No : 136

Revisiting the Philippine Aptitude Classification Test:
Analysis of Potentially Biased Test Items

Cecile Gatchalian, Center for Educational Measurement, Inc., Philippines, cgatchalian@cem-inc.org.ph
Armi Lantano, Center for Educational Measurement, Inc., Philippines, alantano@cem-inc.org.ph

Key words: test fairness, differential item functioning, item bias, gender differences, linguistic differences

Abstract

Differential item functioning (DIF) indicates that an item is potentially biased. To gain a better understanding of the behavior of the subgroups on potentially biased items, the distractor response analysis can be conducted. In this procedure, the incorrect alternatives are examined for differences in patterns of responses among different subgroups of a population. Bias is inferred to be present in an item when the subgroups are differentially attracted to the distractors of the item.

This follow-up study on the fairness of the Philippine Aptitude Classification Test (PACT), a battery of tests measuring different kinds of aptitudes or abilities, aims to gain more substantive evidence of item bias in the test. It further investigates the two PACT subtests identified in the previous study as having the most number of items with DIF, namely, the Perceptual Acuity and the Verbal Filipino. The Rasch model technique was applied on another sample of examinees, and distractor response analysis was used to examine the response pattern across the incorrect options of items with DIF among the gender and the geographic subgroups of the PACT takers. The analysis of response patterns provided insights on group differences, and implications to test construction were drawn.

Introduction

The consideration of fairness is important in sound testing practice. A test should be constructed such that differences between groups' test scores are due to differences in what the test measures. Test developers have to ensure that items possessing unfair biases will be detected and replaced by items that are less likely to contain such biases. Developing tests that are as fair as possible contributes to the societal goal of equal opportunities for all.

A well-made test addresses the concern on fairness from test design to testing outcomes. It is subjected to careful review and empirical checks to minimize bias. Bias happens when the psychometric properties of a test or when the manner in which it is used results in different meanings for scores earned by members of different subgroups (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 1999). Evidence of bias may be found by studying how the test scores are used or by examining the test items themselves.

Differential item functioning (DIF) is an indicator that an item is potentially biased. To gain a better understanding of the behavior of subgroups on potentially biased test items, Osterlind (1983) recommends that, after an initial investigation of DIF in all of the test items, a distractor response analysis be conducted on the items with DIF. In this procedure, the incorrect alternatives are examined for differences in patterns of responses among the different subgroups. Bias is inferred to be present in an item when the subgroups are differentially attracted to the distractors of the item.

Background

The Philippine Aptitude Classification Test (PACT) is a battery of tests that measure different kinds of aptitudes or abilities important in college and vocational work. The preliminary investigation of its fairness (Franco & Lantano, 2009) resulted in the detection of items with gender and geographic DIF. Gender and geographic location were chosen as subgroupings as local studies (Caoli-Rodriguez, 2007; Hicap, 2006 as cited in Franco & Lantano, 2009) showed significant achievement differences among Filipino students across these variables.

Among the three DIF-detection procedures used, the Rasch model technique was the most sensitive. Out of 210 items, it identified 85 items with gender DIF and 55 items with geographic DIF. The subtest Perceptual Acuity contained the most items with gender DIF, and the subtest Verbal Filipino contained the most items with geographic DIF.

The present study aims to gain more substantive evidence of item bias in the PACT. It further investigates the two PACT subtests identified in the previous study as having the most number of items with DIF, namely, the Perceptual Acuity and the Verbal Filipino. Applying statistical procedures, it seeks to find out how the test items behave for the different subgroups. Specifically, it describes the items that are found to be easier for the different subgroups and determines if the difficulty of an item for a subgroup stems from an internal characteristic of the item. Based on technical criteria, potentially biased items will be detected and corrected or replaced by items that are less likely to contain bias.

Methods

Subjects

A sample of 2,008 third year high school students was randomly chosen from the total PACT takers of school year 2008-2009. The sample has an approximately 6:4 ratio of both male – female and National Capital Region (NCR) – other regions (non-NCR) subgroups.

The NCR is one of the 17 regions of the Philippines. The division of the Philippine archipelago into regions is based on the geographical, cultural, and ethnological characteristics of the provinces (Philippine Travel Photos, n.d.). The NCR is mainly the Metropolitan Manila. It is the center of politics, economics, society, and culture of the country.

Instrument

The PACT is a battery of tests that measure different kinds of aptitudes or abilities important in college and vocational work. The subtest Perceptual Acuity measures the ability to visually detect the pattern of change or the differences in given figural or semantic stimuli. It consists of two item types: Figure Series and Proofreading, each with 15 items. The subtest Verbal Filipino measures the ability to understand Filipino, the national language of the Philippines. It consists of two item types: Talasalitaan (vocabulary) and Mga Salitang Magkaugnay (verbal analogies). Talasalitaan is further subdivided into Kasingkahulugan (synonym) and Kasalungat (antonym). Verbal Filipino consists of 19 items: 6 Kasingkahulugan, 4 Kasalungat, and 9 Mga Salitang Magkaugnay.

Procedure

The Rasch model technique was used to identify the items that exhibit DIF. The subtests Perceptual Acuity and Verbal Filipino were analyzed for gender and geographic DIF, respectively. With the Rasch model, the item difficulty of each item was calibrated separately for each of the gender (male vs. female) and geographic location (NCR vs. non-NCR) subgroups. The difficulty indices were tested for statistical significance using the chi-square. The items that showed a significant difference between the subgroups were flagged as potential DIF items.

Moreover, the items that consistently exhibited DIF in the initial study and in this study were analyzed by examining the response pattern across the incorrect options among the gender and the geographic location subgroups of the PACT takers. Employing the distractor response analysis, it determined significant difference between two groups. It involved the following steps: (a) preparing a matrix of choice response alternatives for the test items under study, (b) placing the data in a series of 2x2 contingency tables for significance testing, and (c) hypothesis testing. In this study, the Mantel-Haenszel procedure was used to test if there existed significant difference in the proportions selecting distractors on a test item between identified subgroups of the population.

Results

DIF Analysis

The analysis on Perceptual Acuity for the gender subgroups resulted in the identification of 14 items with DIF. (Refer to Table 1.) The results of 13 of these items were consistent with those in the initial study. Six (6) were Figure Series items, while seven (7) were Proofreading items. All of the Figure Series and one Proofreading items were consistently easier for males. Whereas, most of the Proofreading items (6 out of 7) were consistently easier for females. These results indicate that Figure Series items were more difficult for female examinees, whereas Proofreading items were more difficult for male examinees.

Table 1. Summary of Gender DIF Analysis

Item No.	Subtest	Item Difficulty			Easier for	Comparison With Previous
		Male	Female	Difference		
3	Figure Series	-1.08	-0.63	-0.45*	male	consistent
5	Figure Series	-0.18	0.23	-0.40*	male	consistent
6	Figure Series	-0.12	0.22	-0.34*	male	consistent
7	Figure Series	-0.23	0.05	-0.28*	male	consistent
8	Figure Series	-0.49	-0.13	-0.35*	male	consistent
9	Figure Series	-0.27	0.10	-0.37*	male	consistent
16	Proofreading	-1.42	-1.81	0.39*	female	consistent
18	Proofreading	-1.07	-1.50	0.43*	female	consistent
19	Proofreading	-0.60	-0.93	0.33*	female	consistent
21	Proofreading	-0.57	-0.99	0.42*	female	consistent
22	Proofreading	-0.48	-0.95	0.46*	female	consistent
25	Proofreading	-0.50	-0.86	0.37*	female	consistent
29	Proofreading	0.98	1.28	-0.30*	male	consistent
30	Proofreading	1.41	1.74	-0.32*	male	

* Significant at 0.01 alpha level

On the other hand, the analysis on Verbal Filipino for the geographic subgroups resulted in the identification of 11 items with DIF. (Refer to Table 2.) The results of 10 of these items were consistent with those in the initial study. Both Talasalitaan (vocabulary) and Mga Salitang Magkaugnay (verbal analogies) have 5 items with DIF. Majority of these items (8 out of 10) were consistently easier for the NCR subgroup.

Table 2. Summary of Geographic DIF Analysis

Item No.	Subtest	Item Difficulty			Easier for	Comparsion With Previous
		NCR	non-NCR	Difference		
1	Talasalitaan	-1.79	-1.34	-0.45*	NCR	Consistent
4	Talasalitaan	0.18	1.54	-1.36*	NCR	Consistent
9	Talasalitaan	-0.93	-0.36	-0.58*	NCR	Consistent
10	Talasalitaan	-0.91	-0.54	-0.37*	NCR	Consistent
11	Talasalitaan	-1.09	-1.62	0.53*	non-NCR	Consistent
19	Mga Salitang Magkaugnay	-1.06	-0.19	-0.87*	NCR	Consistent
20	Mga Salitang Magkaugnay	-0.15	0.74	-0.89*	NCR	Consistent
21	Mga Salitang Magkaugnay	-0.09	0.27	-0.36*	NCR	Consistent
22	Mga Salitang Magkaugnay	-0.67	-0.22	-0.45*	NCR	Consistent
23	Mga Salitang Magkaugnay	-0.24	-0.57	0.32*	non-NCR	Consistent
25	Mga Salitang Magkaugnay	0.70	0.41	0.29*	non-NCR	

* Significant at 0.01 alpha level

Distractor Response Analysis

Results of the distractor response analysis showed that 11 of the 13 items with gender DIF have significantly different response patterns for male and female examinees. (Refer to Table 3.) All of the Figure Series items and one Proofreading item were found biased against

female examinees. In these items, a significantly large number of female examinees were attracted to the incorrect options, indicating unfamiliarity with the concept reflected by the items. Conversely, most of the Proofreading items (6 out of 7) were found biased against male examinees. This indicates that male examinees have a significant tendency to favor other options over the correct options in these items. Hence, male examinees exhibit less familiarity with the concept reflected in the said items.

Table 3. Summary of Distractor Response Analysis for Gender

Item Number	Subtest	Correct Option	Easier for	Attractive Distractor(s) to the Other Group
3	Figure Series	3	Male	4,5
5	Figure Series	4	Male	2
6	Figure Series	3	Male	Not significant
7	Figure Series	1	Male	Not significant
8	Figure Series	5	Male	2
9	Figure Series	2	Male	5
16	Proofreading	1	Female	2,3,5
18	Proofreading	1	Female	2,3
19	Proofreading	3	Female	1,2
21	Proofreading	2	Female	1,3
22	Proofreading	2	Female	1,3
25	Proofreading	2	Female	1,3
29	Proofreading	3	Male	1

Based on the result of distractor response analysis on Verbal Filipino, only those items with geographic DIF favoring the NCR group were found significant. (Refer to Table 4.) Both items found to be easier for the non-NCR subgroup did not have significantly different response pattern when compared with their NCR counterpart. Hence, Verbal Filipino significantly favored the NCR subgroup.

Table 4. Summary of Distractor Response Analysis for Geographic Location

Item Number	Subtest	Correct Option	Easier for	Attractive Distractor(s) to the Other Group
1	Talasalitaan	3	NCR	2,4
4	Talasalitaan	1	NCR	3,4,5
9	Talasalitaan	1	NCR	4
10	Talasalitaan	3	NCR	2,4
11	Talasalitaan	4	non-NCR	Not significant
19	Mga Salitang Magkaugnay	5	NCR	3
20	Mga Salitang Magkaugnay	2	NCR	3
21	Mga Salitang Magkaugnay	5	NCR	4
22	Mga Salitang Magkaugnay	3	NCR	1,5
23	Mga Salitang Magkaugnay	4	non-NCR	Not significant

The distractor response analysis on Verbal Filipino further revealed 8 items significantly showing bias between the NCR and the non-NCR subgroups. In these items, one to three

distractors obtained significantly large number of responses from the non-NCR subgroup. Thus, these results indicate that the NCR is more familiar with the concepts reflected in the said items.

Three examples of test items with gender or geographic bias are found in the Appendix. The choice-response matrix of each item is also presented.

Discussion

Gender DIF in Perceptual Acuity

The Perceptual Acuity subtest contains two item types: Figure Series and Proofreading. The Figure Series items consist of visual stimuli, while Proofreading items consist of verbal stimuli. The Figure Series items were easier for males, while the Proofreading items were easier for the females.

The behavior of the gender groups on the two different item types is consistent with the findings of scientific research on the differences between the genders in regard to the brain and cognitive functioning (Bush, n.d.; Weiman, 2004). Males are found to be right-hemisphere dependent, specializing in visual tasks, such as the Figure Series items. On the other hand, females have been shown to be left-hemisphere dependent, performing better in verbal tasks. Thus, Proofreading is easier for them.

The difference in the performance of the genders on the Perceptual Acuity subtest is illustrated in the sample analyses of a Figure Series and a Proofreading items in the Appendix. In the Figure Series item, significantly more males chose the correct answer, while significantly more females chose two distractors. In the Proofreading item, the behavior of the genders reversed. Significantly more females correctly answered the item, while significantly more males chose three distractors.

Geographic DIF in Verbal Filipino

The Verbal Filipino subtest was easier for the NCR examinees than for the non-NCR examinees. This result may be attributed to the diversity in the linguistic background of the Filipino people. There are 175 individual languages used in the different regions of the country (Philippines, n.d.). Filipino, the national language, is primarily based on Tagalog, which is just one of the eight major languages of the Philippines (Santiago & Tiangco, 1991). The facility of NCR examinees in Verbal Filipino can be explained by the fact that Tagalog is mainly spoken in the region. Filipino is less familiar in non-NCR as there are other languages dominant in these regions. Non-NCR examinees may not have sufficiently acquired the Filipino language, resulting in their low performance on the test.

The unfamiliarity of the non-NCR examinees with some of the words in the Verbal Filipino subtest is illustrated in the sample analysis of a Mga Salitang Magkaugnay (verbal analogies) in the Appendix. A significantly smaller percentage of non-NCR examinees correctly answered the item. Moreover, two distractors were significantly chosen by more non-NCR examinees. These two words are also related to the third word in the stem, but only the correct option is related to the third word in the same way the second word is related to the first word in the stem. The non-NCR examinees seem to have difficulty in determining the fine distinctions between the options.

Conclusion

In this study, differentially functioning items in Perceptual Acuity and Verbal Filipino were further analyzed. The results illustrated gender differences in cognitive functioning and diversity in the linguistic background of the Filipino people.

Through the distractor response analysis, bias in several items was confirmed. In some items, a single distractor is significantly more attractive to a subgroup. These distractors have to be replaced in order to correct the bias. In a couple of cases, an item has two or more significantly more attractive distractors. These items will have to be replaced.

In addition, the following implications to test construction can be drawn:

(1) In developing test items, the inclusion of visual stimuli, even in achievement tests, should be considered so as not to disadvantage male examinees.

(2) In developing tests in Filipino, the use of words familiar nationwide has to be seriously considered.

Test fairness is one of the standards for educational and psychological tests. The selection of items based on discrimination and difficulty indices alone is not sufficient to ensure this. DIF and distractor analysis provide additional information on which items are appropriate to the target examinees.

References

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

Bush, T. (n.d.). Gender differences in cognitive functioning: a literature review. Cyber-Journal of Sport Marketing. Retrieved June 7, 2010, from http://fulltext.ausport.gov.au/fulltext/1997/cjism/v1n1/tbush1.htm

Franco, L. M. & Lantano, A. S. (2009). Study of fairness in the Philippine Aptitude Classification Test. Paper presented at the 35th IAEA Annual Conference. Retrieved March 15, 2010 from http://static.icms.com.au/iaea2009/papers/323.doc

Osterlind, S. J. (1983). Test item bias. Newbury Park, CA: Sage.

Philippine Travel Photos. (n.d.). Retrieved June 19, 2010, from http://www.philippine-portal.com/regional -information/regions.html

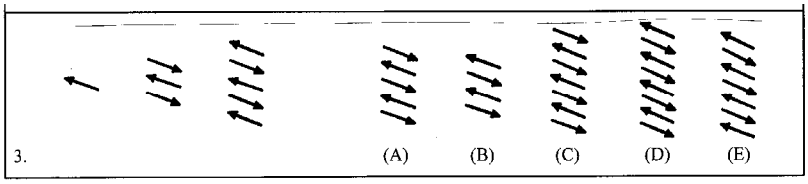
Philippines. (n.d.). Retrieved June 24, 2010, from http://en.wikipedia.org//wiki/Philippines

Santiago, A. O. & Tiangco, N. G. (1991). Makabagong balarilang Filipino. Manila, Philippines: Rex Book Store.

Weiman, H. (2004, May 10). Gender differences in cognitive functioning. Retrieved June 7, 2010, from http://homepages.luc.edu//~hweiman/Gnederdiffs.html

Appendix
Example of Biased Items with Corresponding
Choice-Response Matrix

FIGURE SERIES ITEM



Choice-Response Matrix

	Response Choice					
Groups	A	B	C*	D	E	Omits
male	3.44%	5.40%	70.33%	9.02%	11.07%	0.74%
female	3.78%	3.55%	64.26%	14.09%	13.52%	0.80%

PROOFREADING ITEM

16. Security Pacific Bank
333 South Hope Street
Los Angeles, California

(A) Security Pacific Bank
333 South Hope Street
Los Angeles, California

(B) Security Pacific Bank
333 South Cope Street
Los Angeles, California

(C) Security Pacific Bankers
333 South Hope Street
Los Angeles, California

(D) Security Pacific Bank
333 South Hope Street
Los Añgeles, California

(E) Securities Pacific Bank
333 South Hope Street
Los Angeles, California

Choice-Response Matrix

	Response Choice					
Groups	A	B	C*	D	E	Omits
male	76.24%	8.49%	7.16%	3.44%	3.91%	0.76%
female	83.29%	6.16%	4.86%	2.49%	2.13%	1.07%

MGA SALITANG MAGKAUGNAY ITEM

22. BAYANI : MAGITING ::
PULUBI :

(A) mangmang
(B) mahiyain
(C) gusgusin
(D) baliw
(E) gala

Choice-Response Matrix

	Response Choice					
Groups	A	B	C*	D	E	Omits
NCR	21.44%	3.47%	67.97%	2.12%	4.66%	0.34%
non-NCR	25.46%	5.51%	48.96%	3.18%	16.03%	0.86%