CISRO / OCRA

Control of exam validity
Implementation of Harmonized LLQP

August 2014

Context

By January 2016, the new pan Canadian qualification program will be implemented to replace the existing programs. It is important to plan the transition towards the Harmonized Life License Qualification Program (LLQP) in order for its implementation to run smoothly.

The LLQP derives its characteristics from the current qualification processes and the changes it will bring will vary from one jurisdiction to the next. A modular approach with an open-book exam based on standardized study material will require some adjustment from both Course Providers and candidates. Consequently, it will be important to ensure that the new exam adequately evaluates the achievement of the revised qualification standards.

Therefore, the objective of the following exam validity control mechanisms is to ensure that well-prepared candidates who have mastered all evaluated competencies may successfully pass all modules.

First and foremost, the validity of the new exam will be controlled by a rigorous exam question development process. It will ensure all questions selected to compose the exam meet the highest testing standards. This development process started more than two years ago, with the occupational analysis workshops that helped defined what it is that the Harmonized LLQP needs to evaluate. In other words, measures taken before the exam is implemented will have already played a role in ensuring its validity.

Despite this, the exam's real (vs. theoretical) level of difficulty will be confirmed only once the exam is administered. The challenge lies more specifically with the first candidates that will write the new exams. After only a few weeks, the educational specialists will have collected sufficient data to analyze the exams' characteristics. It is anticipated that the data will confirm expectations about the exam's level of difficulty. Nonetheless, the data will allow specialists to bring adjustments, if required.

The following transitional measures, implemented with the first exam administration, will effectively ensure the first candidates are treated fairly. It will serve to control exam validity and regulate the level of difficulty of exam questions. This, in turn, will guarantee LLQP scores are valid as of the first administration.

Notwithstanding these transitional measures, LLQP exam questions will continue to be monitored on an on-going basis throughout their administration to control their statistical qualities.

Transitional exam validity control mechanisms

To ensure their validity, initial candidates' results will be analyzed according to the following principles until sufficient statistical data is gathered on all exam questions. Following this analysis, potentially problematic exam questions may be cancelled (and results adjusted) if it is found that they do not provide a reliable measure of candidates' competency.

The following criteria will determine whether transitional control mechanisms must be applied – as stated earlier, standard control mechanisms of exam question validity, which are explained later in this document, will apply at all times.

- Results will be adjusted (by cancelling potentially problematic questions) only if a module's pass rate is below 70%;

- Results will be adjusted until the following conditions are met, as reviewed weekly by the measurement and evaluation specialists:

  - A minimum of 300 responses have been collected for each exam question;
  - An average pass rate of 70% or more has been achieved for 4 consecutive weeks for each modular exam format.

It will be important to react quickly and to collect sufficient data to produce reliable indices leading to the identification and correction of problematic questions. Consequently, the following measures will be implemented to attain the targeted pass rate as quickly as possible:

- The number of exam formats administered for each module will be limited until sufficient statistical data is gathered on each question to build supplemental formats;

- Transitional control mechanisms may apply to only one of a module's formats if it is deemed more difficult than the other(s).

Both standard and transitional control mechanisms will rely on the difficulty and discrimination indices. These will be used to identify the questions that should be monitored more closely or cancelled altogether and/or replaced without delay.

- The difficulty index is a measure of how many candidates successfully answer the question. The percentages in the diagram that follows indicate the proportion of candidates having successfully answered the question.

- The discrimination index refers to a question's ability to differentiate between strong and weak candidates:

  - A positive index indicates that strong candidates tend to successfully answer the question whereas weak candidates do not;
  - A null index indicates that the question does not provide any differentiation;
  - A negative index indicates that strong candidates tend to fail the question whereas weak candidates do not.

The following benchmarks will be used to determine a question's status:

- All questions passed by a proportion of candidates of 70% or more should be kept;

- All questions passed by a proportion of candidates of 40% or less should be cancelled and/or replaced;

- Questions passed by between 40% and 70% of candidates should be kept, monitored and potentially revised, or cancelled/replaced, based on the discrimination index:

  - Questions with a positive discrimination index should be kept;
  - Questions with a null index should be monitored and potentially revised;
  - Questions with a negative index should be cancelled and/or replaced.

Particular attention will be paid to questions marked as "Keep and monitor" to identify means of improving their quality.

Moreover, further analysis may be undertaken to ensure the reliability of data used to determine that a question should be cancelled. For example, if a Course Provider or jurisdiction is over-represented in the data collected, comparative evaluations will be done to ensure its candidates' results are representative of those of other Course Providers or jurisdictions. This will avoid cancelling valid questions due to skewed data.

Exam question validity indicators
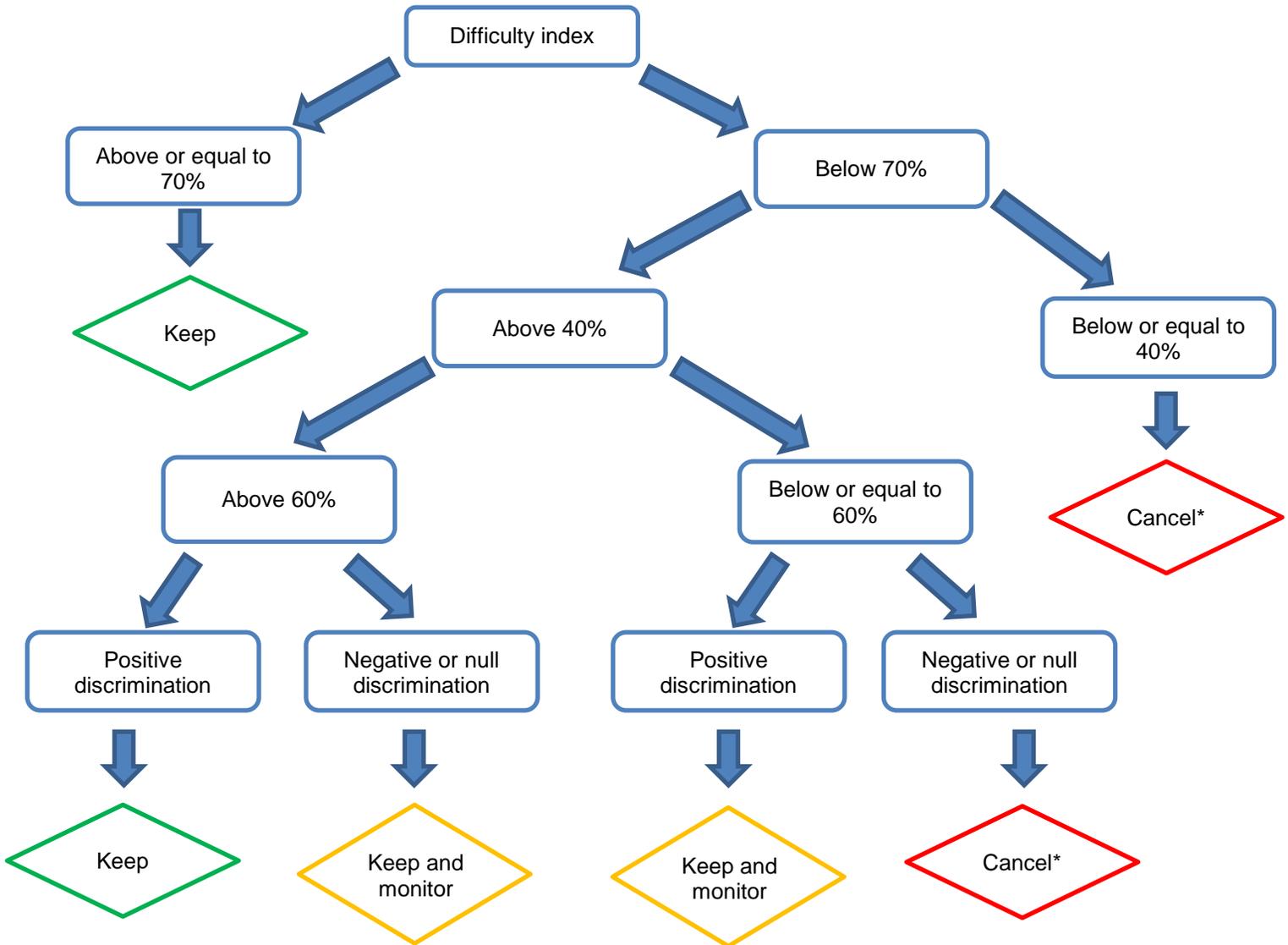
* Further analysis may be required to confirm that a question should be cancelled.

The application of the above control mechanisms will lead to the exclusion from initial candidates' scores of unduly difficult questions and of those unable to differentiate between candidates that are well prepared and those that are not.

Other measures to monitor the Harmonized LLQP exam

The adjustment of initial exam results will not be the only measure put in place by the exam development team to ensure its measurement tool is appropriate. Standard control mechanisms will also apply.

As stated earlier, the exam development process itself is structured to support the reliability and validity of the exam. From the start, the occupational analysis workshops brought together a representative sample of life insurance agents whose input led to the elaboration of the Competency Profile and the drafting of the Curriculum. These documents respectively address the tasks that a life insurance agent should be able to carry out and the skills that should be evaluated upon career entry. Furthermore, the questions will be rigorously drafted and validated, through the involvement of competent industry subject matter experts (SMEs) working in different jurisdictions. And questions will be selected by measurement and evaluation specialists based on the specifications of the Curriculum. All of this will ensure the exam is valid and relevant to professional practice.

Once it is implemented, the exam will be continuously and very closely monitored, over and above the analysis required to regulate exam results. Questions presenting measurable weaknesses, observed through statistical analysis, will be revised or replaced. Moreover, candidate and stakeholder comments and suggestions will initiate verifications and possibly additional validations with SMEs. In turn, these may lead to questions being revised or even cancelled despite their favorable statistical characteristics.

Conclusion

Direct administration, rather than pretesting, is being recommended to implement the Harmonized LLQP exam. But this does not mean that the exam's qualities will be completely unknown and uncontrolled. Both quantitative and qualitative control mechanisms will be implemented.

From the start, the exam will have been developed according to a rigorous process. This will have allowed the appropriateness of its construct and content to be confirmed repeatedly. As for the various steps leading to its elaboration, they will have guaranteed its validity and relevance to entry-level professional practice.

Moreover, the regulation of initial candidates' exam results will mitigate the impact of any potential measurement errors by cancelling statistically inappropriate questions. This administrative measure will be applied until a targeted pass rate is achieved through corrections stemming from the monitoring of exam formats. In turn, this will help reduce the impacts of the implementation of the Harmonized LLQP on the certification rate of new agents. It will ensure a smooth transition, beneficial to all stakeholders, all the while guaranteeing a quick confirmation of the validity and reliability of the licensing exam.

Bibliography

American Educational Research Association, American Psychological Association and National council on Measurement in Education (1999). Standards for educational and psychological testing. Washington: American Educational Research Association.

Bertrand, R., et Blais J. G. (2004). Modèles de mesure : L'apport de la théorie des réponses aux items. Sainte-Foy : PUQ.

Crocker, L. and J. Algina.. (1986). Introduction to Classical and Modern Test Theory. California: Wadsworth Group/Thomson Learning.

De Landsheere, G. (1992). Dictionnaire de l'évaluation et de la recherche en éducation : 2e édition. Paris : Presses Universitaires de France, Coll. «Les grands dictionnaires».

Kane, M., T. Crooks and A. Cohen (1999). Validating Measures of Performances. Educational Measurement: Issues and Practice, Vol. 18, No 2, pp, 5-17.

Laurier, M. D. (2005). Les principes de la mesure et de l'évaluation des apprentissages: 3e édition. Montréal: Gaëtan Morin Éditeur.

Legendre, R. (2005). Dictionnaire actuel de l'éducation : 3e édition Montréal: Guérin, Éditeur ltée.

Louis, R. (1999). L'évaluation des apprentissages en classe : 1re édition. Laval: Éditions Études Vivantes.

Morissette, D. (1984). La mesure et l'évaluation en enseignement: 1re édition. Sainte-Foy: Les Presses de l'université Laval.

Moss, P.A. (1995). Themes and Variations in Validity Theory. Educational Measurement: Issues and Practice, Vol. 14, No 2, pp, 5-13.