Available Online at ESci Journals
International Journal of Educational Studies
ISSN: 2312-458X (Online), 2312-4598 (Print)
http://www.escijournals.net/IJES

GENDER STEREOTYPES INFLUENCE ON ASSESSMENT IN A MULTIDISCIPLINARY PLATFORM

Faiz M. Marikar
General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka.

A B S T R A C T

Gender based evaluation in assessment standardizing is a long fed issue to tackle the examiners fairness among gender stereotypes. It reduces variation among the student's evaluation and testing procedure that could adversely affect others. However, individual marking with who became lectures after been a mother, father, pregnant, male and female with a proper guideline will make the assessment somewhat fair than the group marking or conference marking. Using a common marking scheme in evaluation is not always a correct approach. Although testing accommodations are by now standard practice in most large-scale testing programmes even though gender will somewhat influence on stereotype questions, for the most part, these practices lie outside formal educational measurement theory. This article was building on recent research in group marking with pregnant women, mother, father, male who are not married and females who are not married with multiple disciplinary approaches will influence the final raw marks. In this study males who are not married gave the highest marks for the importance of breast feeding. The rest were given lower marks. After that group marking was assessed which is much effective than the individual marking. Marking grid which was based on common matching examinees will give us high accuracy. The present focus is an assessment for special populations, but it is argued that the principles apply more broadly.

Keywords: Gender influence, Individual marking, Conference marking, Group marking, Assessment.

INTRODUCTION

Fairness in the examination is not an easy task, according to personal, geographical and even influential factor will lead to an 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 it is not an easy task which will lead to inferior the quality of marking. Furthermore, in such situations conference marking will be the ideal and test bias can be minimizing but not for the zero level. In this study we identified a problem of independent marking without a proper mechanism will lead to trouble marking. Furthermore, we have identified a question related to milking and observed the results and it is gender biased. 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 participation is very important.

When examination assessing the capacity of individuals, attention is paid to the knowledge, skills and attitudes that each person has regarding gender equality and the empowerment of women and the integration of these to their examination work. The information that will be gathered entails how much they know and understand about particular concepts (for example, gender equality, women's empowerment, etc. towards paper marking), policies (such as organizational gender strategy, gender mainstreaming strategy) and procedures (how gender equality should be reflected in paper marking), as well as how capable they are of implementing all these and other processes.

Individual and group fairness: The 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

* Corresponding Author:
Email: faiz.marikar@fulbrightmail.org

is used to specify a broad range of assessments including tests, scales, inventories, and instruments. Twelve criteria pertaining to fairness are given in the Standards. Nine refers to groups or subgroups, and three to individuals or aspects of individuals, suggesting the broad categories of the 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 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 (Chowdhry; 2016). 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 arguing 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).

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, and 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. Fairness requires coverage of larger group and still more issues when it comes for the individual assessment and it is fairly distributed among groups. One such issue is how to make the marking scheme among individuals according to gender. 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 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 the shape of the breast. No evidence 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 propagandas. 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 lotof 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 been feeding their infants which own milk for ages. The west influence changed the practice in this country to last few years and it in the 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 an 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.

Recent findings in American shows that breast milk promotes better growth. Better growth in infancy means better growth citizen. The optimum temperature of breast milk of human is also important. The baby has same temperature like mother and he wishes to drink it."

Research Context and Participants: Twenty students 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 responders were employed at the university as lectures.

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 ideas.

Data Analysis: To analyze the marks of the answer script, we compared informal reasoning displayed by individuals representing the highland 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 20 responders in the group of evaluation from different speciality and from different faculties. Initially, questionnaire was given among them and asked them to give me the raw marks. I have not given them any marking schemes to evaluate. 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. First round males who are not married were given the highest scores which is 60% and lowest grade 45% was given by the females who are not married. Importance of the breast feed was highly evaluated by males and not be females. Pregnant women, mother and the father were given somewhat similar and its values were 55, 53 and 52 which is somewhat similar (Table 1). When I found that female gender is reluctant to breast feed their babies. After that I have grouped them into 4 according to the gender male, females, more males and less females, less males and more females and asked them to prepare a marking scheme and re-score them according to the marking scheme. There is no big difference among the categories, but pregnant women gave more marks than the other groups (Table 1).

Table 1. Frequency analysis of marks according to gender stereotypes.

	Male1	Female1	Pregnant1	Father1	Mother1	Male2	Female2	Pregnant2	Father2	Mother2
Mean	45.14	60.00	55.00	52.33	53.33	46.14	58.80	60.00	49.00	52.00
Median	45.00	60.00	55.00	55.00	45.00	47.00	59.40	60.00	52.00	45.00
Mode	40	60	55	40	40	34	52	60	40	45
Std. Dev.	11.936	6.325	0	11.240	18.930	9.924	5.845	0	7.937	12.124
Range	37	20	0	22	35	27	16	0	15	21
Minimum	25	50	55	40	40	34	52	60	40	45
Maximum	62	70	55	62	75	61	68	60	55	66
Average	60	45	55	52	53	59	46	66	49	52

Group marking is somewhat interesting to be discussed, group one, two, three and four mean marked values are as follows 55.4, 49.25, 45.40 and 56.00group three which is lower than the other groups (Table 2). [Interesting there was high Standard deviation in group three which is 13.22 which is significantly high with the other groups which lead to the difference to obtain the least mean value. The range of marking in group three is very significant which 34 is not acceptable. The probable cause for value significant because they are not in the same discipline and not even belongs to the same gender.

After analyzing the results, it was shown to the examiners and highlighted the importance of marking scheme (Table 3). The second phase of the study was based on marking grid and asked them to prepare to mark scheme and do the marking. Interestingly we found there is homogenous of marking among two groups and the standard deviation was closer and there are no significant differences in the group 1 and group 4 (Table 2). Mean values were 55.40 and 56.60 in group one and group four respectively and marks given from both groups are in a homogenous. Furthermore, we checked why the reason behind of this type of variation is.

Table 2. Frequency analysis of group marking.

	Group1	Group2	Group3	Group4
Mean	55.40	49.25	45.40	56.60
Median	55.00	49.50	40.00	56.00
Mode	48	37	40	49
Std. Deviation	5.320	10.012	13.221	6.693
Range	13	24	34	17
Minimum	48	37	34	49
Maximum	61	61	68	66
Average	55	49	45	56

Table 3. Marking schemes

Group one Marking Scheme	Group two Marking Scheme	Group three Marking Scheme	Group four Marking Scheme
Introduction (20 Marks) about breast milk Sri Lankan situation Foreign situation	Organization and Flow (20 Marks)	Introduction (20 Marks)	Content (60 Marks) Relevance (10) Analysis (10) Adequate of facts (30) Accuracy of facts (10)
Importance of breast feeding (50 Marks) protect against infection Provide Vitamins/Ca best milk	Language and Grammar (20 Marks)	Problem Statement (20 Marks)	Language (20 Marks)
Research report (10 Marks) Present past	Introduction (15 Marks)	Purpose (20 Marks)	Presentation (20 Marks)
Present threat (5 Marks)	Milk companies provides only milk Importance (15 Marks)	Significance (10 Marks)	
Actions against these importers (5 Marks) USA/China	If it is not (15 Marks)	Statistics (20 Marks)	
Our duty (10 Marks)	Present role of promoting Breast Feeding Conclusion (15 Marks)	Style (10 marks)	

Observations are very clear two marking schemes developed from the two groups are not compatible with each other.

According to Table 3four 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 three, therefore, they received the highest mean value of 56. Group two and three where they went wrong is they had given highest portion for the unnecessary section.

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 same answer script because their specialities are different. Second with or without knowing the brain gives a somewhat similar mark

according to gender stereotypes. Because the expectations were not met in individual marking tried in group discussion and make a common marking scheme. Four groups had discussions separately and prepared four 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 four different marking grids gave four different marks for the same answer, which is doubtful. Further analysis revealed that the difference is mainly because different specialities lectures prepared to mark grid. 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 detection techniques based on internal criteria: Do item bias procedures obscure test fairness issues. *Differential item Functioning*, 397-413.

Chowdhry, M. A., (2016). Perspective on productive classroom practices in Science, *International Journal of Educational Studies* 3(3), 78-86.

Ferdman, B. M. (1989). Affirmative action and the challenge of the color-blind perspective. In *Affirmative action in perspective* (pp. 169-176). Springer, New York, NY.

Helwig, R., & Tindal, G. (2003). An experimental analysis of accommodation decisions on large-scale mathematics tests. *Exceptional Children*, 69(2), 211-225.

Messick, S. (1989). Validity In. R. Linn (Ed.) Educational measurement (13-103). New York, NY: American Council on Education and Macmillan.

Office of Management and Budget. (1997). *Revisions to the standards for the classification of federal data on race and ethnicity.* Federal Register Notice (62FR58782- 89). Washington, DC.