

Floating cut-scores and deleted questions

The National Knowledge Exam (NKE) is prescribed as a requirement by for the CHRP designation by the Board of Directors of HRP A pursuant to the authority granted to the Board by the *Human Resources Professionals Association of Ontario Act, 1990*. HRP A has delegated the development and scoring of the NKE the Canadian Council of Human Resources Associations (CCHRA). Nonetheless, HRP A remains fully responsible and accountable for exams administered to its members that are developed and scored through the CCHRA.

This delegation to CCHRA of the development and scoring of the NKE comes with certain terms and conditions. The most important stipulation with respect to the delegation of the development and scoring of the NKE to CCHRA is that everything to do with the exams must meet or exceed the standards set out in the *Standards for Educational and Psychological Testing* jointly published by the AERA, APA, and NCME and the standards set out in the *NCCA Standards for the Accreditation of Certification Programs* published by the National Commission for Credentialing Agencies (NCCA).

The *Standards for Educational and Psychological Testing* are available in the HRP A Resource Centre and the *NCCA Standards for the Accreditation of Certification Programs* are available from the following web site:

<http://www.credentialingexcellence.org/AccreditationServices/CertificationAccreditation/StandardsInterpretations/tabid/93/Default.aspx>

The standards that credentialing exams must meet are much more stringent than those which typically apply to exams used in college and university courses. The development and scoring of credentialing exams requires very specific expertise in psychometrics. To ensure that exams meet or exceed professional and technical standards, CCHRA has engaged the services of an external psychometric consultant and established an Exam Board comprised of psychometric experts from the different regions of Canada including representation from Ontario. This Exam Board oversees the exam process at CCHRA including the work of the external psychometric consultant. Four out of five members of the CCHRA Exam Board have Ph.D.s in psychometrics or related disciplines.

Through participation in all key CCHRA committees, and because of the technical expertise assembled by CCHRA, HRP A is confident that the exams developed and scored on behalf of HRP A will meet or exceed all applicable professional and technical standards.

A number of changes in our testing processes and procedures, some more visible than others, were introduced in order to meet the *Standards for Educational and Psychological Testing* jointly published by the AERA, APA, and NCME and the standards set out in the *NCCA Standards for the Accreditation of Certification Programs*. Of these changes, two appear to be most misunderstood: floating cut-scores and the deletion of flawed questions. A cut-score is the score needed to pass the exam. A floating cut-score refers to the fact that the cut-score is set for each form of the exam based on the difficulty of the exam and thus can vary somewhat from one form of the exam to another. The deletion of flawed questions refers to the practice of discarding questions based on post-administration analyses. The

deletion of flawed questions means that the exam score may be calculated based on fewer questions than were administered in the exam.

Although each of these practices has been questioned by some exam-writers, it is important to note that both of these practices are not only consistent with applicable technical standards, they are in fact required by these technical standards. Let's review each in turn.

Floating cut-scores

Both the *Standards for Educational and Psychological Testing* and the *NCCA Standards for the Accreditation of Certification Programs* are clear on the matter—fixed cut-scores should not be used for credentialing tests. The cut-score needs to be set in such a way that takes into account the relative difficulty of the exam. Since October 2009, CCHRA has used Angoff panels to set the cut-scores for its exams. The Angoff method is the most widely used method for setting cut-scores on credentialing tests.

Although it is the intention to create exams of similar difficulty, it is inevitable that some exams will vary somewhat in difficulty. When an exam is somewhat more difficult, the cut-score will be correspondingly lower; when an exam is somewhat less difficult, the cut-score will be correspondingly higher. This means that a given score, say a 69, could be at or above the cut-score on one form of the exam and below the cut-score on another. This is because the same score doesn't mean the same level of proficiency for all forms of the exam.

For those looking for a more extensive explanation of how cut-scores are set, the following references may be useful:

- 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.

The important point here is that floating cut-scores for the NKE were established according to professional and technical standards.

The deletion of flawed questions

The deletion of flawed questions is standard practice among psychometric experts. First, it is important to note that the purpose and intent of the exam is to make credentialing decisions. Psychometricians work towards maximizing the validity and reliability of the examination. The NKE is used to decide whether an exam-writer possesses the requisite level of knowledge in HR to be certified as a CHRP.

Such a decision creates the following decision table:

Decision Accuracy

		Decision Based on Observed Scores	
		Fail	Pass
True State	Sufficient Knowledge	Incorrect Decision	Correct Decision
	Insufficient Knowledge	Correct Decision	Incorrect Decision

Every aspect of the examination process, including the scoring of the exam, is designed with the objective of minimizing the number of decision errors. Minimizing such errors is in the public interest. Thus, the manner in which the exam is scored is designed to minimize the number of decision errors. Discarding questions based on a post-exam review reduces the number of incorrect credentialing decisions; not doing so would increase the number of decision errors.

As it turns out, on virtually any exam it will be the case that some questions do not contribute to the accuracy of the overall test. This is related to the concept of ambiguous questions. There is a difference between 'ambiguous questions' and 'flawed questions,' however. Ambiguity is a subjective impression on the part of exam-writers. There are problems with this subjective definition of ambiguity, however. One problem is that different exam-writers will find different questions ambiguous. Another problem with the subjective approach to ambiguity is that exam-writers with less knowledge will find more questions ambiguous.

Psychometricians have worked out procedures for identifying the contribution each question makes to the accuracy of the decision made on the basis of the total score. This process is called *item analysis*, and is the crux of the post-exam review. The purpose of the post-exam review is to identify flawed items—*flawed* being defined here as 'not making a contribution to overall decision accuracy.'

It would be a highly unusual if exam-writers did not feel that at least some of the questions were ambiguous. Nonetheless, the subjective impressions of exam-writers are not reliable indicators that a question is problematic. Psychometric indicators provide a more reliable and objective basis for making decisions as to whether a question is making a contribution to the accuracy of the test as a whole.

Having established that a question does not contribute to the overall accuracy of the exam, psychometricians agree that the removal of the question is the best course of action. The deletion of questions based on post-exam analyses is an accepted psychometric practice which improves the accuracy of the exam.

One of the reasons why one does not hear of discarded questions on college or university exams is that very few instructors (probably only psychometrics instructors) ever conduct post-exam item reviews. If these post-exam item reviews were conducted, likely a number of questions would be found to be flawed.

No exam developer sets out to create flawed questions. At the time the NKE was assembled, all of the questions were judged to be viable questions by the exam review panel. Before the test was administered, none of the questions were thought to be ambiguous by the exam review panel.

What is important here is not so much the number of questions that were discarded, but the fact that the questions that were retained comprise a valid and reliable basis for making credentialing decisions. The external psychometric consultant engaged by CCHRA and the Exam Board were of the opinion that (1) that the deletion of questions based on psychometric indicators improved the accuracy of the test, and (2) that the remaining questions comprised a valid and reliable basis for making credentialing decisions.

The bottom line

HRPA is responsible and accountable for the exams which are developed and scored on its behalf. While HRPA has delegated the development and scoring of the NKE to the CCHRA, HRPA is involved in various parts of this process. HRPA has stipulated that all work done by CCHRA on its behalf meet or exceed the standards set out in the *Standards for Educational and Psychological Testing* and the *NCCA Standards for the Accreditation of Certification Programs*. CCHRA has engaged the services of an external psychometric consultant and constituted an Exam Board comprised of psychometricians recommended by the member Provincial Associations to oversee the development and scoring of exams. For these reasons HRPA has full confidence that all work done on its behalf by CCHRA meets or exceeds the *Standards for Educational and Psychological Testing* and the *NCCA Standards for the Accreditation of Certification Programs*.

There is no doubt that some of the accepted practices in psychometrics are not intuitively obvious to everyone. Be that as it may, HRPA does not see a choice here; the exams prescribed by HRPA must meet or exceed professional and technical standards as determined by experts in the field.