RESEARCH ARTICLE

SENSE OF ASSESSMENT FAIRNESS IN A MULTIDISCIPLINARY PLATFORM

*Faiz MMT Marikar

Staff Development Centre, General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka

ARTICLE INFO	ABSTRACT
	Assessment standardizing is a long fed issue to tackle the examiners fairness. It reduces variation among the student's evaluation and testing procedure that could adversely affect others. However, individual marking with a proper guidelines will make the assessment somewhat fair than the group marking or conference marking. Equivalent grid want make the evaluation equal. Although testing accommodations are by now standard practice in most large-scale testing programmes, for the most part these practices lie outside formal educational measurement theory. This article was building on recent research in group marking with multiple disciplinary which is much affective that the individual marking. Marking grid which was based on common matching examinees will give us high accuracy. The present focus is assessment for special populations, but it is argued that the principles apply more broadly.

INTRODUCTION

Fairness in examination is not an easy task, according to personal, geographical and even influential will leads to unfair test. Other types of unfairness are not easy to characterize. For example, in a vulnerable community if they asked to submit the report of marking within a short period of time want be an easy task which will leads to inferior the quality of marking. Furthermore in such situations conference marking will be the ideal and test bias can be minimize but not for the zero level. In this study we identified a problem of independent marking without a proper mechanisms will leads to trouble marking. Therefore introduced the group marking or conference marking for minimizes the test bias. What we found was the important related issue is how high the burden of proof should be administrated from the top management participations is very important.

Individual and group fairness

The 1999 Standards for Educational and Psychological Testing (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999) is intended to provide criteria for the evaluation of tests, testing practices, and the effects of test use, where the term test here is

used to specify a broad range of assessments including particular tests, scales, inventories, and instruments. Twelve criteria pertaining to fairness are given in the Standards. Nine refer to groups or subgroups, and three to individuals or aspects of individuals, suggesting the broad categories of group and individual fairness (Ferdman, 1989). The goal is to prevent influences irrelevant to the test to create advantages or disadvantages that result in higher or lower test scores (Messick, 1989). In sum, if a test or test item is equitable, it is presented to individuals under impartial conditions, meaning that no individual student is favored over another in demonstrating what they know or understand.

Helwig and Tindall (2003) carried out such an experiment with read-aloud accommodations for a mathematics test for both general and special education students to determine whether accommodations provided a better assessment of the proficiency of the latter. Another aspect of individual fairness involves treating test takers with dignity and sensitivity. This aspect of testing may have no counterfactual stated in terms of alternative tests or test conditions: It is no defence of a charge of unfairness in this regard to argue that examinees were treated badly but equitably. It is also no defence to demonstrate equity if the objectives upon which a test is based are themselves faulty Bouville (2008).

Corresponding author: Faiz MMT Marikar,
Staff Development Centre, General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka.

Group fairness

The categorization of individuals into groups must be done with caution, and this point is illustrated below with respect to race. In October 1997, the Office of Management and Budget (OMB) in the US released new categories for collecting data on race and ethnicity (OMB, 1997). Ethnicity is at least as nuanced as race, involving issues of language, race, place of origin, values, and heritage. The charge of unfairness is often, if not inevitably linked to social bias. Beyond these causes, there are also effects to consider, and this raises a central question in test fairness of how cause and effect are linked. Quantitative methods can improve an understanding of the link between cause and effect, but arguments solely based on the authority of statistical methods are both flawed and obfuscating (Camilli, 1993).

MATERIALS AND METHODS

In this study we have examined teachers' capabilities of test fairness in a given answer script. To emphasize the hypothesis fairness requires coverage of larger group and stills more issues when it comes for the individual assessment and it is fairly distributed among groups. One such is issue is how to make the marking scheme among individuals. Second, test fairness is examined in the group framework after the preparation of group marking scheme.

Questionnaire Used

Part of the mock examination answer script was given as a questionnaire to be analyzed by method of paper marking to the lectures and it is as follows.

"It is very difficult to get breast milk contaminated. Contamination of baby food leads to infection and poor growth. The final thing can be a weak nation. Europe's women don't like breast feeding. They say it can change shape of breast. No evidences for this situation. It does not drop with breast feeding mothers like breast feeding. They feel happy when child drinks. The child also feels happy. Multinational companies make mothers stop breast feeding. They give free milk to mothers to stop them their own breast milk. This was done to for them to get more money. We should not be fooled by propaganda's. Last year American University students boycotted a big American firm that makes infant food same reason. If American people can protest against their own company why we cannot do like that? We should encourage our mothers to breast feed as long as possible. That is why if they do like that lot of money can be used for other things and our country can benefit.

Breast milk is the best milk. When we discuss the statement we have to remember that mothers have being feeding their infants which own milk for ages. The west influence changed the practice in this country to last few years' decades and it in good that we go back to what those we have been used to a long time.

Breast milk was the best composition to suit the needs of humans. As the saying goes breast milk is best for kid which cow's milk is best for calves his requirements of the infants can be easily met if breast milk is used.

A big problem in infant is infections. Breast milk can counteract infection in childhood because pressure if gamma globulins. Many children end up in problems due to infections in childhood we prevent them with breast milk.

Recently researches in American shows that breast milk promote better growth. Better growth in infancy mean better growth citizen. Correct temperature of breast milk of human is also important. The baby has same temperature like mother and he wishes drinking it."

Research Context and Participants

Twelve lecturers took part in this study. Six faculties including Law, Medical, Allied Health Sciences, Engineering, Management and Defence equally represent from General Sir John Kotelawela Defence University, Sri Lanka, which was located in Ratmalana, Sri Lanka. All participants were employed at the university as lectures. The University is a socially and economically diverse community in Western Province of the country and this is the one and only Defence University in this country. They have been expected to know the answers to the questions asked in the questionnaire used in this study.

Administration of Questionnaire among the Students

The hard copy of answer script which is used as a questionnaire was administered among each of the respective participants at the University teacher training programme. Given 30 minutes and within given period they have to mark the answer script. Care was taken to avoid exchanging the participants now considered as examiners information or ideas.

Data Analysis

To analyze the marks of the answer script, we compared informal reasoning displayed by individuals representing high and low level of marking with the standard deviation. The validity of the marking was independently assessed by two observers. For statistical analysis, we transformed all our data using the basic statistical analysis package.

RESULTS AND DISCUSSION

There were two groups and each group was consisted with 6 members in this study. In first marking individually was given a chance to mark the answer script according to their wish. There was no moderation, observation or marking scheme to give marks. In this study one answer script was photocopied and given to all participants. Group one mean marked value is 56.17 and group two it is 53.50 which is lower than the first group. Interesting there were high Standard deviation in group one which is 23.43 which is significant high with the group two which comes 10.672. The range of marking in group one is 35-95 and group two 35-65 which gives a high standard deviation. The value was significant because they are not in the same discipline.

After analyze the results it was shown to the examiners and highlighted the importance of marking scheme. Second phase of study was based on marking grid and asked them to prepare a marking scheme and do the marking. Interestingly we found there is homogenous of marking among two groups and standard deviation was closer and there are no significant differences and value is 3.764 in group one and 3.817 in group two. Mean values were 59.17 and 40.83 in group one and group two respectively and values are significant deviate. Furthermore we checked why the reason behind of this type of variation is. Observations are very clear two marking schemes developed from the two groups are not compatible with each other.

Table 1. Frequency analysis of marks

	FirstGpA	SecondGpA	FirstGpB	SecondGpB
Mean	56.17	59.17	53.50	40.83
Median	52.50	60.00	56.50	40.50
Mode	65	60	35(a)	45
Std. Deviation	23.413	3.764	10.672	3.817
Range	60	10	30	10
Minimum	35	55	35	35
Maximum	95	65	65	45

a Multiple modes exist. The smallest value is shown

Table 2. Marking schemes

Group one Marking Scheme	Group two Marking Scheme
Introduction (20 Marks) about breast milk Sri Lankan situation Foreign situation	Organization and Flow (20 Marks)
Importance of brest feeding (50 Marks) protect against infection Provide Vitamins/Ca best milk	Language and Grammer(20 Marks)
Research report (10 Marks) Present past	Introduction (15 Marks)
Present threat (5 Marks) Milk companies provides only milk	Importance (15 Marks)
Actions against these importers (5 Marks) USA/China	If it is not (15 Marks)
Our duty (10 Marks) Present role of promoting Breast Feeding	Conclusion (15 Marks)

According to Table 2 two different types of marking schemes were observed. Observations were clear that why there were two deviations in second marking. Comprehensive marking scheme was provided by the Group one and therefore they received the highest mean value of 60. Group two where they went wrong is they had given highest portion for firstly organization and flow secondly Language and Grammar which is 20% each which total almost 40%. If it is languages agree but not for the technical papers like this.

Conclusion

Two ways of answer script related to evaluation fairness have been briefly identified and explained in this study. First, a number of pitfalls were identified in individual marking for a same answer script because their specialties are different. Because the expectations were not met in individual marking tried in group discussion and make a common marking schemes. Two groups had discussions separately and prepared two marking schemes. According to the marking second marking was done and observed similar standard deviation among two groups. The results were amazing and significant with individual marking. Still there is something to change because two different marking grids gave two different marks for same answer, which is doubtful. Further analysis revealed that the difference is mainly because different specialties lectures prepared marking grid. Finally both groups were together making one marking scheme helped all to get a common mark (Data was not shown). This exercise was very important to teach the lectures about their capabilities of developing their own marking schemes with the consultations with the senior academics.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999. Standards for educational and psychological testing (2nd ed.). Washington, DC: American Educational Research Association.

Bouville, M. 2008. The obsession with exam fairness. Retrieved from http://www.mathieu. bouville. name/education-ethics/Bouville-exam-fairness.pdf

Camilli, G. 1993. The case against item bias techniques based on internal criteria: Do item bias procedures obscure test fairness issues? In P. W. Holland & H. Wainer (Eds.), Differential item functioning: Theory and practice (pp. 397–417). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ferdman, B. M. 1989. Affirmative action and the challenge of the color-blind perspective. In F. A. Blanchard & F. J. Crosby (Eds.), Affirmative action in perspective (pp. 169–176). New York, NY: Springer-Verlag.

Messick, S. 1989. Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 13–103). New York, NY: American Council on Education and Macmillan.

Office of Management and Budget. 1997, October 30. Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity (Federal Register Notice 62FR58782- 89). Washington, DC: Author.

****