STATE HEALTH SOCIETY, BIHAR

IMPORTANT NOTICE

(Advertisement No. 07/2024)

This is to inform to all candidates that, we have received more than expected number of applications for the post of Community Health Officer (CHO). Numbers of applications are larger than test taking terminals available in Patna, therefore it is to inform the candidates that the exam for CHO will be conducted in multiple batches. In order to address concerns of candidates with regard to varying difficulty level of Questions Papers for different batches, we will adopt scientifically proven statistical method of Equi-percentile Equating Methodology for scoring and evaluation. This methodology is also acceptable in the court of law.

SCORING/EVALUATION:

- The evaluated scores obtained by each of the candidates in different batches (multiple batches) if held will be calculated using Equi-Percentile Equating Methodology.
- Scores up to eight decimal points shall be taken for the purpose of calculations.
- Equi-percentile Equating Method will be used for shortlisting candidates if exams are conducted in multiple batches only. If exams are conducted in single batch – overall score will be considered.
- Candidates can know more about the method in the FAQ's on Equi-percentile Equating Method available on the SHSB website http://statehealthsocietybihar.org/

Deputy Secretary
-cum-Incharge HR
28/11/24

Frequently Asked Questions
Equating of Scores on Multiple Forms

1) Why is Equating of scores on multiple test forms (also known as multiple test Question papers) needed?

In mass conducted and large scale high-stake tests or examinations spread over a window of several days and several administrations, it has not been practically possible to use a single form of test in a subject, either for entrance/admission or for recruitment/achievement. On a single day, one may use a single format with reshuffled item numbers and/or reshuffled options of the items or multiple forms. It is therefore essential that different forms of a test used on different days and/or different administrations/batches in a day should be first of all ensured that they are equivalent to each other, i.e. designed to be equivalent. In other words, every item in a test to a corresponding item in the other format matching the content topic, the ability cluster tested and the difficulty level.

It is, however, a requirement that a single reference test paper should be created with the same content template that will be used to generate several multiple forms.

2) What is Equating and what are its Results?

Equating is a statistical process that is used to adjust scores on test forms so that scores on the forms can be used interchangeably. It adjusts for differences in difficulty among forms that are built to be similar in difficulty and content.

The outcomes after equating are, the reference form test scores remaining as they are and the test scores on each of the other forms are equated to the reference form test scores by a method selected for Equating resulting in test scores in all forms measured to the same scale.

3) Are there several methods of Equating?

Yes, there are several methods of Equating and they are as follows:

a) Mean Equating
b) Median Equating
c) Linear Equating (Based on Mean and S.D.)

d) Equipercentile Equating
e) Equating using Item Response Theory
f) Anchor Test Equating

For the given examination, Equipercentile equating method will only be used:

Equipercentile Equating involves percentile rank or score to be found for all scores in each of the forms and of all forms and clubbed together to generate a merit list. This uses the distributions of scores on the two test forms (X & Y) and finds a set of pairs of equivalent scores such that the proportion of individuals with each score below percentile rank is the same for the two test forms.

Percentile rank for a given raw scores indicates the percentage of a student who score below this mark.

Example: If in two theoretically designed equivalent test forms administered to Group 1 and Group 2, the maximum mark being 20, a student getting 12/20 in Group 1 has (say) 65% of students below his score then his percentile rank is equivalent to 65^{th} Percentile Rank. On the other hand, if in a Group 2, a student getting 13/20, in this case as well having 65% of students scoring below 13 in his Group, then the score 13 has a 65th percentile rank. In the above example, the score of 12 in Group 1 is equal to score of 13 in Group 2, with both having 65^{th} Percentile Rank.

4) Have researchers or psychometricians recommended Equipercentile Equating method?

Yes, Researchers or psychometricians after prolonged and extensive research have come recommended Equipercentile Equating method.

5) Why Equipercentile Equating method is preferred or chosen for the given examination?

Among the methods using Classical Test Theory (CTT), Equipercentile Equating turns out to be the best both from a statistical point of view and from purely a common sense point of view. The constitution of the reference forms, the number of items in it, content-wise, ability-wise and difficulty-wise, to be designed theoretically the

same with every other form. This method is found to be satisfying both statistical righteousness (dharma) and legal propriety (everyone is the same before law, no advantage or disadvantage to anybody and a level playing field).

6) Are there instances of Equipercentile equating getting through a legal scrutiny, in India?

Yes. On the grounds of statistical righteousness and legal propriety, the judgement of court(s) were in favour of Equipercentile Equating method adopted for publishing test results that involved multiple batches or test forms. The Apex Court in India provided a decision in favour of a High Court Judgment, upholding the Equipercentile equating method, thus lending legal authenticity to the Equipercentile equating method for multi-batch examinations.

7) When many multiple forms (supposedly and nearly equivalent to each other) are used, which of them is to be taken as reference group/batch and why?

The usual practice in all multi form testing scenario, a pre-planned carefully constructed (according to an agreed content template) will form the basis for generating equivalent multiple forms. Hence, the reference format is prior decided and used. When this is not the practice as it happens with many cases, the batch with the test scores yielding for its maximum score, i.e. the highest percentile rank/score will be taken as the reference batch. The simple reason is the percentile rank generated for this form will have a large range that can accommodate all the ranges of percentile ranks of other multiple forms test scores so that there are no outliers. This is the practice evolved after considerable research and experimenting with them over a period of time, accepted and recommended as a best practice in the industry.

8) How is the equivalence of multiple test forms ensured in practice?

In practice there are two options, the **first option** is Angoff's method by getting several SMEs to create collectively a sample paper specifying contents, ability to be tested and the guessed value of item difficulty and use this test paper to drive creation of similar other

papers, matching item by item, the specification of content, ability and item difficulty of the first set. This is the long drawn process and it requires quite a bit of hard work, discussion amongst the SMEs, time and cost. The **second option**, a better one at that, is to use the item bank already available with every item specified and coded with content, ability tested and item difficulty with the help of a content template specifying the numbers that are decided to have in each of the content, ability cluster and difficulty. This will ensure multiple forms of same test difficulty with negligible varying range of difficulty, if any.

9) Can any form become an act of circumstantial compulsion and if so, how would the exam body deal with it?

Yes, it is possible and it has happened and it may continue to happen. That it may be a force of circumstance that will render an additional test to be held on a later date for which already a test form is available or is to be created and in either case, the principles outlined in some of these answers will be taken up in so far as creating the additional test paper. However, Equipercentile Equating can be applied even though the subsequent group that has not been planned to have more or less the same number of test takers can be taken up and Equipercentile Equating may again turn out to be the best method.

10) Some Exam Bodies come out with some test takers getting 100 Percentile Rank. Is it at all possible? If not, Why?

This phenomenon has been seen in some test providers without a clear and correct understanding of what percentile rank means. For a given raw score, the percentile rank score gives the percentage of test takers whose scores are below this score. By no stretch of imagination and statistical interpretation, any score can have 100 percent of test takers below this score. This means the concerned score is also included which is absolutely erroneous. They fail to understand the significance of the correct definition of percentile rank which is the percent of midpoint of cumulative frequency at this score level. This means the midpoint of cumulative frequency at the score level is sum of cumulative frequency at that score level and the cumulative frequency of the next lower score divided by 2 and therefore, the percentile rank will be certainly less than 100.

11) Is it possible to generate Percentile Rank manually using Excel or should a software be used? Can Equated Score as well be manually done or with the help of a software?

If percentile rank is generated manually using excel, there is a built-in check that the top score cumulative frequency will be equal to the number of test takers. However, the hidden details of this is not seen in the software used.

For equated scores, it is always preferable to use a software especially developed for it since the manual calculations will be extremely time consuming and prone to errors.

12) Experts have rated Anchor Equating higher than that of Equipercentile Equating. Why it is not recommended for Indian conditions?

It can be appreciated that Anchor Test Equating involves a definite number of items called anchor test items to be placed as an integral part of every form test paper. An instruction given to the test takers will have to say that the score on the anchor test items for any test taker will not be counted for his total score but only used as a means to provide an equating algorithm to be applied to the rest of the item scores in each form. This instruction will not be seriously taken by Indian test takers who may choose not to attempt or take it serious and thus, the whole process will be vitiated. For this reason, Exam Body's choice is not to go for Anchor Equating.

13) It is found that when two test forms are used and scored, it may become inevitable to remove few items from one of the test forms due to inadvertent errors, thereby reducing the number of items from that of the previous test form. How can this be dealt with?

True, this situation is very likely and it can surface at times. In a case study, 100 items were used in each test form and in one form, two items were found to be invalid and unacceptable. Hence, they needed to be removed in which case there were two options to deal with it.

a) Score the test form of 98 items for 98 and work out the percentile rank of these scores.
b) Give 2 marks to every test taker (assuming each item carried one mark for right answer), no matter whether he attempted/answered it right or wrong making it out of 100. The percentile rank for these scores will be found out.

There will not be any change in the percentile rank in both the cases and therefore, the merit list. Certainly, there will be changes in equated scores.

14) In a case study, a strange situation had arisen with a recruitment test with 5 multiple formats, the first day two of them and subsequent day three of them. Test takers were drawn for these 5 test formats from different districts, the numbers being small for each batch. The scores on these multiple forms were equated using the Equipercentile Equating method (the legally sanctioned format) and the merit list created. The merit list is the combination of different batches taking different test forms and from different districts each. A query was sent for seeking district-wise merit list. Is that possible? If so, how and if not, why?

The multiple forms' scores have been converted into percentile rank/scores considering each batch and the test takers. The merit list was derived putting them together, their percentile scores arranged highest to lowest to generate an overall merit list. This is the most accurate method, for the given case in hand. If district-wise merit list is required, they can be taken out of this overall merit list applying the cut-off score on percentile rank as decided by the Exam Body. There is no other way to work out an additional merit list by taking district-wise test takers (they take different formats and PR does not mean anything to them). So, the simple answer is district-wise merit list is to be a part of overall merit list only.