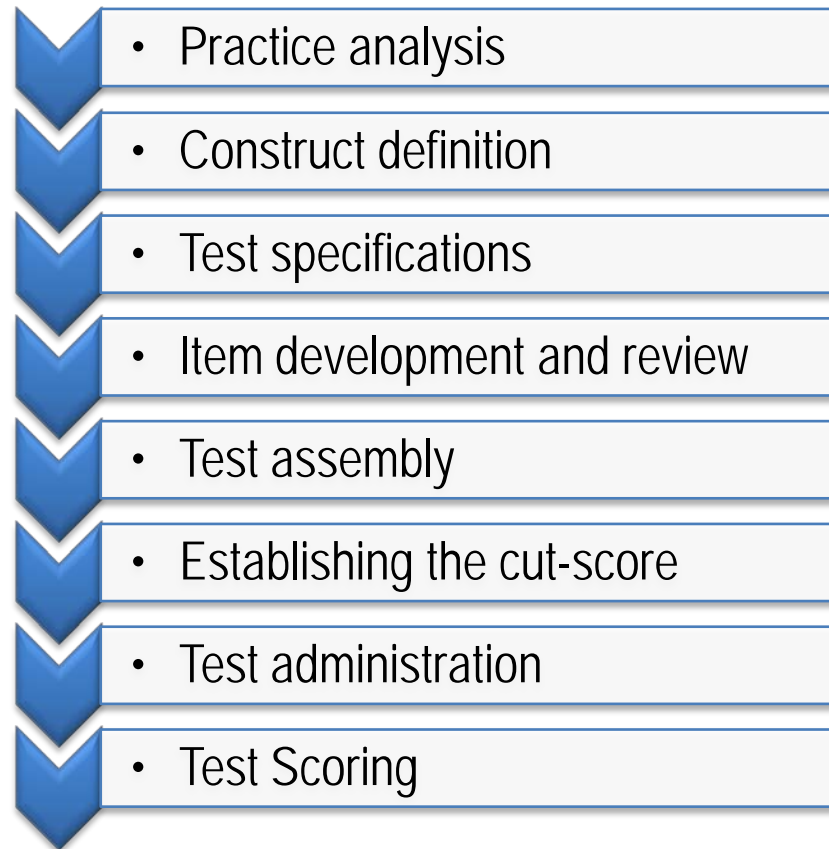
# Reason #3: Facilitates Mutual Recognition

- There is nothing that requires different provincial associations to use the same exam in order to make the designation transportable
- A national standard means comparable competence not identical process
- On the other hand, using the same exam does make comparability easier to establish

# The Independent Board of Examiners

- Within CCHRA, the Independent Board of Examiners (IBE) is responsible for the development and scoring of the exams
- The IBE consists of seven prominent academics in various areas of HR from across Canada

# Test Development Process



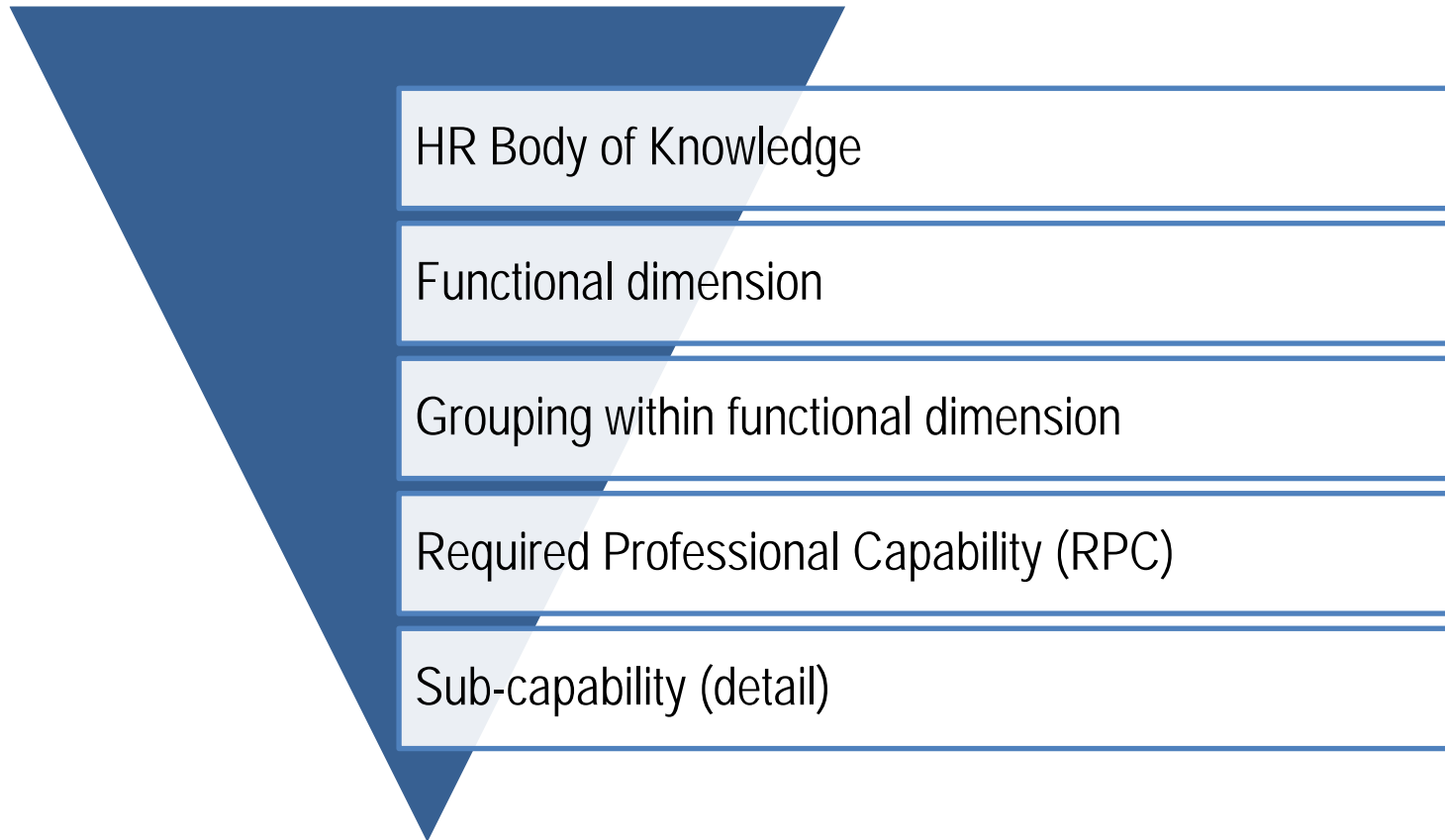
# Practice Analysis

- A formal study of an occupation, the tasks accomplished as well as the knowledge, skills, abilities, and other characteristics needed to perform in that occupation
- Original practice analysis conducted in 1998, updated in 2007
- Optimally, practice analyses are updated every five years

# Output of Practice Analysis

- The output of a practice analysis is a framework that describes what competent members of a profession must know and/or be able to do
- In HR, this framework is called '*The HR Body of Knowledge*' (although many refer to the framework as 'the RPCs<sup>®</sup>')

# The HR Body of Knowledge



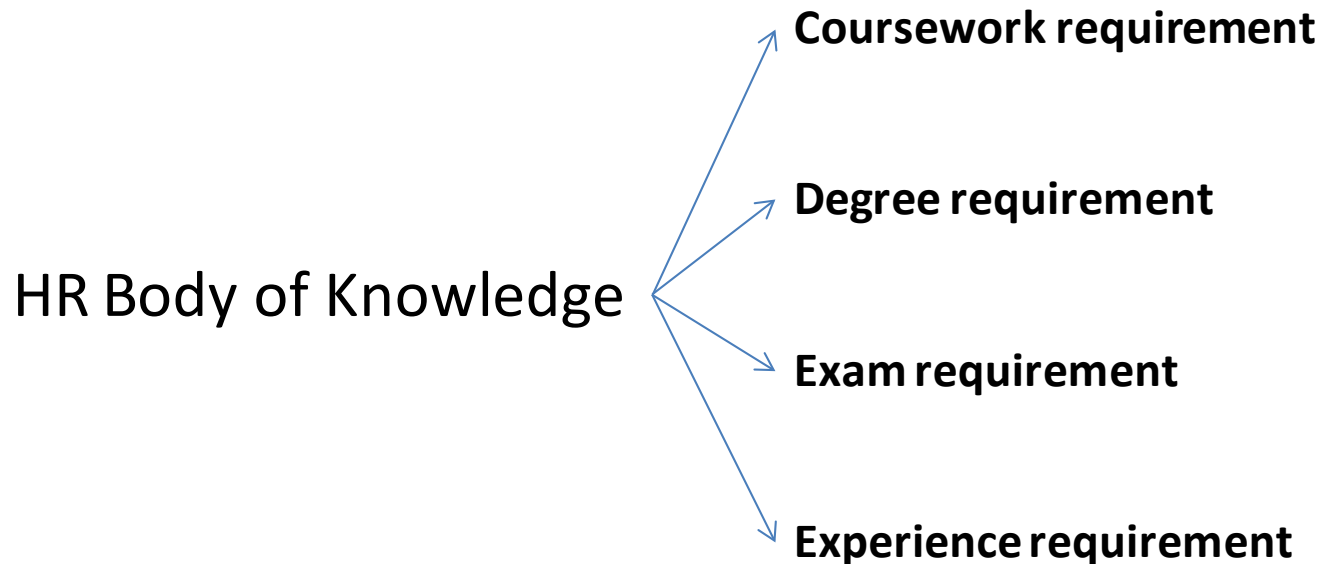
# The HR Body of Knowledge

- The two most important levels in the hierarchy are the functional domains and the Required Professional Capabilities (RPCs<sup>®</sup>)
- The functional domains are important because this level is used to define the test specifications for the exams
- The Required Professional Capabilities (RPCs<sup>®</sup>) are important because they represent the most useful level to describe the individual competencies that every professional in HR should be able to understand and apply in the workplace
- The RPCs<sup>®</sup> are used to develop exam questions

# The Seven Functional Domains

- Professional Practice in HR
- Organizational Effectiveness
- Staffing
- Employee and Labour Relations
- Total Compensation
- Organizational Learning, Training & Development
- Occupational Health & Safety

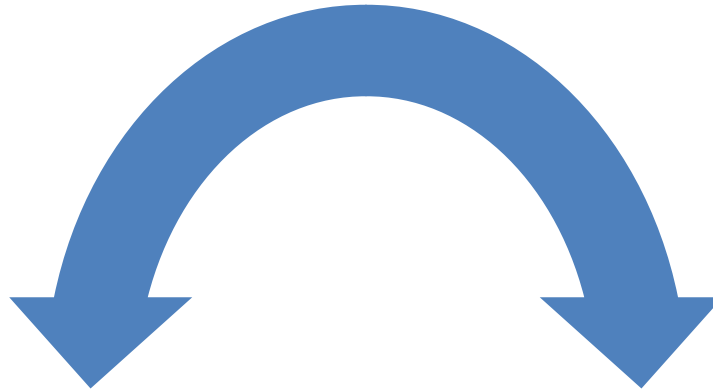
# All Requirements Derive from the HR Body of Knowledge



# Construct Definition

- To be precise, an exam is not a requirement, it is a method of assessment
- The real requirement is what the exam measures—we call these ‘constructs’
- The exam is just a way of getting at a construct
- The two exams—the NKE and the NPPA—although both exams, actually tap different ‘constructs’

# Construct Definition



**Academic knowledge** refers to basic facts, policies, practices, methods, legislation, etc. It is information that can be written into procedures and transferred fairly accurately during the learning process.

**Experiential knowledge** refers to the application of academic knowledge to relevant problems, challenges, and situations. It is expected that this experiential knowledge is gained through experience.

# Sorting the RPCs®

- The RPCs® were sorted into two categories those that were best understood as reflecting academic knowledge and those that were best understood as reflecting experiential knowledge
- The sorting of the RPCs® was accomplished using a rating included in the practice analysis

# Sorting the RPCs®

*What level of proficiency is needed to successfully carry out the RPC?*

Rating	Anchor	Description
1.0	<i>Don't Know</i>	Respondent had no knowledge or experience with the RPC™
2.0	<i>Entry-Level Professional</i>	The RPC™ could be performed by new entrants to the HR field with no experience
3.0	<i>Developing Professional</i>	The RPC™ could be performed by developing professionals who could be expected to master the RPC™ through on-the-job experience or developmental/training programs
4.0	<i>Experienced Professional</i>	The RPC™ could be performed by professionals who had more training and experience

RPCs® with proficiency ratings of 3.22 or greater were deemed experiential RPCs® whereas those with proficiency ratings of 3.21 or less were deemed academic RPCs®.

# Test Specifications

Content Domain	Total Number of RPCs®	Academic Knowledge		Experiential Knowledge	
		Number of RPCs®	Percent of RPCs®	Number of RPCs®	Percent of RPCs®
Professional Practice	42	24	25.0	18	20.0
Organizational Effectiveness	22	8	8.3	14	15.6
Staffing	30	18	18.8	12	13.3
Employee & Labour Relations	25	10	10.4	15	16.7
Total Compensation	25	10	10.4	15	16.7
Org. Learning & Training	26	18	18.8	8	8.9
Occupational Health & Safety	17	8	8.3	9	10.0
<b>Total</b>	<b>187</b>	<b>96</b>	<b>100.0</b>	<b>91</b>	<b>100.0</b>

# NKE and NPPA

## NKE

- Assesses academic knowledge
- Conventional knowledge test
- Four-option multiple-choice
- 150 items

## NPPA

- Assesses experiential knowledge
- Situational judgment test (SJT)
- Four-option multiple-choice
- 60 items (plus 5 un-scored items for pre-testing)

# Item Writing

## NKE

- Academics in the field of Human Resources Management (HRM) at Canadian universities and colleges are enlisted to write questions pertaining to their area of specialization using the knowledge RPCs® as starting point
- Items reviewed for clarity, grammar, and spelling by professional editor with HR editing experience
- Final review by IBE

## NPPA

- NPPA item writing panel identifies critical incidents
- The critical incidents become the scenarios
- Different responses to the scenario are written
- One option is identified by the panel as the best response to the scenario
- Items reviewed for clarity, grammar, and spelling by professional editor with HR editing experience
- Final review by IBE

# Translation

- Items are translated (usually from English into French) by a professional translator familiar with HR terminology and usage

# Test Assembly

- From among the item pools, and respecting the test specifications, items are selected to make up a given form of the test
- Tests are printed
- Tests are shipped to the various provincial HR associations

# Standard Setting

- Standard setting refers to setting the cut-score
- For licensure and certification tests, the standard is absolute and not relative
- Exam-writers do not compete against each other; there is no pass or fail quota. There is no targeted pass rate
- If a particular group of exam-writers is strong, more exam-writers would pass the exam; if a particular group of exam-writers was weak, fewer exam-writers would pass the exam
- The cut-score should represent the same level of proficiency and competence regardless of the group tested

# Angoff Panels

- Starting with the October 2009 exam, CCHRA started using what are known as ‘Angoff panels’ to set the cut-scores for the exams
- An Angoff panel is a method which makes use of the combined judgment of panel members to establish the probability that a candidate at the threshold of competence would be able to answer the question correctly
- Angoff panels are made up of HR professionals who have a good handle on what is required of HR professionals
- Presently, Angoff panels are convened by CCHRA after the tests are assembled, but they can be conducted even before the tests are assembled

# Angoff Panel Data

	Angoff Panel Judge					Across Judges	
	1	2	3	4	5	Average	Standard Deviation
Question 1	.75	.75	.80	.65	.70	<b>.73</b>	<b>.057</b>
Question 2	.65	.70	.75	.65	.80	<b>.71</b>	<b>.065</b>
Question 3	.70	.65	.60	.65	.65	<b>.65</b>	<b>.035</b>
Question 4	.65	.75	.65	.70	.60	<b>.67</b>	<b>.057</b>
Question 5	.55	.50	.45	.65	.55	<b>.54</b>	<b>.074</b>
---	---	---	---	---	---	---	---
Question 146	.80	.80	.80	.70	.60	<b>.74</b>	<b>.089</b>
Question 147	.80	.75	.70	.55	.65	<b>.69</b>	<b>.096</b>
Question 148	.55	.60	.65	.65	.45	<b>.58</b>	<b>.084</b>
Question 149	.65	.65	.70	.75	.65	<b>.68</b>	<b>.045</b>
Question 150	.65	.70	.65	.65	.55	<b>.64</b>	<b>.055</b>
Passing score	<b>101.25</b>	<b>102.75</b>	<b>101.25</b>	<b>99</b>	<b>93</b>	<b>99.45</b>	<b>3.846</b>

# Fixed and Floating Cut-Scores

- The cut-score is determined before the test is administered or scored
- Because the cut-score for the test is derived by adding the probabilities for each question, the cut-score will vary depending on the particular set of items that make up the test
- Each version of the exams will have its own cut-score
- An exam that is made up of somewhat more difficult questions will have a somewhat lower cut-score; an exam that is made up of somewhat easier questions will have a somewhat higher cut-score

# Fixed and Floating Cut-Scores

- In the past, the cut-score for the exam was set at 70%, and adjustments were made to the scores to make this work
- Advice from experts in psychometrics was that it was better to leave the scores alone but to set the cut-score accordingly
- Going forward, instead of adjusting the scores, the cut-score will be allowed to vary based on the relative difficulty of the exam based on the Angoff panel ratings
- Allowing the cut-score to vary obviates the need to make adjustments

# Floating Cut-Scores vs. Adjustments

- In October 2009, in order to keep the cut-score at a fixed 70%, the adjustment was +14 for the NKE and +3 for the NPPA
- This is equivalent to a cut-score of 91 out of 150 or 60.6% for the NKE and 39 out of 60 or 65.0% for the NPPA
- Going forward, instead of adjusting the scores, the cut-score will be allowed to vary based on the Angoff panel ratings

# Switching to Floating Cut-Scores

- The switch to a floating cut-score
  - makes no difference in terms of how one would prepare for the exam
  - makes no difference in terms of test-taking strategy
  - makes no difference in terms of whether one passes or fails the exam
- The switch to a floating cut-score makes the scoring more transparent and is consistent with best practices in testing

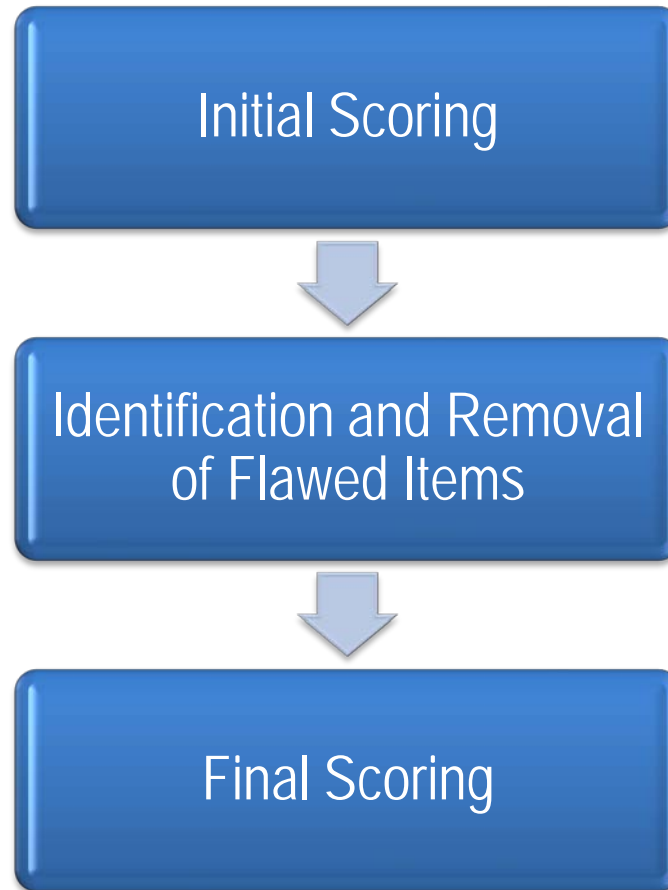
# Scanning the Answer Sheets

- The Scantron® sheets are scanned
- The Scantron® reader flag response sheets with missing (omits) or multiple responses
- These response sheets are reviewed individually

# Scoring the Exams

- Scoring the exams is a multi-step process
- There is an initial scoring of the exam based on all the administered items. The purpose of this initial scoring is to identify flawed items
- There is a post-exam review where items that may be flawed are reviewed to determine whether they are truly flawed or not
- There is a final scoring where the scores and the cut-score are recalculated based on retained items. The purpose of this final scoring is to assign a score to exam-writers

# Scoring the Exams



# Post-Exam Review

- Despite the care and attention that is given in developing items, some items can fail to perform as expected
- In a post-exam review, the performance of each item is examined statistically by conducting what is called an ‘item analysis’
- Item analysis identifies ‘non-performing’ items
- Non-performing items are usually flawed items
- Flawed items reduce the validity and reliability of the exams and should be discarded

# Example of a Flawed Item

Which of the following cities is the capital of Ontario?

- a. Montreal
- b. Ottawa
- c. Saskatoon
- d. Edmonton

# What is an Ambiguous Item?

- It is not uncommon for exam-writers to feel that some of the items were ambiguous
- Psychometricians have ways of identifying ‘non-performing’ or flawed items
- The statistics calculated for each item include difficulty and discrimination indices for each option
- In the context of item analysis, the discrimination index refers to the statistical correlation between performance on the item and performance on the exam as a whole
- In addition, difficulty indices are calculated for candidates at varying levels of overall exam performance, as well as for each linguistic version of the exam
- The pattern of responses across incorrect options (‘distractors’) is also examined

# Final Scoring

- Items deemed ‘flawed’ are removed
- The scores are re-calculated based on the retained items
- The cut-score is also recalculated based on the Angoff panel ratings for the retained items
- The final score and the final cut-score may be based on less than the full number of items
- This represents the ‘definitive’ scoring of the exam

# Rounding

- The scores, and the cut-score are reported in terms of percentages
- Final scores at or above the final cut-score are ‘passes’; final scores below the final cut-score are ‘fails’
- We want to avoid having two candidates with the same percentage score where one passes while the other fails
- Therefore, if two raw scores equal the same rounded percentage cut score, we use the lower of the two

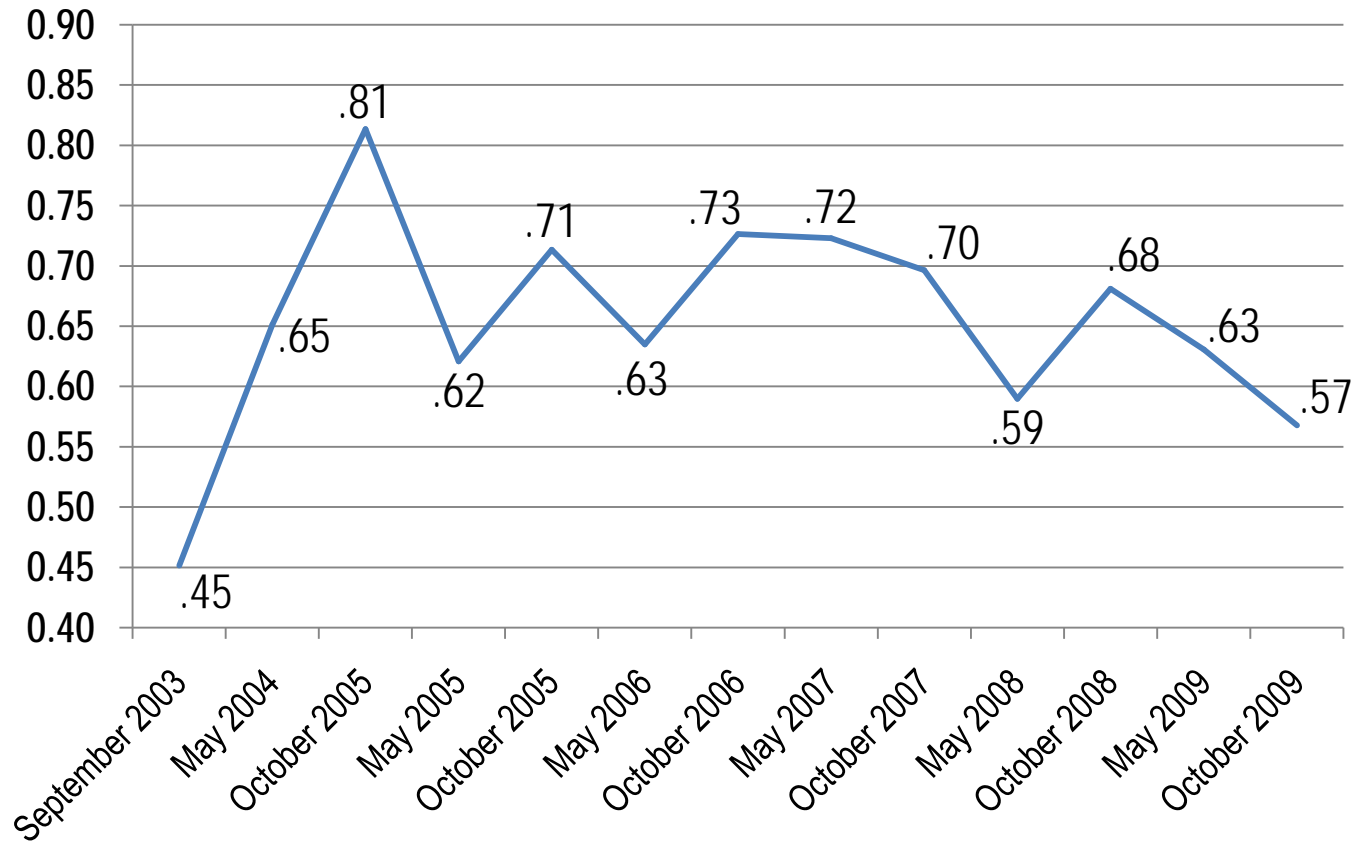
# Manual Rescores

- Despite the quality control procedures, it remains possible that a error could have been made in the scoring of an exam
- Upon request and payment of a fee, CCHRA will conduct a manual rescore of the exam
- A manual rescore is just that, starting with the Scantron<sup>®</sup> sheet, the exam is rescored by hand
- The CCHRA fee for manual rescoring is \$25 for the NKE and \$75 for the NPPA
- For HRPAs exam-writers, all requests for manual rescoring must be made to HRPAs

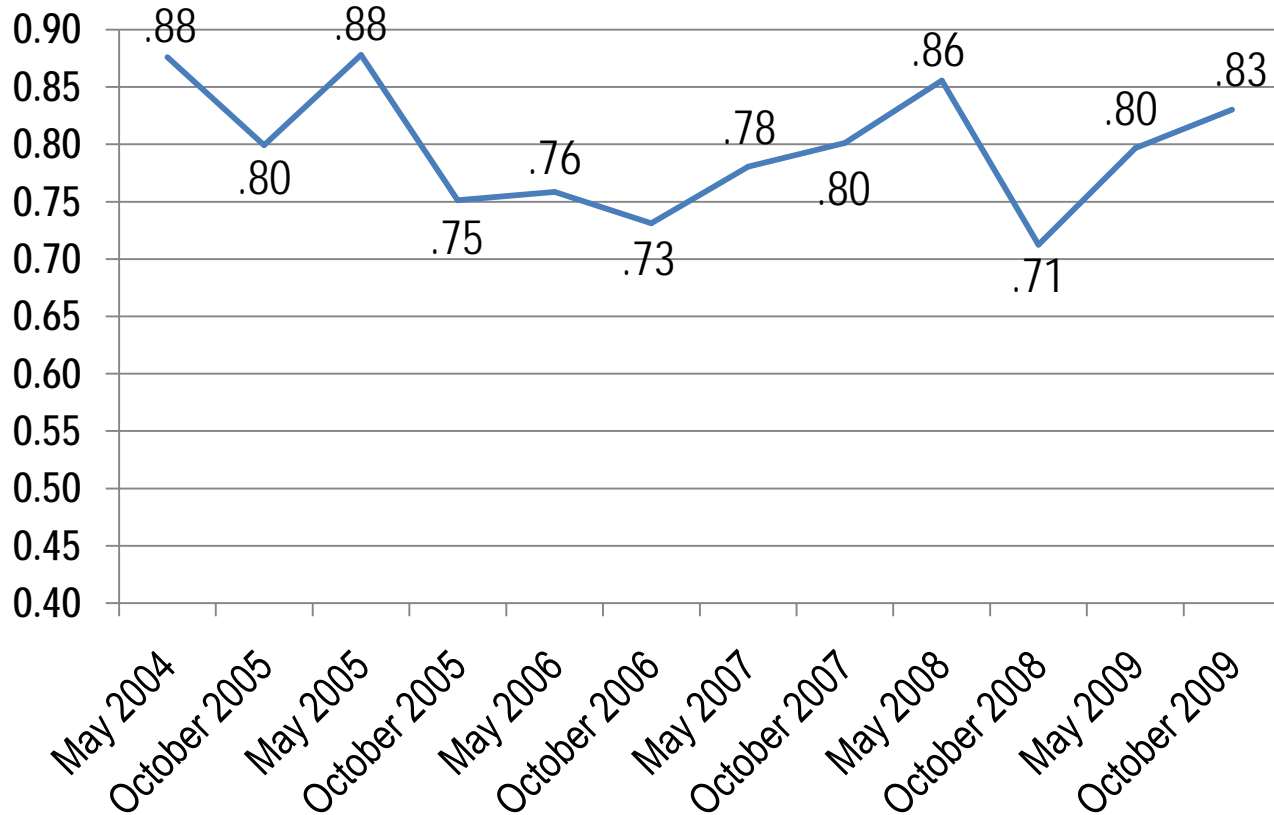
# Discontinuation of the NPPA

- May 2010 was the last administration of the NPPA in Ontario
- The NPPA was replaced with the validation of experience
- Both the NPPA and the validation of experience tap experiential knowledge construct as defined earlier

# NKE Pass Rate



# NPPA Pass Rate



# NCCA Accreditation

- In all its exam processes, HRPA strives to meet or surpass NCCA accreditation standards
- For more on NCCA accreditation standards  
<http://www.credentialingexcellence.org/AccreditationServices/CertificationAccreditation/StandardsInterpretations/tabid/93/Default.aspx>

# Technical Terms

- Practice analysis
- Construct
- Test specifications
- Test assembly
- Standard setting
- Angoff panels
- Fixed and floating cut-scores
- Initial and final scoring
- Post-exam review
- Flawed items
- Item analysis
- Item discrimination
- Item difficulty
- Distractor analysis

# Interested in Knowing More?

- *Various. (2009). Certification: The ICE Handbook* J. Knapp, L. Anderson, and C. Wild (Eds.), The Institute for Credentialing Excellence, Washington, DC.
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Sage: Thousand Oaks, CA.
- Zieky, M. J., Perie, M., & Livingston, S. A. (2008). *Cutscores: A manual for setting standards of performance on educational and occupational tests*. Educational Testing Service: Princeton, NJ.
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. (3<sup>rd</sup> Ed.). Lawrence Erlbaum Associates: Mahwah, NJ.
- National Commission for Certifying Agencies (2004). *Standards for the accreditation of certification programs*. Author: Washington, DC.



Human Resources  
Professionals  
Association

Office of the Registrar  
1902-2 Bloor Street West  
Toronto, ON M4W 3E2

[www.hrpa.ca](http://www.hrpa.ca)

[registrar@hrpa.ca](mailto:registrar@hrpa.ca)

416.923.2324

Toll-free 1.800.387.1311

Fax 416.923.7264