



FAQ on the Scoring of the Exams

July 8, 2010

Claude Balthazard, Ph.D., CHRP
Director, HR Excellence and Registrar
Human Resources professionals Association

1. *There have been some changes in how the exams are scored, haven't there?*

Actually there have been a few changes. In October 2009, CCHRA began using Angoff panels to set cut-scores for the exams. In May 2010, CCHRA switched to a floating-cut score and changed how it handled questions deemed flawed in the post-exam review.

2. *Who decided to make these changes and for what reason?*

The development and scoring of the exams is the responsibility of CCHRA's Independent Board of Examiners (IBE). The reason for that is that it is important that there be independence between those who develop the exams, set the cut-scores, and score the exams and the provincial associations. This is to avoid any possibility that decisions about exam content or the scoring of the exams be influenced by the self-interest of the provincial associations. Having the development and scoring of the exam conducted by an independent body of experts enhances the confidence that exam-writers, members of the profession, and the public have in the exam process and results.

The recommendations of CCHRA's IBE in regards to the development and scoring of the exams are reviewed by CCHRA's Professional Standards Committee and finally approved by CCHRA's Board of Directors which has representation from all CCHRA member associations.

The reason for all these changes is the same—to keep up with best practices in the field of credentialing. The field of credentialing is complex and sophisticated. Best practices evolve over time. These recent changes in the scoring of the exams are evolutionary rather than revolutionary. These changes did not change the difficulty of the exams—it is not easier or more difficult to pass the exams than it was before these changes were made. The changes provide a more precise way of setting cut-scores to take into account the variations in exam difficulty which may occur, a more transparent way of scoring the exams, and a more scoring neutral way of handling flawed items.

3. *How is the cut-score linked to the difficulty of the exam?*

Exam used in the context of certification and licensure are 'criterion-referenced.' This means that the exams are used to determine whether candidates have the level of knowledge or proficiency required to be certified. For exams used in the context of certification and licensure, there can be no quotas or targeted pass rate.

It is important that this standard be the same from one sitting to the next, from one form of the exam to the next. Despite the fact that all exams are developed according to specifications, some forms of the exam may be somewhat more difficult or somewhat easier than others. On the other hand, groups of exam writers also differ. Some groups of exam-writers can be better prepared than others. Ensuring that the standard of competence or proficiency required to pass the exam is consistent from one exam sitting to the next is a task which is confided to psychometricians. Starting October 2009, CCHRA started using what are known as 'Angoff panels' to set the cut-scores for the exams.

4. *What is an Angoff panel?*

An Angoff panel refers to a specific approach to setting the cut-score for an exam to ensure that exam difficulty is consistent across forms of the exam. Angoff panels got their name from William Angoff, who was a psychometrician at ETS for over 40 years and who developed the method.

The intent, of course, is that the exams be of the same difficulty from one sitting to the next; nevertheless, some differences in exam difficulty can occur. One way of thinking about exam difficulty is in terms of the cut-score: a relatively easier exam will have a higher cut-score and a relatively more difficult exam will have a lower cut-score.

Conducting Angoff panels is a more systematic and precise way of setting the cut-score for a particular form of an exam.

5. *How does an Angoff panel work exactly?*

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.

The table below gives an example of what Angoff panel data looks like and how the cut-score is arrived at. The data in the table are entirely fictitious. For each question, Angoff panel members were asked to give the probability that a candidate at the threshold of competence would be able to answer the question correctly.

The Angoff index for a given item refers to the average probability of answering the item correctly as averaged across panel members. Summing the Angoff indices for all the questions included in the exam test gives the proposed cut-score for the whole test. Often this number will fall between two integers; in this case the cut-score is taken as the higher value. For instance, in the table below, the sum of the Angoff indices for all questions was 99.45. However, scores for individual exam-writers can only be integers. Here, the cut-score would be deemed to be 100—a score of 100 would be above the cut-score and a score of 99 would be below the cut-score. Given that scores are reported as percentages the next step is to convert this raw-score cut-score into a percentage cut-score. A score of 100 expressed as a percentage is 66.67%, this would be rounded to the closest percentage which is 67%.

Sometimes, the cut-score needs to be adjusted to eliminate inconsistencies due to rounding. Let's say, for example, that the raw-score cut-score in the example above had worked out to be 100.65. In this case, the raw score cut-score would round to 101, producing a percentage cut-score of 67.33 which rounds to 67%. Those exam-writers that scores 100 also have a rounded percentage score of 67%. So, both scores of 100 and 101 convert to the same rounded percentage score of 67%. It would appear inconsistent that some with a rounded percentage score of 67% would be deemed to have passed the exam whereas others with the same rounded percentage score of 67% would be deemed to have failed the exam. To avoid this apparent inconsistency, the raw-score cut-score would be adjusted to 100, although this makes no difference in terms of the rounded percentage cut-score.

Because the cut-score for the test is derived by adding the probabilities of a correct answer for each question, the cut-score will vary depending on the particular set of items that make up the test. Each version of the NKE 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.

	Angoff Panel Judge					Across Judges	
	1	2	3	4	5	Average	Standard Deviation
Question 1	.75	.75	.80	.65	.70	0.73	.057
Question 2	.65	.70	.75	.65	.80	0.71	.065
Question 3	.70	.65	.60	.65	.65	0.65	.035
Question 4	.65	.75	.65	.70	.60	0.67	.057
Question 5	.55	.50	.45	.65	.55	0.54	.074
---	---	---	---	---	---	---	---
Question 146	.80	.80	.80	.70	.60	0.74	.089
Question 147	.80	.75	.70	.55	.65	0.69	.096
Question 148	.55	.60	.65	.65	.45	0.58	.084
Question 149	.65	.65	.70	.75	.65	0.68	.045
Question 150	.65	.70	.65	.65	.55	0.64	.055
Passing score	101.25	102.75	101.25	99	93	99.45	3.846

6. *What is a floating cut-score?*

The problem with a fixed cut-score is that it doesn't take into account the relative difficulty of the exam. Starting May 2010, CCHRA is moving to what is called a 'floating cut-score.' Using a floating cut-score means that the cut-score could vary from one administration of the exam to the next—for a more difficult exam the cut-score would be lower; for an easier exam, the cut-score would be higher.

7. *In October 2009, CCHRA used an Angoff panel but the cut-score was still 70%, how come?*

To take into account the relative difficulty of an exam, either of two things could be done: (1) keep the cut-score at 70% but adjust the scores up or down depending on the relative difficulty of the exam, or (2) adjust the cut-score and leave the scores as they are. In October 2009, the Angoff panel proposed a cut-score of 94 out of 150 (63%) for the NKE. The approach taken in October 2009 was to add 11 points to all the scores. This works out to the same thing as using a cut-score of 63%.

8. *What difference does it make whether a fixed or floating cut score is used?*

For exam-writers, the switch from a fixed cut-score to a floating cut-score is something that makes no difference:

- There is no difference in how exam-writers would prepare for the exam
- There is no difference in how an exam-writer would respond to exam questions
- There is no difference in the pass/fail outcome

Although it would be possible to keep on making adjustments to the scores such that the cut-score would remain at 70% for all sittings, this is not the preferred way to go. Psychometricians prefer to leave the scores as they are because it is a more transparent way of doing things.

9. *I wrote the exam in October 2009 and got a score of 69% and failed because the cut-score was 70%; now the cut-score is 67%, if that cut-score had been used in October 2009, I would have passed—how is that fair?*

In October 2009, an adjustment was made to the scores so that the cut-score could be kept at 70%. You can't compare the two.

All in all, we are going to have to get used to the fact that what is a passing score in one sitting may not be a passing score for another sitting. You can't say "I want to write a relatively easier exam with the cut-score for a relatively more difficult exam"—that would be unfair. Indeed, the whole idea behind floating cut-scores is increased fairness.

10. *I noticed that the NKE was scored out of 148, what is that all about?*

That has to do with the fact that two flawed items were removed from the scoring of the exam. In the test development process, every effort is made to ensure that items are of high quality. However, it does happen that some questions don't work out.

"Non-performing items" is a term psychometricians use to refer to questions that didn't perform as expected based on statistical indices. For instance, one important aspect that psychometricians look at is 'item discrimination.' Item discrimination refers to the ability of a question to differentiate between exam-writers at different levels of knowledge or proficiency. Item difficulty, which is the proportion of exam-writers that answer the question correctly, is also considered.

Questions that fail to perform as expected are flagged for further review. One of the tasks of the Independent Board of Examiners is to review the questions that were flagged as 'non-performing'

on statistical grounds to determine whether these items were flawed in some way. Information such as the pattern of responses across incorrect response options (“distractors”) is examined. As a result of their post-exam review, the IBE may determine that some questions were indeed flawed.

Flawed items detract from the reliability and validity of the exams. The best way of handling flawed items is simply to throw them out—that is to act as if these questions were not there. The scores and the Angoff cut-score are recalculated leaving these flawed questions out. This is the most ‘scoring neutral’ approach to handling flawed items.

For the May 2010 exams, two questions on the NKE and three questions on the NPPA were deemed flawed. Therefore, the NKE was scored out of 148 and the NPPA was scored out of 57.

11. *So when is the final cut-score set?*

A preliminary cut-score can be calculated after the Angoff panel has finished its work. However, the final cut score is not known until the post-exam review is completed. Only once the final set of items has been arrived at can the final cut-score be determined. The final cut-score is only known a short time before the scores are released.

12. *How does rounding work?*

The idea with rounding is to round once and as late as possible. The reason is that we don’t want rounding errors to compound. Final scoring is based on the items that have been retained after the post-exam review. For the May 2010 exams, two questions were discarded for the NKE and three questions for the NPPA. This means that the NKE was scored out of 148 and the NPPA out of 57. For each exam-writer the number of correct answers across the retained items is calculated. This number of correct answers is divided by the number of retained questions. This gives a proportion of correct answers. For instance, an exam-writer with a score of 99 on the NKE would get a proportion correct of $99/148$ which is .6689189. This number is converted to a percentage score by multiplying it by 100. Here the percentage score would be 66.89189%. This percentage score is rounded to the closest percentage integer, so 66.89189% becomes 67%. This rounded percentage score is the score that is reported.

The cut-score is also reported as a rounded percentage score. If the exam-taker score is equal or greater than the cut-score that exam-writer is deemed to have passed the exam.

13. *How do I find out more about all of this stuff?*

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*. (3rd Ed.). Lawrence Erlbaum Associates: Mahwah, NJ.

National Commission for Certifying Agencies (2004). *Standards for the accreditation of certification programs*. Author: Washington, DC.