International Education Journal, 2007, 8(2), 237-248.
ISSN 1443-1475 © 2007 *Shannon Research Press*.
http://iej.com.au

Is the Aggression Questionnaire bias free? A Rasch analysis

Sabry M. Abd-El-Fattah[1]
University of South Australia, School of Education *sabry.abd-el-fattah@unisa.edu.au*

Buss and Perry (1992) developed the Aggression Questionnaire (AQ) to assess aggressiveness as a personality trait in high school and college samples. The AQ has been used by researchers in United States, Italy, Germany, Netherland, Japan, Canada, and Greece. The present study is reported on an Arabic adapted version of the AQ among a sample of 510 Egyptian high school students. An exploratory factor analysis technique defined four factors: physical aggression (9 items), verbal aggression (5 items), anger (7 items), and hostility (8 items). The correlation among the four factors ranged from 0.38 to 0.49. A confirmatory factor analysis revealed that the AQ could be described by four first levels factors that were linked by a higher order factor of general aggression. Rasch analysis showed that the AQ was bias free. Relevance of these findings to the assessment of the trait aggressiveness is discussed.

Aggression questionnaire, Egyptian adolescents, bias, Rasch analysis

INTRODUCTION

Aggression describes an overt behaviour intended to harm another person (Bushman, Cooper, & Lemke, 1991). Buss and Perry (1992) published a self-report measure of trait aggressiveness; the Aggression Questionnaire (AQ). The AQ consisted of 29 items. An exploratory factor analysis of responses from 406 college students yielded four correlated factors: (a) physical aggression, (b) verbal aggression (c) anger, and (d) hostility. The hostility factor represented a combination of resentment and suspicion items. The correlation coefficients among the four factors of the AQ ranged from 0.25 to 0.48. Subjects rated their response to each item of the AQ on a 5-point scale ranging from 1 (Extremely uncharacteristic of me) to 5 (Extremely characteristic of me). Thus scores from the four factors of the AQ could be summed to obtain a total score, which represents a respondent's overall level of aggressiveness.

The AQ showed acceptable psychometric properties as indicated by the test-retest reliability over a period of nine weeks being 0.80 for physical aggression, 0.76 for verbal aggression, 0.72 for anger, 0.72 for hostility, and 0.80 for overall AQ. A confirmatory factor analysis (CFA) showed that the AQ could be described by four first-level factors (i.e., physical aggression, verbal aggression, anger, and hostility) that were linked by a higher order factor (i.e., general aggression). Buss and Perry reported that the factorial structure of the AQ was invariant when compared with the factor loadings emerged from an exploratory factor analysis across two samples of college students and across gender.

The invariance of the four-factor structure of the AQ has been validated in a number of studies in several countries. For example, Fossati, Maffei, Acquarini, and Di Ceglie (2003) reported that the four-factor structure of the AQ was invariant in a sample of Italian university students. Bernstein and Gesn (1997) found that the four-factor structure of the AQ was invariant in a sample of

[1] The author would like to thank Ali Abd-Allah and Samya Ibrahim for their valuable help with data collection. Special thanks for Rebecca Woodford, University of Birmingham, UK, and Donald Armstrong, University of Pennsylvania, United States, for their comments on an early version of this manuscript.

American university students and that the factorial structure was not an artefact of differences in items distributions. Similar results were reported by von Collani and Werner (2005) in a sample of German university students, and Tsorbatzoudis (2006) in a sample of Greek high school students.

Other studies, however, have reported adequate fit of the AQ only after some items were discarded and reviews could be found in Williams, Boyd, Cascardi, and Poythress, (1996). For example, Harris (1995) validated the four-factor structure of the AQ in a sample of Canadian university students after removing two items from the hostility scale. Similarly, Meesters, Muris, Bosma, Schouten, and Beuving (1996) suggested discarding three items from the hostility scale when working with a sample of Dutch university students. Furthermore, Nakano (2001) conducted a validation study on a Japanese adapted version of the AQ. Although Nakano found the Japanese version of the AQ to be psychometrically adequate, his results indicated a better fit of the four-factor structure when two items were removed from the physical aggression scale.

Furthermore, Vigil-Colet, Lorenzo-Seva, Codorniu-Raga, and Morales (2005) argued that some items of the AQ may be culturally or linguistically biased. The reanalysis of the data collected in different cultures and languages indicated that some items should be discarded. Vigil-Colet et al. developed a new short version of the AQ by removing Items 4 and 7 of the physical aggression scale; Item 3 of the verbal aggression scale; Items 4, 5, and 7 of the anger scale; and Items 2, 3, and 6 of the hostility scale. The resulting scale showed an adequate fit to the four-factor structure and an internal consistency similar to that of the full version of the AQ.

AIM OF THE STUDY

Considering previous research findings, it is possible to suggest that the factorial structure of the AQ needs to be further investigated in different contexts. The present study is reported from an Arabic adapted version of the AQ among a sample of Egyptian high school students. One goal of the present study is to investigate the factorial structure of the AQ within an Egyptian context. A second goal is to test for gender bias of the AQ across Egyptian males and females groups using the Rasch analysis procedure.

METHODS

Participants

Subjects of the present study included 510 (265 males and 245 females) second year students enrolled in two high schools in El-Minia, Egypt during 2006. The median age of students was 16.3 years with a range from 16 to 18 years. Students were recruited to participate during their normal classes at their schools. Participation was voluntary and 32 students from the approached sample declined to participate in data collection. Because only two schools were involved in the data collection, no allowance has been made for the slight cluster sample design of the study, although the use of two schools can be used to provide replication in the analysis.

Measurements

The Aggression Questionnaire (AQ)

Buss and Perry (1992) developed the AQ as an updated version of an earlier scale, the Hostility Inventory (Buss & Durkee, 1957). The version of the AQ employed was a self-reported measure that consisted of 29 items and four subscales: physical aggression (9 items), verbal aggression (5 items), anger (7 items), and hostility (8 items). Subjects rated their response to each item of the AQ on a 5-point scale that ranged from 1 (Extremely uncharacteristic of me) to 5 (Extremely characteristic of me).

Procedures

The author translated the 29 items of the AQ from English to Arabic. Applying a blind-back-translation strategy, two qualified translators, working without referencing to the English version of the AQ, independently translated the Arabic version back to English. All the translators were accredited with the British-Egyptian Centre in El-Minia, Egypt. Other three qualified translators independently compared the original English version of the AQ to the new English version that was translated back from Arabic, and rated the match between the two versions on a scale from 1 to 10. A score of 1 represented poor match, whereas a score of 10 represented perfect match. The average percentage of match between the two versions of the AQ was 96 per cent which could be considered acceptable (see, Brislin, Lonner, & Thorndike, 1973). The AQ was administered to the sample of the study in the eleventh week of the 2006 school year.

RESULTS

Exploratory Factor Analysis

An exploratory factor analysis with oblique rotation of the AQ identified four correlated factors: physical aggression (9 items, Cronbach $\alpha = 0.82$), verbal aggression (5 items, Cronbach $\alpha = 0.81$), anger (7 items, Cronbach $\alpha = 0.83$), and hostility (8 items, Cronbach $\alpha = 0.80$). The percentage of variance explained by a specific factor was 19 per cent for physical aggression, 14 per cent for verbal aggression, 12 per cent for anger, and 11 per cent for hostility. The correlation coefficients, presented in Table 1, among the four extracted factors range from 0.38 to 0.49. The factor loadings for the four factors of the AQ are recorded in Table 2.

Table 1: Correlation among the four factors of the Aggression Questionnaire (N=510)

Factors	1	2	3	4
1. Physical	–			
2. Verbal	0.49*	–		
3. Anger	0.45*	0.48*	–	
4. Hostility	0.41*	0.38*	0.40*	–

Note. $p < 0.05$

Unidimensionality

In order to employ the Rasch model to test for gender bias of the AQ, it was necessary to examine whether or not the items of the AQ were unidimensional since the unidimensionality of items was agreed to be one of the requirements for the use of the Rasch model (Hambleton & Cook, 1977; Anderson, 1994).

Consequently, a confirmatory factor analysis procedure is employed to test the unidimensionality of AQ items. Confirmatory factor analysis is a statistical procedure that is employed for investigating relations between a set of observed variables and the underlying latent variables (Byrne, 2001; Kim & Mueller, 1978). Thus, confirmatory factor analysis assumes that the observed variables are derived from some underlying source variables (Kim & Mueller, 1978). Factor analysis may also be used as an appropriate method for identifying the minimum number of hypothetical variables that account for the observed covariation, and thus as a means of exploring the data for possible data reduction (Kim & Mueller, 1978). However, one of the main purposes of confirmatory factor analysis is to examine the common underlying dimensions associated with a number of observed variables.

The Mplus 4.0 program (Muthen & Muthen, 2006) was used to run a confirmatory factor analysis of the AQ using the full information maximum likelihood estimation procedure (Bollen, 1989). The analysis showed that a nested model (see Figure 1) in which the AQ items were assigned to four specific correlated first-order factors of Physical Aggression, Anger, Verbal Aggression, and Hostility, as well as a general higher order factor, which was labelled as *Aggression.* This provided the best fitting model, χ^2 (371, $N = 510$) = 385.6, $p = 0.29$, Root-Mean-Square Error of

Approximation (RMSEA) = 0.01, Standardized Root-Mean-Square Residual (SRMR) = 0.02, Adjusted Goodness of Fit Index (AGFI) = 0.99, Parsimonious Goodness of Fit Index (PGFI) = 0.29, Tucker-Lewis Index (TLI) = 0.99, Parsimony Ratio (PRATIO) = 0.85, and Parsimony Normed Fit Index (PNFI) = 0.83.

Table 2: Exploratory factor analysis of the Aggression Questionnaire (N=510)

Factor/Statement	Factor loadings
Physical Aggression	
1. Once in a while I can't control the urge to strike another person.	0.72
2. Given enough provocation, I may hit another person.	0.68
3. If somebody hits me, I hit back.	0.66
4. I get into fights a little more than the average person.	0.63
5. If I have to resort to violence to protect my rights, I will.	0.59
6. There are people who pushed me so far that we came to blows.	0.55
7. I can think of no good reason for ever hitting a person.*	0.51
8. I have threatened people I know.	0.48
9. I have become so mad that I have broken things.	0.46
Eigenvalue	5.3
Verbal Aggression	
1. I tell my friends openly when I disagree with them.	0.61
2. I often find myself disagreeing with people.	0.58
3. When people annoy me, I may tell them what I think of them.	0.53
4. I can't help getting into arguments when people disagree with me.	0.48
5. My friends say that I'm somewhat argumentative.	0.44
Eigenvalue	2.6
Anger	
1. I flare up quickly but get over it quickly.	0.64
2. When frustrated, I let my irritation show.	0.61
3. I sometimes feel like a powder keg ready to explode.	0.59
4. I am an even-tempered person.*	0.58
5. Some of my friends think I'm a hothead.	0.55
6. Sometimes I fly off the handle for no good reason.	0.53
7. I have trouble controlling my temper.	0.49
Eigenvalue	4.0
Hostility	
1. I am sometimes eaten up with jealousy.	0.62
2. At times I feel I have gotten a raw deal out of life.	0.60
3. Other people always seem to get the breaks.	0.57
4. I wonder why sometimes I feel so bitter about things.	0.54
5. I know that "friends" talk about me behind my back.	0.50
6. I am suspicious of overly friendly strangers.	0.48
7. I sometimes feel that people are laughing at me behind my back.	0.47
8. When people are especially nice, I wonder what they want.	0.45
Eigenvalue	4.2

Note * The scoring of these items was reversed.

All the hypothesized regression path coefficients of the AQ model, presented in Table 3, were statistically significant because the critical ratio (CR) for a specific regression path coefficient was > ±1.96 (Byrne, 2001). The correlation between the error terms associated with two observed variables of the physical aggression scale (i.e., Items 1 and 2, $r = 0.29$) could be justifiable on the basis that correlated error terms often indicated some type of meaning redundancy between the measured variables (see, Abd-El-Fattah 2006; Abd-El-Fattah & Barnes, 2007; Abd-El-Fattah & Yates, 2007; Byrne, 2001).

Rasch Analysis

It is common within classical test theory to sum individual item response values to obtain a total score. However, this approach has been criticised and reviews have been made by Andrich (1978), Masters (1988), and Wright and Masters (1982). For example, Bond and Fox (2001) highlighted that the summing of individual item response values had two underlying assumptions.

First, each item was measured on an equal interval scale. Thus, each item was contributing equally to the underlying trait. Second, the distances or the steps among the response categories were equal for an item and through all items of a scale, that is, the level of the underlying trait required to move from one response category to another was the same for an item and was equal across all items of a scale. Bond and Fox concluded that those two assumptions were counterintuitive and mathematically inappropriate.

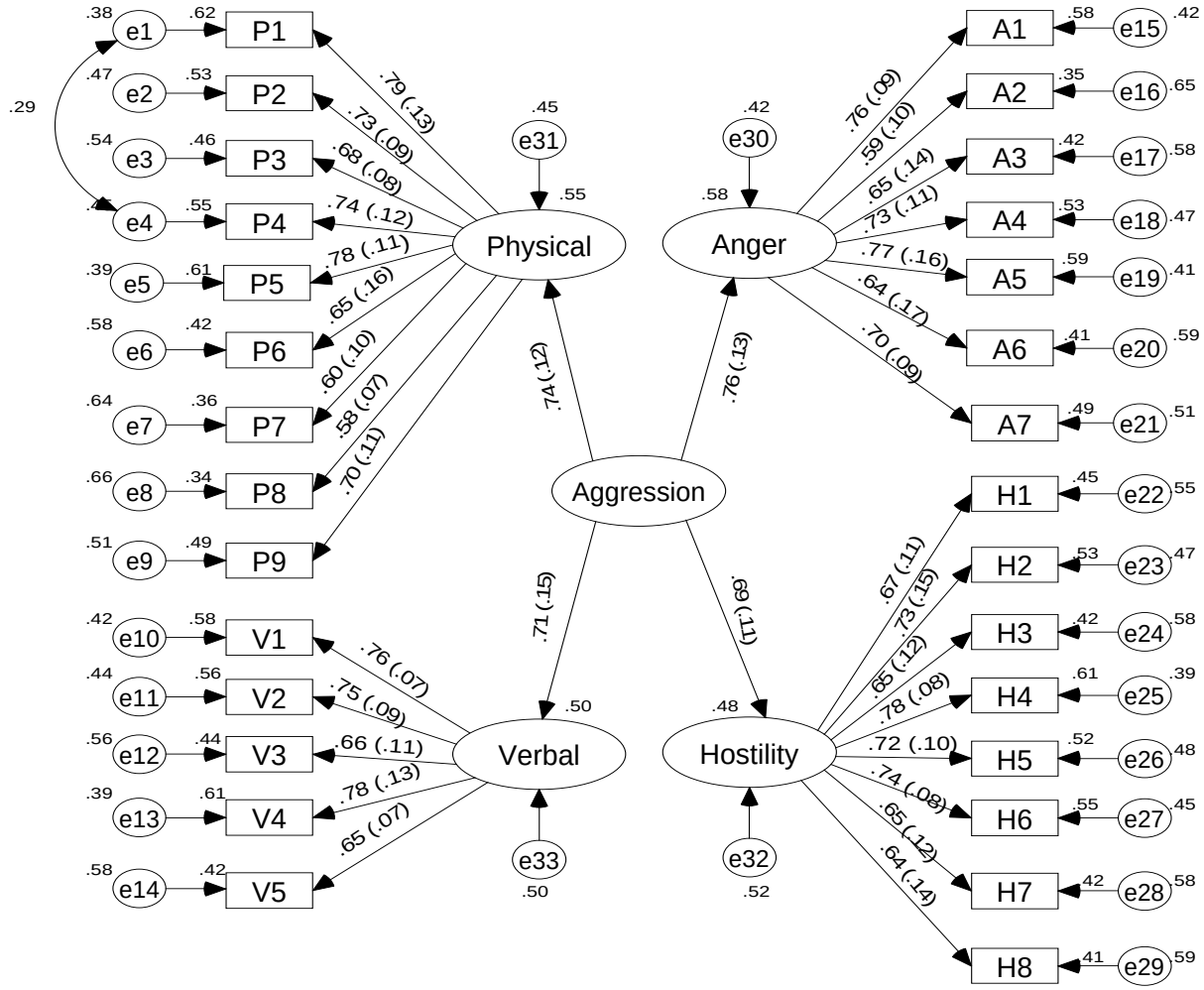

Figure 1: A second-order confirmatory factor analysis of Aggression Questionnaire

The basic Rasch model is a dichotomous response model (Rasch, 1960; Wright & Stone, 1979) that represents the conditional probability of a binary outcome as a function of a person's trait level (*B*) and an item's difficulty (*D*). The Rasch dichotomous response model is given by:

$$p_{ni} = \frac{\exp\left(\beta_n - \delta_i\right)}{1 + \exp\left(\beta_n - \delta_i\right)}$$

where P_{ni} is the probability of an endorsed response (a 'yes' response to an item), βn is the trait (or ability) parameter of person n, and δi is the difficulty of endorsing item i. When $\beta n > \delta i$, $\beta n = \delta i$, and $\beta n < \delta i$, the chances of a 'yes' response is greater than 50 per cent, equal to 50 per cent, and less than 50 per cent, respectively.

Andrich (1978, 1988) is credited for extending Rasch dichotomous response model to the rating scale. The rating scale model is an additive linear model that describes the probability that a specific person (*n*) will respond to a specific Likert-type item (*i*) with a specific rating scale step

(x). It is important to note that the Likert scale can be modelled with either the rating scale or the partial credit model (Masters, 1988; Wright & Masters, 1982). The partial credit model allows the item format and the number of categories to vary from item to item (e.g., some items are scored with a 5-point scale and others with a 6-point scale). When the item format is inconsistent from item to item, the partial credit model is useful in providing estimates of the psychological distance between each set of the ordinal categories (Masters, 1988). However, the rating scale model restricts the step structure to be the same for all items (Wright & Masters, 1982). In essence, the rating scale models are a subset of the partial credit models (Andrich, 1978).

Table 3: Standardized path coefficients, standard error, critical ratio, error variance, and R^2 of the second-order confirmatory factor analysis of the Aggression Questionnaire (N = 510)

Paths	Path coefficient	Standard error	Critical ratio	Error variance	R^2
Physical Aggression					
1	0.79	0.13	6.1	0.38	0.62
2	0.73	0.09	8.1	0.47	0.53
3	0.68	0.08	8.5	0.54	0.46
4	0.74	0.12	6.2	0.45	0.55
5	0.78	0.11	7.1	0.39	0.61
6	0.65	0.16	4.1	0.58	0.42
7	0.60	0.10	6.0	0.64	0.36
8	0.58	0.07	8.3	0.66	0.34
9	0.70	0.11	6.4	0.51	0.49
Verbal Aggression					
1	0.76	0.07	10.9	0.42	0.58
2	0.75	0.09	8.3	0.44	0.56
3	0.66	0.11	6.0	0.56	0.44
4	0.78	0.13	6.0	0.39	0.61
5	0.65	0.07	9.3	0.58	0.42
Anger					
1	0.76	0.09	8.4	0.42	0.58
2	0.59	0.10	5.9	0.65	0.35
3	0.65	0.14	4.6	0.58	0.42
4	0.73	0.11	6.6	0.47	0.53
5	0.77	0.16	4.8	0.41	0.59
6	0.64	0.17	3.8	0.59	0.41
7	0.70	0.09	7.8	0.51	0.49
Hostility					
1	0.67	0.11	6.1	0.55	0.45
2	0.73	0.15	4.9	0.47	0.53
3	0.65	0.12	5.4	0.58	0.42
4	0.78	0.08	9.8	0.39	0.61
5	0.72	0.10	7.2	0.48	0.52
6	0.74	0.08	9.3	0.45	0.55
7	0.65	0.12	5.4	0.58	0.42
8	0.64	0.14	4.6	0.59	0.41
Aggression					
Physical	0.74	0.12	6.2	0.45	0.55
Verbal	0.71	0.15	4.7	0.50	0.50
Anger	0.76	0.13	5.8	0.42	0.58
Hostility	0.69	0.11	6.3	0.52	0.48

The simple dichotomous response model can be extended to provide an appropriate model for use with polytomous response categories by the addition of an additional difficulty parameter; either a second δ parameter or a τ parameter. The Rasch rating scale model is given by:

$$p_{nij} = fn\left(\frac{\exp(\beta_n - \delta_i - \tau_j)}{1 + \exp(\beta_n - \delta_i - \tau_j)}\right)$$

Or

$$p_{nij} = \frac{\sum_{k=1}^{mi} \exp \sum_{j=1}^{k} (\beta_n - \delta_{ij})}{1 + \sum_{k=1}^{mi} \exp \sum_{j=1}^{k} (\beta_n - \delta_{ij})}$$

where *n* = subscript for persons, *i* = subscript for items, and *j* = response categories (0, 1, 2).

In the present analysis, the QUEST program (Adam & Khoo, 1993) was used to run the Rasch analysis for the AQ. All the reported results were obtained from the QUEST program. The RUMM program (Andrich, Sheridan, & Luo, 2000), however, was used to plot the Item Characteristic Curve and Category Probability Curve with thresholds for an example item of the AQ.

Item fit statistics

One important item fit statistics was the infit mean square (INFIT MNSQ). The infit mean square measured the consistency of fit of the cases to the Item Characteristic Curve (ICC) for each item with weighted consideration given to those cases close to the 0.5 probability level. The acceptable range of the infit mean square statistic for each item of the AQ was taken to be from 0.77 to 1.30 (Adams & Khoo, 1993). Items that had infit mean square above 1.30 indicated that the relevant items did not discriminate well, and below 0.77 indicated that the relevant items provide redundant information. Items that had INFIT MNSQ outside the acceptable range must be deleted from the analysis (Wright & Stone, 1979). Figure 2 shows that, in the present analysis, no items of the AQ had been deleted because all items had an INFIT MNSQ value within the acceptable range of 0.77 to 1.30. Specifically, the range of the INFIT MNSQ for all items ranged from 0.83 to 1.18.

The RUMM program could divide the examined sample into a specified number of groups or Class Intervals (CIs) for each item. The average ability of individuals within each CI was calculated and represented by a dot on the ICC for each item. If an item fit the Rasch model, the dots should fall on or as close as possible to the ICC. Any deviations of any of these dots from the ICC represented a difference between the observed mean ability of the CI that these dots represent and the expected mean ability of the CI as predicted by the Rasch model. In the present analysis, the RUMM program divided the sample of the study (N = 510) into six CIs that were plotted along the ICC for each item. Figure 3 shows the ICC for Item 3 of the AQ.

Figure 4 shows the Category Probability Curve and thresholds for Item 3 of the AQ. The thresholds reflect the item difficulty for each item. According to Bond and Fox (2001), a threshold is "the level at which the likelihood of failure to endorse a given response category (below the threshold) turns to the likelihood of endorsing the category (above the threshold)" (p. 234). For example, in the case of four response categories, there are three thresholds that mark the boundaries between the four response categories: SD (*Strongly Disagree*)-D (*Disagree*)-A (*Agree*)-SA (*Strongly Agree*) and all are ordered. That is, the data are regarded as ordinal and the Rasch model transform the counts of the endorsement of these ordered Likert categories into interval scales (Bond & Fox, 2001).

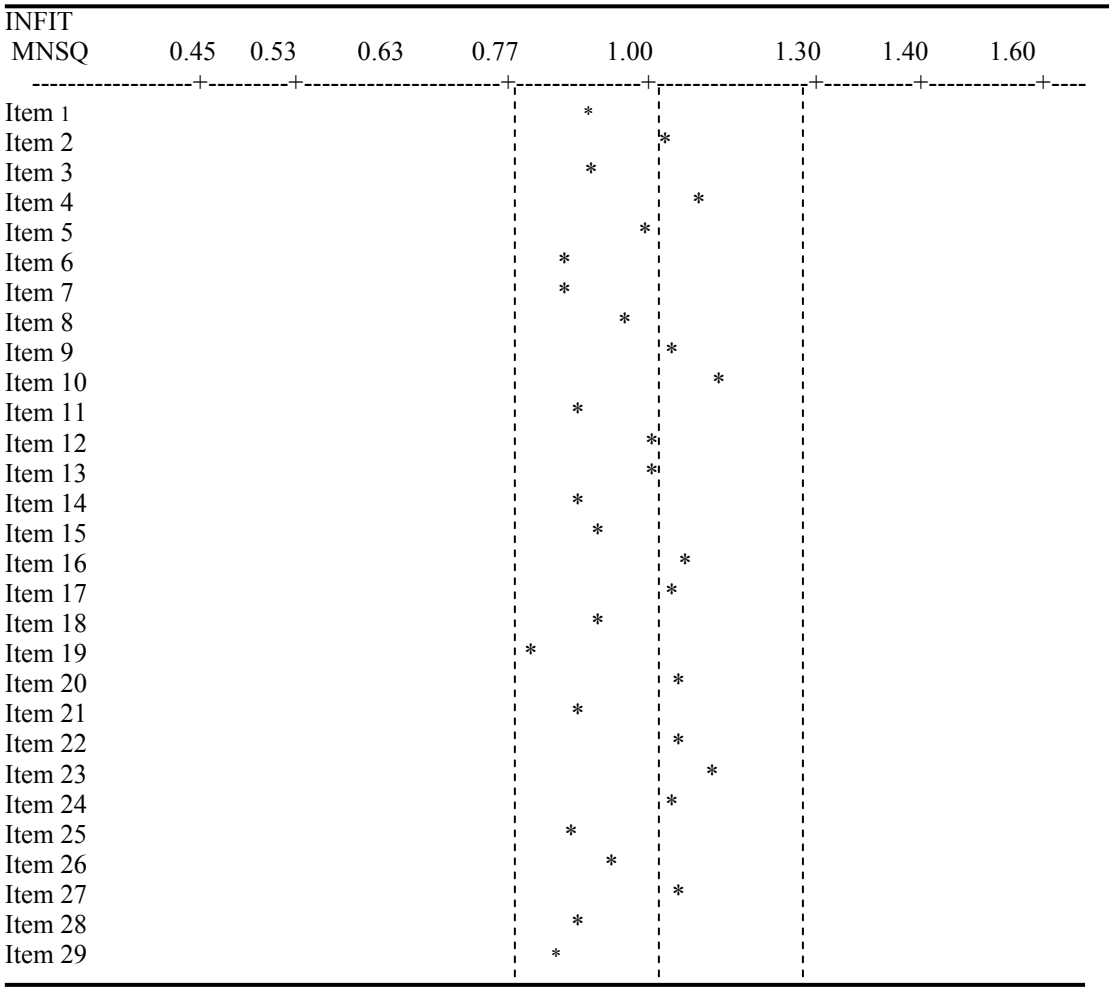

Figure 2: Plot of all Infit Mean Squares for all items of the AQ

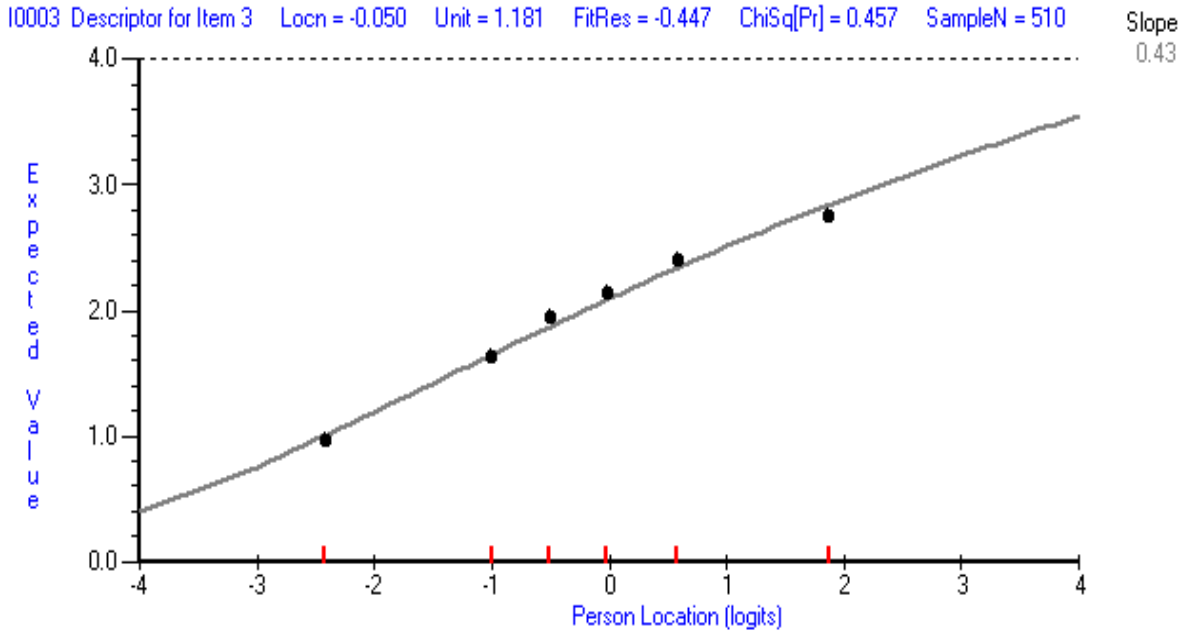

Figure 3: Item Characterise Curve for Item 3 of the AQ

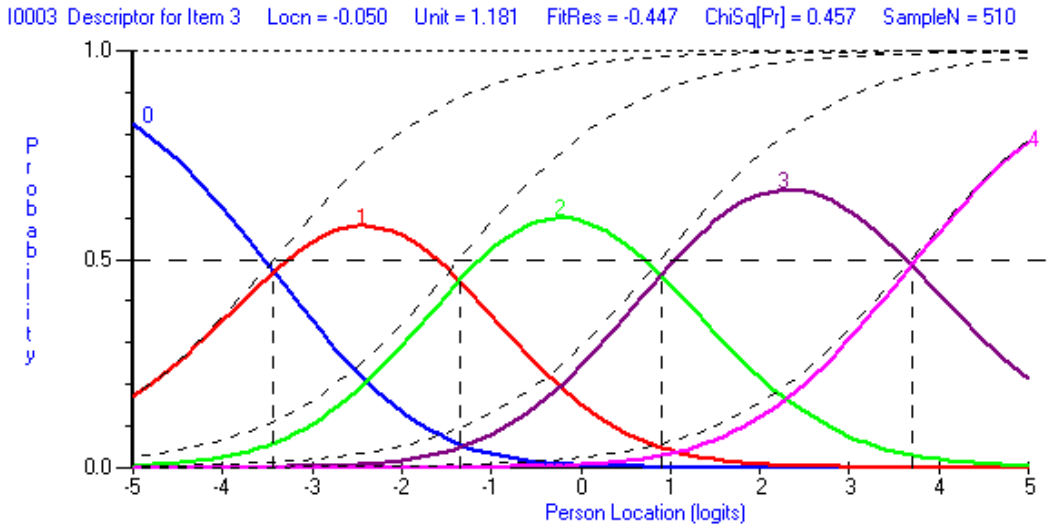

Figure 4: Category Probability Curve and thresholds for item 3 of the AQ

Case Estimates

It is also important when investigating the fit of the Rasch scale to data to examine the estimates for each case. The case estimates give the performance level of each student on the total scale. In order to identify whether the cases fit the Rasch scale or not, it is important to examine the case OUTFIT mean square statistic (OUTFIT MNSQ) which measures the consistency of the fit of the persons to the student characteristic curve for each student, with special consideration given to extreme items. In the present analysis, the general guideline used for interpreting *t* as a sign of misfit is if $t > \pm 5$ (Wright & Stone, 1979). Thus, if the OUTFIT MNSQ value for a person had a *t*-value greater than $\pm$ 5, that person did not fit the scale and was consequently deleted from the analysis. In the present analysis, no person was deleted because the *t*-value for all cases fell within the acceptable range of $\pm$ 5. Specifically, in the present analysis, the OUTFIT MNSQ for all cases had *t*-values between - 2.8 to + 3.7, and since the normal *t*-value tests were not being employed, as is stated above, no cases were deleted.

Gender Bias

Differential item functioning (DIF) might result in an unfair advantage to members of one group over the members of another group (Lord, 1980). Therefore, it was necessary to ensure that every item was functioning identically across all groups of interest. Item response theory (IRT) was a preferred method for detecting DIF (Lord, 1980). Detecting DIF was based on comparing the ICCs of a specific item, which were estimated separately in each group. If a given item was unbiased, then the ICCs for that item should be the same. When the estimated ICCs of the relevant item differed between the groups of interest by more than sampling error, then DIF was suspect (Lord, 1980).

The QUEST program produced a plot of standardized differences between the performances of the groups of interest for each item. An item that had a *t*-value > $\pm$ 2 indicated significant differences in performance between the groups of interest and the relevant item needed to be further investigated in order to identify the cause of the bias (Wright & Stone, 1979). Figure 5 shows that, in the present analysis, no gender bias was detected for any item of the AQ because all items had standardized differences between males and females groups within the acceptable range of $\pm$ 2. Specifically, in the present analysis, the standardized difference between males and females groups ranged from - 1.84 to +1.78

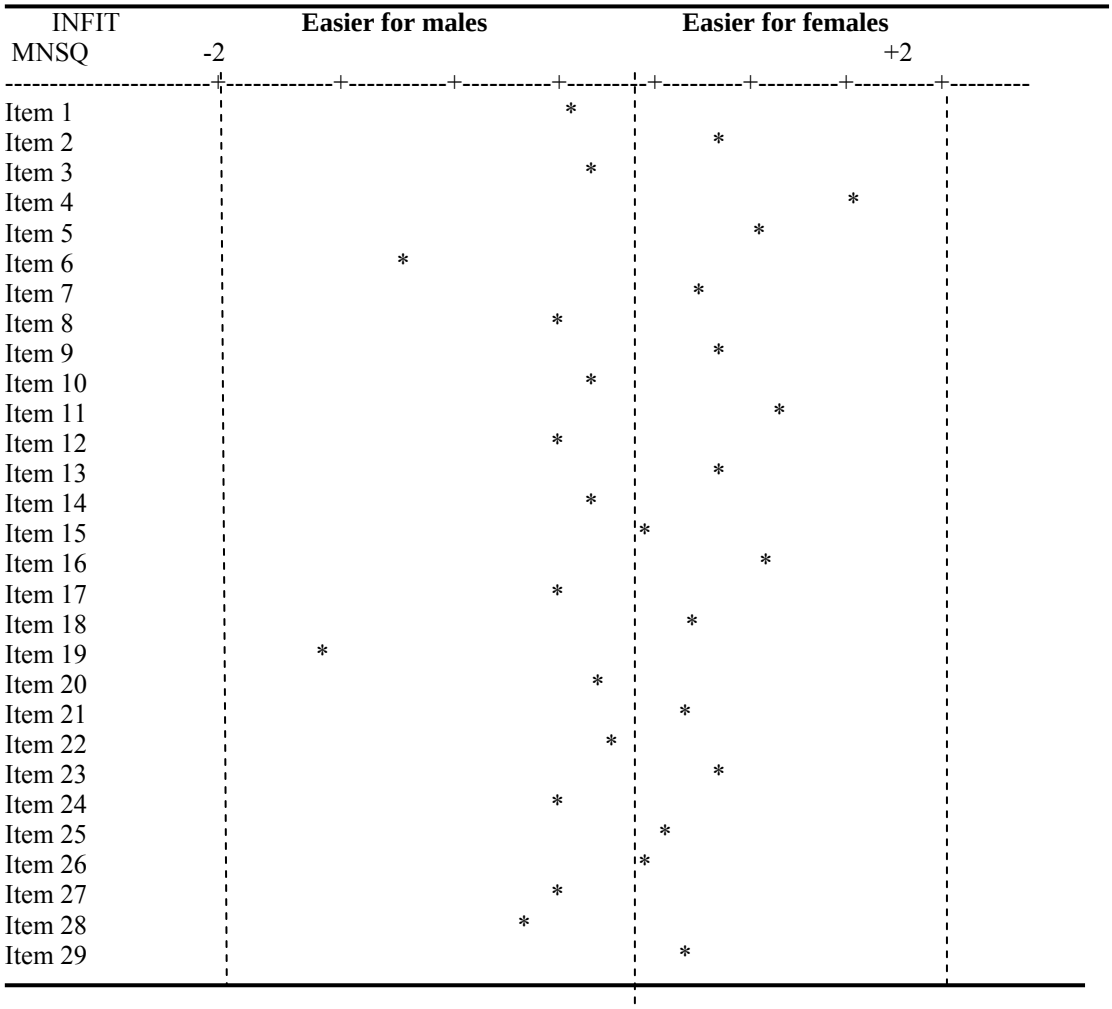

Figure 5: Plot of the standardized differences for males and females groups for all items of the AQ

DISCUSSION

One goal of the present study was to investigate the factorial structure of the AQ within an Egyptian context. A second goal was to test whether the AQ was free of gender bias using the Rasch analysis procedure. The findings of the study showed that the AQ could be described by four first-level factors (i.e., physical aggression, verbal aggression, anger, and hostility) that were linked by a higher order factor (i.e., general aggression). These results seem to be consistent with the original four-factor structure of the AQ as described by Buss and Perry (1992). In addition, these results were in line with findings from other research studies that had replicated the four-factor structure of the AQ and found it to be invariant in different cultures and contexts such as United States (Bernstein & Gesn, 1997), Greece (Tsorbatzoudis, 2006), Italy (Fossati, Maffei, Acquarini, & Di Ceglie, 2003), and Germany (von Collani & Werner, 2005). In a manner different from other studies that suggested removing some items of the AQ to achieve a better goodness-of-fitness of the four-factor structure (see, Harris, 1995; Meesters, Muris, Bosma, Schouten, & Beuving, 1996; Nakano, 2001), the findings of the present study did not suggest that any item of the AQ should be discarded.

A second finding of the present study showed that the AQ was free of gender bias. This implied that all items of the AQ seemed to function in a highly similar way across males and females groups. This finding seemed to be inconsistent with other research findings that had employed factorial invariance procedures and recommended developing a shorter version of the AQ because

some items seemed to be culturally or linguistically biased (see, Vigil-Colet, Lorenzo-Seva, Codorniu-Raga, & Morales, 2005).

In summary, the AQ seemed to represent a promising measure of the trait aggressiveness. The AQ showed satisfactory psychometric properties and could be described by four first-level factors that were linked by a higher order factor of general aggression. In addition, the AQ seemed to be free of gender bias.

REFERENCES

Abd-El-Fattah, S. M. (2006). Effects of family background and parental involvement on Egyptian adolescents' academic achievement and school disengagement: A structural equation modelling analysis. *Social Psychology of Education, 9*, 139-157.

Abd-El-Fattah, S. M., & Barnes, A. (2006). Con*firmatory factor analysis of Bath County Computer Attitude Scale within an Egyptian context: Testing competing models*. Paper presented at the Australian Association for Research in Education Conference, Adelaide, South Australia.

Abd-El-Fattah, S. M., & Yates, G. C. R. (2006). *Implicit Theory of Intelligence Scale: Testing for factorial invariance and mean structure.* Paper presented at the Australian Association for Research in Education Conference, Adelaide, South Australia.

Adams, R. J. & Khoo, S.T. (1993). Quest- The Interactive Test Analysis System. Hawthorn, Victoria: ACER.

Anderson, L. W. (1994). Attitude measures. In T. Husen (ed), *The International Encyclopaedia of Education, Vol. 1*, (second ed.), pp. 380-390. Oxford: Pergamon.

Andrich, D. (1978). Rating formulation for ordered response categories. *Psychometrika, 43*, 561-573.

Andrich, D. (1988). *Rasch Models for Measurement*. Newbury Park CA: Sage.

Andrich, D., Sheridan, B., & Luo, G. (2000). RUMM2010: A Windows interactive program for analyzing data with Rasch unidimensional Models for Measurement [Computer Program]. Perth, Western Australia: RUMM Laboratory

Bernstein, I. H., & Gesn, P. R. (1997). On the dimensionality of the Buss/Perry Aggression Questionnaire. *Behaviour Research and Therapy, 35*, 563-568.

Bollen, K. A. (1989). Structural equations with latent variables. New York: John Wiley.

Bond, T. G. & Fox, C. M. (2001). Applying the Rasch model: Fundamental *measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum Associates

Brislin, R. W., Lonner, W. J., & Thorndike, E. M. (1973). *Cross-cultural research methods*. New York: John Wiley.

Byrne, B. (2001). *Structural equation modelling with AMOS: Basic concepts*, *applications, and programming*. New Jersey: Lawrence, Erlbaum Associates.

Bushman, B. J., Cooper, H. M., & Lemke, K. M. (1991). Meta-analysis of factor analyses: An illustration using the Buss-Durkee Hostility Inventory. *Personality and Social Psychology Bulletin, 17*, 344-349.

Buss, A. H., & Durkee, A. (1957). An inventory for assessing different kinds of hostility. *Journal of Consulting Psychology*, 31, 343-349.

Buss, A. H., & Perry, M. (1992). The Aggression Questionnaire. *Journal of Personality and Social Psychology*, 63, 452-459.

Fossati, A., Maffei, C., Acquarini, E., & Di Ceglie, A. (2003). Multi-group confirmatory component and factor analyses of the Italian version of the Aggression Questionnaire. *European Journal of Psychological Assessment, 19*, 54-65.

Harris, J. A. (1995). Confirmatory factor analysis of the Aggression Questionnaire. *Behaviour Research and Therapy*, 33, 991-993

Hambleton, R. K. & Cook, L. L. (1977). Latent trait models and their use in the analysis of educational test data. *Journal of Educational Measurement*, *14*, 75-96.

Kim, J & Mueller, C. W. (1978). *Factor analysis statistical methods and practical issues*. London: Sage.

Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Hillside, NJ: Lawrence Erlbaum Associates.

Masters, G. N. (1988). The analysis of partial credit scoring. *Applied Measurement in Education*, *1*, 279-297.

Meesters, C., Muris, P., Bosma, H., Schouten, E., & Beuving, S. (1996). Psychometric evaluation of the Dutch version of the Aggression Questionnaire. *Behaviour Research and Therapy, 34*, 839-843.

Muthen, L. & Muthen, B. (2006). *Mplus 4.0* [Statistical Program]. Los Angeles, CA: Muthen & Muthen.

Nakano, K. (2001). Psychometric evaluation on the Japanese adaptation of the Aggression Questionnaire. *Behaviour Research and Therapy, 39*, 853-858.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Chicago: University of Chicago Press.

Tsorbatzoudis, H. (2006). Psychometric evaluation of the Greek version of the aggression questionnaire. *Percept Motor Skills, 102*, 703-718.

Vigil-Colet, A., Lorenzo-Seva, U., Codorniu-Raga, M. J., & Morales, F. (2005). Factor structure of the Aggression Questionnaire among different samples and languages. *Aggression Behavior, 31*, 601-608.

von Collani, G., & Werner, R. (2005). Self-related and motivational constructs as determinants of aggression. An analysis and validation of a German version of the Buss-Perry Aggression Questionnaire. *Personality and Individual Differences, 38*, 1631-1643.

Williams, T. W., Boyd, J. C., Cascardi, M. A., & Poythress, N. (1996). Factor structure and convergent validity of the Aggression Questionnaire in an offender population. *Psychological Assessment, 8*, 398-403.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.

Wright, B. D & Stone, M. H. (1979). *Best test design: Rasch measurement*. Chicago: Mesa Press.