On the Relationship between Test Format, Attitudes Towards and Performance in a Statistics Test

Fadia Nasser, Ph.D.
Tel Aviv University, School of Education & Beit Berl College
Ramat Aviv 69978
Tel Aviv
Israel
fadia@post.tau.ac.il

1. Introduction

In view of the complexity of conveying statistical knowledge, especially to students in the humanities, various teaching strategies and various assessment methods have been recommended in order to improve and to adequately assess students' performance (Cobb, 1992). Gal and Garfield (1997) stressed that assessment methods should not only assess specific skills but also identify students' understanding of the major ideas in statistics and their aptitude to choose and apply statistical methods properly in real life situations. Multiple-choice questions and short answer open-ended questions are commonly used to assess students' performance in introductory statistics courses. Despite arguments against the viability of such item formats to assess students' statistical reasoning processes, when adequately constructed they are able to assess students' knowledge, their understanding and their ability to use the acquired knowledge in new situations (Barrnett-Foster & Nagy, 1996; Gal & Garfield, 1997).

As for students' preferences and attitudes towards test formats, the existing literature indicates that, in general, multiple-choice and short open-ended questions are most preferred and student scores were most determinant of student preferences regarding tests (Yunker, 1999).

This study examined the interrelationship between test format, students' attitudes and performance in an introductory statistics test comprised of multiple-choice, open-ended and output-based questions.

2. Method

The sample consisted of 105 graduate students in the school of education of a major university in Israel. The massive majority of participants were female students (84%). Introductory statistics for all participants was a required course.

Attitudes towards each of the three types of test formats were measured by 12 semantic differential scales with 7 points, where 1 indicates the most positive and 7 the most negative attitude towards the corresponding test format. These items tapped emotional, , intellectual and appraisal reactions to the test. Cronbach's α coefficients corresponding to students' attitudes towards multiple-choice, open-ended and output item formats were .88, .87, and .88 respectively.

Students' performance was measured by an achievement test, which included 24 multiple-choice items, 9 open-ended items, and one SPSS followed by four open-ended questions to be answered on the basis of information included in the output.

3. Results and discussion

Table 1 presents the means and standard deviations for participants' attitudes toward each of the three test formats and for their scores on each of these formats. Results indicate that participants expressed favorable attitudes towards all three formats. However, on average, they were more favorable of multiple-choice questions, and expressed no differential preference for either one of the two open-ended types. ANOVA results indicated no significant differences in participants' attitudes towards the three assessment formats.

As to achievement, the mean score on the multiple-choice questions was the lowest, while the mean score on the open-ended questions was the highest. The mean score on the output

Table 1

 Means and standard deviations for participants' attitudes toward each of the three assessment formats and for their scores on each of these formats

Variable	Mean	SD
Attitudes		
Towards multiple-choice format (1-7)	2.88	.94
Towards open-ended format (1-7)	3.03	.96
Towards output format (1-7)	3.07	1.09
Achievement		
Multiple-choice (0-100)	80.61	13.08
Open-ended (0-100)	93.03	9.63
Output (0-100)	86.00	18.91

questions was midway between the scores on the two other formats. In other words, on average, students scored lowest on the test format that they liked most. Results from ANOVA indicated that the mean scores in the three assessment formats differed significantly. Follow up post hoc analysis indicated that the mean scores on each pair of the three assessment formats were also statistically significant.

As to the correlation between attitudes towards and scores in test format, there was an inverse correlation between attitudes and scores, for both the multiple-choice and output formats (-.397 and -.196 respectively). A negligible correlation was found between attitudes towards and scores in the open-ended assessment format. Results from this study demonstrate the viability of using various assessment formats to adequately assess students' statistical knowledge and reasoning processes. In addition, it was shown that the most favorable format is not necessarily the one in which students perform the best.

REFERENCES

Barrnett-Foster, D., & Nagy, P. (1996). Understanding student strategies to test questions of varying format. Higher Education, 32, 177-198.

Cobb, G. W. (1992). Teaching statistics. In L. Steen (Ed.), Heeding the call for change: Suggestions for curricular action. Washington, DC: The Mathematics Association of America.

Gal, I., & Garfield, J. (1997). Curricular goals and assessment challenges in statistics education. In I. Gal, & J. Garfield (Eds.), The assessment challenge in statistics education. Amsterdam: IOS Press.

Yunker, D. B. (1999). Adding authenticity to traditional multiple-choice test format. Education, 120, 82-87.

RÉSUMÉ

La présente étude examine la relation entre les attitudes des étudiants envers différents formats d'évaluation et leur performance dans un cours d'introduction en statistiques. Les formats d'évaluation se réfèrent á trois genres de questionnaire: á choix multiples, ouvert, et output-based. L'échantillon comprenait 105 étudiants en maîtrise, la majorité des étudiantes. L'outil de recherche comprenait deux parties: un questionnaire mesurant les attitudes envers les trois formats d'évaluation et un examen en statistiques comprenant des questions dans les trois formats d'évaluation. Les résultats indiquent que les étudiants ont favorisé le questionnaire á choix multiples, alors que leurs notes par ce moyen furent les plus basses. Des différences significatives ont apparu entre les moyennes des notes obtenues par les trois formats d'évaluation. De plus, il y avait une corrélation inverse entre attitudes et notes pour le formats á choix multiples et output-based. Les résultats de la présente étude montrent que l'utilisation de différents moyens d'évaluation est viable et peut contribuer á mieux évaluer les connaissances et les processus de raisonnement des étudiants en statistiques. De même, l'étude montre que le meilleur format d'évaluation n'est pas nécessairement celui dans lequel les étudiants performent le mieux.