Assessment Analysis: How It Is Done

Thomas Puthiaparampil

Corresponding author: Prof Thomas Puthiaparampil pthomas@unimas.my
Institution: 2. Universiti Malaysia Sarawak (UNIMAS)
Categories: Assessment, Medical Education (General)

Received: 01/08/2017
Published: 04/08/2017

Abstract

This article elaborates assessment analysis, a topic recently introduced in another article published in January 2017 by the same author in AMEE MedEdPublish. The formulae used to generate difficulty index and discrimination index of examiner-scored assessments as well as whole papers like MCQ are described in simple terms. The formulae used for assessment analysis are aligned to those established decades ago for item analysis. A step-by-step guideline is included to enable those interested to practice assessment analysis. The formulae, an Excel worksheet showing index calculations and the interpretation of index values are also included.

Keywords: Assessment Analysis

Background

Assessment analysis was introduced in the article "FROM ITEM ANALYSIS TO ASSESSMENT ANALYSIS: INTRODUCING NEW FORMULAE" by Puthiaparampil T. et.al, published on 11/01/2017, 6 (1), Paper No: 7, DOI: https://doi.org/10.15694/mep.2017.000007.

This new concept was discussed and two sets of formulae (one using binary method and the other using actual scores) compared in the above article (Puthiaparampil T. et.al 2017). Another study also compared the difficulty index and discrimination index of MCQ paper with long case in 2014 (Taib and Yussof, 2014). Literature review by the author did not show other studies using the concept of assessment analysis. This ‘Practical Tips and Guidelines’ describes the formulae for difficulty index (DIFi) and discrimination index (SISi) for assessments using actual scores, and gives a step-by-step guideline on how they are used.

What is assessment analysis?

Assessment analysis in short is a method to ascribe difficulty index and discrimination index values to whole
assessments, especially those scored by examiners.

The need for assessment analysis

Several assessment tools are employed in all the clinical end of posting examinations and final professional examinations of the Faculty of Medicine and Health Sciences, Universiti Malaysia Sarawak (UNIMAS), for the past two decades. The assessment tools employed are: Multiple True/False Questions (MCQ), One Best Answer Questions (BAQ), Modified Essay Questions (MEQ), Objective Structured Clinical Examination (OSCE), Long Case and Short Case. See table 1 for details.

It has been observed consistently that the passing rate or marks scored in various assessments differ widely. For example, MEQ 1 has 100% passing rate, while MEQ 2 shows only 20% passing rate. All these years such observations remained subjective in the absence of a method to quantify them using indexes like those used for MCQ and BAQ.

A glimpse into the history

In the optical mark reader (OMR) scored papers each item and its components are automatically given difficulty index and discrimination index values, which indicate how well the candidates performed in that item (DIFI), and whether the item was answered correctly more by the good students or by the poor students (DISi).

The quality of the item is generally judged using these indexes as they provide some numerical values for this purpose. These indexes are also useful in modifying the questions accordingly for future use (Miller et al. 2013; Linn et.al. 1995). Such a facility has been lacking in the examiner-scored assessments like MEQ, OSCE, Long case and Short case. In the absence of such indicators the performance and quality of these assessments remained an approximation.

Difficulty index of OMR-scored items indicates the fraction of candidates who got the answer correct in a particular item (Miller, et.al. 2013; Matlock-Hetzel, 1997). DIFI is independent for each item and assessment. It is not influenced by the quality of other questions in the same paper or other assessments in the same examination.

Discrimination index, on the other hand, is influenced by all the items in the same assessment. For calculating the DISi the examinees are sorted into an upper cohort (27.5% top scorers) and a lower cohort (27.5% bottom scorers) based on their scores (Miller et al., 2013; Kelly, T. L. 1939).

DISi is a comparison of the performance of upper cohort versus lower cohort for a particular item. In a good item the upper cohort is expected to perform better than the lower cohort. If the lower cohort did better than the upper cohort in an item, that item is considered flawed. Easy questions which are answered by most candidates will have high DIFI and low DISi. Difficult questions which are answered by only good students will have low DIFI and high DISi.

The formulae for assessment analysis follow the same principles. However, some modifications were made to make them usable for examiner-scored assessments (Puthiaparampil T. et.al 2017). See table 2. for the formulae. The main differences between item analysis and assessment analysis as described in the above paper are given below.
1. Item analysis uses a binary criterion for scoring. If the answer is correct, the item scores 1 (full mark), and if the answer is incorrect scores 0 (Nil). Examiner-scored assessments like MEQ, OSCE, long case or short case are scored from 0 to 100. Although it is possible to use a binary criterion (pass/fail) in such assessments also, the use of actual scores make the indexes more realistic. Therefore, actual scores were used for DIFi and DISi calculations in assessment analysis.

2. The upper and lower cohorts for generating DISi in item analysis is based on the scores in a single assessment like MCQ or BAQ. But for the same purpose in assessment analysis, the grand total marks of the entire examination are used.

The use of grand total marks for determining the upper and lower cohorts for generating DISi was justified on the following grounds in the article by Puthiaparampil T. et.al in 2017:

In the end of posting and professional examinations six assessments tools are employed to assess the performance and competence of the candidates, because a single tool is inadequate for this purpose. Not a single assessment is capable of testing the competence in all three domains like cognitive, psychomotor and affective, required for evaluating a medical doctor.

A step-by-step guide to generating DIFi and DISi of assessments.

The DIFi and DISi of assessments like MEQ, OSCE, long case and short case are calculated using Excel worksheets as no software is available for it.

- Tabulate all the marks obtained for all assessments by all the candidates in an Excel worksheet
- Convert all marks into out of 100
- Get the grand total marks for each candidate.
- Convert the grand total marks also into out of 100
- To avoid confusion, copy and paste-special into sheet 2 all 'out of 100' marks
- Sort the candidates top to bottom according to grand total marks
- Demarkate the top 27.5% (upper cohort) and the bottom 22.5% (lower cohort) using different colours
- If 27.5% happens to be a number with fraction, round it off to a whole number
- The number of candidates in the upper and lower cohorts should be the same
- Calculate DIFi: Total marks scored by all candidates in one assessment ÷ total marks offered (100 x number of candidates)
- Calculate DISi: (total marks scored by upper cohort in one assessment minus total marks scored by the lower cohort in that assessment) ÷ 100 x number of candidates in a cohort

How are these indexes interpreted?

Table 3 shows the index values and their interpretation as introduced by Ebel and Frisbe (Ebel and Frisbe 1972). DIFi indicates how easy or difficult the question was for the candidates based on their performance. DISi is rather a judgement on the quality of the question based on the performance of the good students compared to the performance of the poor students. However, these are not absolute values. The index values would change when the same question is tested on another batch of students. The judgement of the subject expert would still be the final
verdict.

Another method

Another method for evaluating the index values is to use the benchmark DISi provided by the grand total marks. Those assessments with DISi below the benchmark are considered inferior, while those with index values equal to or above the benchmark as superior. In the example given, the DISi of MEQ 1, 3, 4, MCQ and BAQ were higher than the benchmark. DIFI is independent and so needs no benchmark.

Take Home Messages

- Assessment analysis is a newly introduced concept with great potential to provide index values to examiner-scored assessments as well as whole papers like MCQ and BAQ.
- The index formulae are aligned to existing item analysis formulae with relevant modifications to make them usable for assessments.
- Assessment analysis uses the grand total scores of the entire examination to sort the candidates into upper and lower cohorts.
- The calculation of indexes can be easily done manually on Excel worksheets.

Notes On Contributors

Professor Thomas Puthiaparampil is the sole author of this article.

The original article "From Item Analysis to Assessment Analysis: Introducing New Formulae" had two co-authors apart from Thomas Puthiaparampil. They are Prof. Md Mizanur Rahman and Dr Isabel Fong Lim.

Acknowledgements

My thanks to the dean and deputy deans of The Faculty of Medicine and Health Sciences, UNIMAS for the approval of the project and permission to use the examination results for this project.

Bibliography/References

Ebel, R. L. (1972). Essentials of educational measurement. Englewood Cliffs, NJ: Prentice Hall (Retrieved from book on 16 January 2016)

Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. Journal of educational psychology, 30(1), 17. Retrieved from: http://scholar.google.com [Accessed on 17 January 2016].

https://doi.org/10.1037/h0057123
Matlock-Hetzel, S. (1997). Basic Concepts in Item and Test Analysis. Retrieved from http://scholar.google.com on 17 January 2016

Miller M. D., Linn R. L., Gronlund N. E., (2013). Measurement and assessment in teaching 10th Ed. Upper Saddle River, N J: Prentice Hall, 2013. Retrieved from book [Accessed on 26 January 2016]

Puthiaparampil T, Rahman M, Fong Lim I, From Item Analysis to Assessment Analysis: Introducing New Formulae, MedEdPublish, 2017, 6, [1], 7, 

https://doi.org/10.15694/mep.2017.000007

Taib, F., & Yusoff, M.S.B. (2014). Difficulty index, discrimination index, sensitivity and specificity of long case and multiple choice questions to predict medical students' examination performance. Journal of Taibah University Medical Sciences, 9(2), 110-114. Retrieved from http://scholar.google.com on 13 November 2016

https://doi.org/10.1016/j.jtumed.2013.12.002

## Appendices

### Table 1. Assessment tools employed

|   | End of posting exam | Professional exam | Marked /scored by |
|---|---------------------|-------------------|-------------------|
| 1 | MCQ True/False 20   | MCQ True/False 60 | Optical mark reader software |
| 2 | BAQ 10              | BAQ 50            | Examiners         |
| 3 | MEQ 4 (2 hours)     | MEQ 5 (2.5 hours) |                   |
| 4 | OSCE 10 stations    | OSCE 20 stations  |                   |
| 5 | Long Case 1         | Long Case 1       |                   |
| 6 | Short cases 2       | Short cases 3     |                   |

### Table 2. Formulae for assessment analysis and expansions of abbreviations

| Abbreviations and expansions                     | Formulae             |
|--------------------------------------------------|----------------------|
| TN = total number of candidates                  | DIFi = msT ÷ mmT     |
| nC = total number in a cohort                    | DISi = (UC-LC) ÷ nC x 100 |
| UC = upper cohort                                |                       |
| LC = lower cohort                                |                       |
| msT = marks scored total (all candidates)        |                       |
| mmT = maximum marks total (TN x 100)             |                       |
| DIFi = difficulty index                         |                       |
| DISi = discrimination index                     |                       |
Table 3. Interpretation of index values according to Ebel and Frisbe (1972)

| Difficulty index | Discrimination index |
|------------------|----------------------|
| 0 – 30% (0 – 0.30) | Difficult            | >0.40 | Excellent |
| 31 – 79% (0.31 – 0.79) | Moderate           | 0.2 – 0.39 | Good    |
| ≥80% (≥0.80)          | Easy                | 0.01 – 0.19 | Fair    |

Figure 1 - An Excel Worksheet showing how assessment analysis is done

| S. No. | MFQ1 | MFQ2 | MFQ3 | MFQ4 | MCQ | RAG | LC | SC | OSCE | GT |
|---------|------|------|------|------|-----|-----|----|----|------|----|
| 1       | 100  | 100  | 100  | 100  | 100 | 100 | 100| 100| 100  | 100 |
| 2       | 10   | 10   | 10   | 10   | 10  | 10  | 10 | 10 | 10   | 10  |
| 3       | 10   | 10   | 10   | 10   | 10  | 10  | 10 | 10 | 10   | 10  |
| 4       | 10   | 10   | 10   | 10   | 10  | 10  | 10 | 10 | 10   | 10  |
| 5       | 10   | 10   | 10   | 10   | 10  | 10  | 10 | 10 | 10   | 10  |

Declarations

The author has declared that there are no conflicts of interest.
