Northumbria Research Link

An analysis of differential item functioning by gender in the Learning Disability Screening Questionnaire (LDSQ)

FINAL ACCEPTED MANUSCRIPT BEFORE COPY EDITING

DO NOT CITE

Abstract

The Learning Disability Screening Questionnaire (LDSQ; McKenzie & Paxton, 2006) was developed as a brief screen for intellectual disability. Although several previous studies have evaluated the LDSQ with respect to its utility as a clinical and research tool, no studies have considered the fairness of the test across males and females. In the current study we, therefore, used a multi-group item response theory approach to assess differential item functioning across gender in a sample of 211 males and 132 females assessed in clinical and forensic settings. Although the test did not show evidence of differential item functioning by gender, it was necessary to exclude one item due to estimation problems and to combine two very highly related items (concerning reading and writing ability) into a single literacy item Thus, in addition to being generally supportive of the utility of the LDSQ, our results also highlight possible areas of weakness in the tool and suggest possible amendments that could be made to test content to improve the test in future revisions.

Keywords: Learning Disability Screening Questionnaire; LDSQ; differential item functioning; measurement invariance; intellectual disability.

Introduction

Intellectual disability (ID) is defined with respect to three criteria: significant impairment in intellectual functioning, significant impairment in adaptive functioning, and onset before adulthood (British Psychological Society, 2000). From a societal perspective, ID is an important disorder affecting approximately 10 in 1000 individuals and involving the utilisation of large amounts of public health resources (Maulik, Mascarenhas, Mathers, Dua, & Saxena, 2011). Accurately identifying individuals with ID is an important psychometric challenge. Correct identification of individuals with an ID facilitates the provision of appropriate services and resources to individuals in need of this additional support. However, false identifications waste resources and potentially expose the wrongly diagnosed individual to social stigma. Given the potential consequences of mis-classifying individuals as affected versus unaffected by ID, there are stringent guidelines regarding the assessment of the disorder. An assessment should be conducted by a trained professional with intellectual deficits measured using a standardised assessment such as the Wechsler Adult Intelligence Scale Fourth Edition (WAIS-IV: Wechsler, 2008) in adults. It is also recommended that adaptive functioning is also assessed using a standardised assessment such as the Adaptive Behaviour Assessment System Second Edition (ABAS-II: Harrison & Oakland, 2003) or Vineland Adaptive Behaviour Scales Second Edition (Vineland II: Sparrow, Balla, & Cicchetti, 2005). The necessary rigor with which individuals are assessed makes the process of assessment time and resource intensive as well as demanding and potentially stressful for the individual being assessed. For resource stretched psychological services and individuals with a potential ID alike, it is, therefore, desirable to target full diagnostic assessments only to those individuals with a genuine risk of having an ID.

The Learning Disability Screening Questionnaire (LDSQ: McKenzie & Paxton, 2006) was developed with this goal. Note that while the LDSQ measures 'intellectual disability' it

was developed at a time when this construct was referred to by the label of 'learning disability' in the UK. This was prior to the adoption of the term 'intellectual disability' [see Schalock, Luckasson, & Shogren (2007) for a discussion regarding changing terminologies]. The LDSQ is a brief screening tool for identifying ID that can be used by a range of professionals without the need for extensive training. It provides a quick estimate of whether an individual is likely to have an ID or not and, therefore, aids in identifying those individuals who should undergo full assessment for ID. The content of the scale refers to key areas of intellectual and adaptive functioning that are associated with the presence of ID. Across seven dichotomously scored items the areas of practical living, occupational, and academic skills are assessed. Abbreviated item contents are provided in Table 1.

The utility of the LDSQ as a clinical and research tool has been evaluated in several previous studies. In terms of its utility as a screening tool, an initial validation study in a clinical sample reported sensitivity and specificity values of 91% and 87% respectively at the cut-off point specified for suspected ID in a community sample (Paxton, McKenzie, & Murray, 2008). Similar discriminative ability was later reported in a forensic sample (sensitivity = 82% and specificity = 88%; McKenzie, Michie, Murray, & Hales, 2012). A subsequent study evaluated the ability of the scale to accurately classify individuals in terms of the British Psychological Society severity categories of ' no ID', 'significant impairment' (IQ between 55 and 69) and 'severe impairment' (IQ less than 55) [Murray & McKenzie, 2014]. This study found that, while the performance of the scale in making this 3 category distinction was statistically significant, it was not clinically significant in the sense of providing sufficient certainty of classification to justify a formal clinical categorisation. However, Murray and McKenzie (2013) did identify some additional benefits to the scale in terms of clinical and research applications. They found that the 6 out of 7 of the items exhibiting invariant item ordering: a property which means that information about levels of

severity is contained not only in overall scores, but in individual item responses. For example, if an individual is unable to tell the time (an easy item), it is possible to infer that their impairment is relatively severe and that it is unlikely that they can read or write (harder items), even if this information is not available.

In terms of applications of the LDSQ, it has been used in clinical and criminal justice services across the UK, in Europe, Japan, and Australia. It was independently piloted by the Department of Health for use in a range of criminal justice services and is currently being used by a number of probation, prison and police services, as well as ID services in the UK.

To date, however, no studies have examined the important issue of whether the LDSQ has equivalent measurement properties in males and females i.e. that it does not show test bias with respect to gender.

There are broadly two forms of test bias that can be defined: prediction invariance and measurement variance. Prediction invariance refers to when the relation between test scores and some criterion is equal across groups, whereas measurement invariance refers to when the relation between the test score and the latent attribute measured by the test is equal across groups. Borsboom, Romeijn and Wicherts (2008) demonstrated that under most circumstances when one holds this implies violation of the other. Thus, a choice must be made between conceptualising test bias in terms of measurement or predictive invariance. The authors provide several arguments for preferring the measurement invariance approach, an important one being that only a measurement invariant test implies that individuals with the same latent trait level should have the same expected test score.

More specifically, measurement invariance can be defined as the conditional distribution of observed scores given the underlying latent variable of interest being independent of group membership (Kim & Yoon, 2012):

$$f(X|W,G) = f(X|W)$$

where X represents the observed score, W represents the underlying latent variable, and G represents group membership. The consequences of a lack of invariance in the context of research studies and selection has been widely discussed (e.g. Meredith & Teresi, 2006).

In the context of measuring a trait such as ID, the absence of measurement invariance can result in the test being biased against one of the groups, for example, with males having higher expected scores than females for the same standing on the latent trait. Thus, the test could systematically over-estimate the level of impairment of one gender relative to the other. It is important to check, rather than assume that a test is measurement invariant. For example, it is known that there exist general population sex differences in cognitive profiles which can be broadly characterised as a relative strength for males in visuospatial processing (especially rotation) and females for verbal processing (e.g. Johnson & Bouchard, 2007; Silverman et al., 2000). If the items of the LDSQ refer to skills disproportionately relying on one of these abilities (e.g. it has 2 literacy items presumably relying somewhat on verbal ability) then test bias could result. In addition, violations of invariance can also provide theoretical insights into group differences (Millsap, 2012). It was, therefore, the aim of the current study to test measurement invariance by gender in the LDSQ.

Method

Participants

Data were gathered from forensic and clinical settings. All participants had English as their first language and the majority were White and from Scotland. The forensic services comprised a community ID forensic service and a forensic in-patient secure unit, both part of a specialist forensic service in Scotland, and a prison based in England. The data from the clinical settings came from two community ID services from two health board areas in Scotland. The LDSQ was completed from existing case-notes and information provided by clinical psychology staff for all those who had been assessed to determine whether they had

ID or not. Although the clinical sample was significantly older than the forensic sample [t(174.03)=-3.08, p<.05], the actual mean ages were similar at 36.6 and 31,5 respectively. Similarly, the two samples differed significantly on FSIQ [t(148.92)=-2.77, p<.05] but both had mean IQs in the 'mild impairment' range. The mean FSIQ for the clinical sample was 63.5 and the mean FSIQ for the forensic sample was 68.2.

Data on a total of 211 male participants were available sampled from forensic (n = 86), clinical (n = 125) settings. The mean age at assessment (based on n = 208 with available age data) of 34.02 (SD = 14.09). The mean estimated full scale IQ (based on n = 201 with available data on IQ) was 64.49 (SD = 11.06). The distribution of IQ scores in this group is shown in Figure 1.

Data for a total of 132 female participants were available, sampled from forensic (n = 24) and clinical (n = 108) settings. The mean age at assessment (based on n = 113 with available age data) of 30.81 (SD = 14.20). The mean estimated full scale IQ (based on n = 131 with available IQ data) was 65.58 (SD = 14.46). The distribution of IQ scores in this group are shown in Figure 2.

<center>(Insert Figures 1 & 2 about here)</center>

Measures

The LDSQ is a 7 item dichotomously scored scale designed to identify those who may have ID. Previous research and preliminary analyses in the current study suggested that two of the items 'read' and 'write' exhibit a high degree of collinearity. The estimate from the current sample with complete responses to both items (n = 286) is *rho* = 0.95. This can cause estimation problems, therefore, we combined these items into a single 'literacy' item which was scored as a '0' if an individual did not possess both reading and writing skills and '1' if they possessed both reading and writing skills. For 15 individuals who had data on only one

of these two items, coding on 'literacy' was determined by the observed response given the evidence that reading and writing skills tend to cluster together strongly.

The scale is appropriate for individuals who are 16 and over and is designed to be used by a range of professionals without the need for a particular qualification or training. Items refer to whether an individual has certain abilities or needs from the domains of literacy, independent living, contact with specialist services, schooling, and employment. Comprehensive details of the development and validation of the scale can be found in McKenzie and Paxton (2006).

Statistical Procedure

Item response theory analysis

To assess for item bias we used a multi-group item response theory (IRT) model. A single uni-dimensional model was specified, representing the theoretical assumption that the LDSQ measures a uni-dimensional construct of ID. This assumption has been supported in previous psychometric analyses of the LDSQ (Murray & McKenzie, 2013). Further, with only 7 items (6 with 'Read' and 'Write' combined into 'Literacy'); any model with more than one latent variable would have poor factor determination. The specific IRT model employed was the 2 parameter logistic (2PL) model which defines the probability of endorsing an item in terms of a logistic function of the difference between the ability of an individual and the difficulty of an item:

$$P_j(\theta_i) = \frac{\exp[a_j(\theta_i - b_j)]}{1 + \exp[a_j(\theta_i - b_j)]}$$

where θ_i is the latent trait level for individual i, and a_j and b_j are the discrimination and difficulty parameters for item j respectively.

Single group models

All analyses were run in IRTPRO 2.1.21. Models were estimated using the Bock-Aitken EM algorithm. We began by fitting the model to each of the gender groups individually. Item fit was assessed using the $S\text{-}\chi^2$ proposed by Orlando and Thissen (2000). It is based on a comparison of observed and expected item responses given an individual's level of the latent trait. Expected frequencies are computed using the joint likelihood distribution of all possible sum scores across all possible response patterns for each sum score. The degrees of freedom for the statistic is the number of sum score categories minus the number of item parameters. If necessary, cells can be collapsed so that a minimum expected frequency of 1 can be achieved.

Local dependence was assessed using the $LD\chi^2$ test proposed by Chen and Thissen (1997). This statistic compares the observed to expected frequencies of the two-way contingency tables between response to each item and each of the other items. These are transformed into z-scores by subtracting the degrees of freedom and dividing by the square root of two times the degrees of freedom. Both this statistic and the z-scores are approximately χ^2 distributed. Guidance in the IRTPRO manual suggests values greater than 10 are deemed to be highly problematic, values between 5-10 potentially problematic, and values less than 5 suggest local independence.

Differential item functioning was then assessed using a Wald test following the suggestions of Woods, Cai, and Wang (2013). Woods et al. (2013) tested the performance of two different Wald tests for the identification of DIF implemented in IRTPRO. The first test, Wald-2 (Langer, 2008) is a two-step procedure. In the first step, the reference group (here males) mean and standard deviation are fixed to 0 and 1 respectively to identify the scale, all item parameters are fixed, and the mean and standard deviation in the focal group (here females) are estimated. In step two, the means and standard deviation for the focal group are

fixed to the values from step 1, allowing the Wald test to be computed on an item by item basis.

Woods et al. (2013) found Wald-2 performed poorly in their simulations with increased type I error rate. Their results suggested an alternative statistic, Wald-1 (Cai, Thissen, & du Toit, 2011), performed better. Here, a single item is selected as an anchor. The mean and standard deviation are fixed at 0 and 1 in the reference group, and item parameters on the anchor item are also fixed. The mean and standard deviation in the focal group, and the item parameters of all other items are freely estimated.

However, there remains debate as to the most appropriate way to select anchor items. Commonly, previous studies can be used to select items for which there is no evidence of DIF to act as the anchor item. In the current application, no such prior evidence exists. As such, here we follow the suggestion of Woods et al. (2013) and first perform DIF analysis using the Wald-2 test to identify an item displaying least evidence for DIF. This item was then selected as the anchor item in a second analysis of DIF based on the Wald-1 statistic.

In both cases, the test compares the differences in parameter estimates (here difficulty and discrimination in the 2PL model) across the reference and focal group, with degrees of freedom equal to the number of parameters being assessed. Freeing a single parameter within a given item similarly provides a 1 degree of freedom test. Conducting this test on a difficulty parameter represents a test of uniform bias (bias that does not vary with the level of the construct) whereas the same test on a discrimination parameter represents a test of non-uniform bias (bias that varies with the level of the construct).

Results

Descriptive statistics

Proportion of endorsement for the LDSQ items by gender group are provided in Table 1. For the purpose of the analyses, reverse scored items were re-coded such that endorsing an item indicated a greater level of ability (less impairment).

(Insert Table 1 about here)

Differential Item Functioning

Initially, on fitting the 2PL, the standard error for difficulty parameter of the 'Time' item in the male group was an extremely large value (9.32), indicating estimation problems. Given this and the fact that we had a large amount of missing data for this item, we repeated analyses excluding this item. Resulting parameter estimates for the 2PL fit to each gender group individually are provided in Table 2. These all appeared to converge to reasonable values.

(Insert Table 2 about here)

Item fit statistics are provided in Table 3. No items showed evidence of mis-fit. Inspection of the LDχ2 test for all item pairs indicated no evidence of LD. Values ranged from ±0.0 to 2.6 for males and ±0.0 to 0.7 for females.

(Insert Table 3 about here)

Results of the DIF analyses are provided in Table 4. The left hand columns show the results for the two-step Wald-2 analysis. Here, no evidence for DIF was present in any of the items. We selected the 'School' item to act as an anchor for the second analysis as it was associated with the largest p-value (see Table 4). We note this decision was somewhat arbitrary in this instance as there was no evidence of DIF in any item. The results from the second analysis are shown in the right-hand columns. Confirming our conclusions from the initial analysis, there was no evidence of DIF for any of the items.

(Insert Table 4 about here)

Discussion

11

The primary aim of the current study was to investigate differential item functioning (DIF) across gender in the LDSQ. If DIF had been present, it would have indicated that the items performed differently across the two groups with the result that scores for males and females on the LDSQ would not have been comparable. Our results indicated that no DIF was present in a core of items of the LDSQ, therefore, supporting the continued use of the LDSQ to assess ID across males and females.

However, we were unable to assess DIF for the entire set of LDSQ items. First, we had to combine the items referring to reading ability and writing ability into a single 'literacy' item.   The necessity to combine Read and Write items was based on an extremely high correlation (rho = 0.95) between the items. Examining the response patterns for these two items shows that 299 participants (87%), gave the same response for Read as they did Write (both yes = 152; both no = 105; both missing = 42). From a purely statistical perspective, such high correlations between items can cause estimation problems. However, from a clinical perspective, the correlation between the two abilities is interesting. During the initial development of the LDSQ, separate items for reading ability and writing ability were included because of clinical observations suggesting that (albeit limited) reading ability could exist in the absence of any writing ability and vice versa. While there were some individuals in the sample for whom this was true, our results suggest that dissociations between reading and writing ability are rare. For future revisions of the LDSQ it may, therefore, be more efficient to drop one of the redundant items or replace them with a single literacy item.

The Time variable was also dropped from the analysis due to an extremely standard error estimates for the item parameters, indicating estimation problems. This is undesirable, because this means that we were unable to conduct a test of DIF that included all LDSQ items. It is unfortunate that we were unable to assess this item because previous research has suggested that this item is a useful indicator of ID. For example, it has shown a high

discrimination parameter (a scalability coefficient of 0.75 in a Mokken analysis) in previous research (Murray & McKenzie, 2013). Furthermore, because it represents a skill that is relatively easier than the other skills measured by the LDSQ (see the endorsement rates in Table 1), it contributes to the reliable measurement of ID at the more severe levels of impairment. On the other hand, the large amount of missing data for the time item could be an indicator that the clinical utility of the item is limited in other ways. For example, it may indicate that raters have difficulty in judging with a high degree of certainty whether an individual is able to tell the time.

We were unable to assess this possibility in the current study because some of the data were collected retrospectively and relied on case notes or clinician knowledge of a case. Missing data in these instances could simply reflect not happening to have recorded this information in case notes, rather than any inherent difficulty in assessing it. Future data collection could, therefore, aim to explore a possible trade-off between the how informative this item is about the presence of intellectual disability and how easy it is to assess in practice. Of course, will also be an important future direction to assess whether the time item exhibits DIF across males and females.

A further limitation of the current study is the sample size of 211males and 132 females. Given that the LDSQ contains a relatively small number of items (making for a comparatively simple IRT model) we deemed these sample sizes to be sufficient. However, samples of this size would still be considered to be relatively small for the analysis of DIF meaning that current study was only powered to detect DIF of a moderate to large effect size: smaller more subtle DIF effects may have gone undetected in the current study. In addition, due to confidentiality restrictions we had only limited information about the demographic characteristics of the sample. As such, it would be beneficial to attempt to replicate the DIF analysis in larger, better characterised samples.

Conclusion

Overall, our findings did not suggest the presence of DIF across males and females in the items of the LDSQ. This provides justification for the continued use the LDSQ in males and females in the same way in clinical and research settings. As noted, an important caveat is that we were unable to assess the time item of the LDSQ and used a combined literacy item in place of separate items referring to reading and writing ability.. While it will be beneficial to resolve these issues in future research, our initial investigation of the performance of the LDSQ provides encouraging results.

References

Borsboom, D., Romeijn, J. W., & Wicherts, J. M. (2008). Measurement invariance versus selection invariance: Is fair selection possible? *Psychological Methods*, *13*, 75-98.

British Psychological Society. (2000). *Learning disability: Definitions & contexts*. Leicester: United Kingdom.

Cai, L., Thissen, D., & du Toit, S. H. C. (2011). *IRTPRO: Flexible, multidimensional, multiple categorical IRT modeling [Computer software]*. Lincolnwood, IL: Scientific Software International.

Chen, W.-H., & Thissen, D. (1997). Local dependence indices for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, *22*, 265-289.

Harrison, P., & Oakland, T. (2003). *Adaptive Behavior Assessment System (ABAS-II).* San Antonio, TX: The Psychological Corporation.

Johnson, W. & Bouchard, T.J. Jr. (2007). Sex differences in mental ability: *g* masks the dimensions on which they lie. *Intelligence, 35,* 23-39.

Kim, E. S., & Yoon, M. (2011). Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling*, *18*, 212-228.

Langer, M. (2008). *A reexamination of Lord's Wald test for differential item functioning using item response theory and modern error estimation* (Unpublished doctoral dissertation). Chapel Hill: University of North Carolina.

Maulik, P. K., Mascarenhas, M. N., Mathers, C. D., Dua, T., & Saxena, S. (2011). Prevalence of intellectual disability: a meta-analysis of population-based studies. *Research in Developmental Disabilities, 32*, 419-436.

McKenzie, K., & Paxton, D. (2006). Promoting access to services: the development of a new screening tool. *Learning Disability Practice*, *9*, 17-21.

McKenzie, K., Michie, A., Murray, A., & Hales, C. (2012). Screening for offenders with an intellectual disability: The validity of the Learning Disability Screening Questionnaire. *Research in developmental disabilities*, *33*, 791-795.

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care, 44*, S69-S77.

Millsap, R. E. (2012). *Statistical approaches to measurement invariance*. New York: Routledge.

Murray, A. L., & McKenzie, K. (2013). Estimating the severity of intellectual disability in adults: A Mokken scaling analysis of the Learning Disability Screening Questionnaire. *Psychological Assessment*, *25*, 1002-1006.

Murray, A.L., & McKenzie, K. (2014). The accuracy of the Learning Disability Screening Questionnaire (LDSQ) in classifying severity of impairment: A brief report. *Journal of Intellectual and Developmental Disability*. Available online DOI:10.3109/13668250.2014.936836

Orlando, M., & Thissen, D. (2000). Likelihood-based item fit indices for dichotomous item response theory models. *Applied Psychological Measurement, 24*, 50-64.

Paxton, D., McKenzie, K., & Murray, G. C. (2008). Putting screening tools to the test. *Learning Disability Practice*, 11, 14–18.

Schalock, R. L., Luckasson, R. A., & Shogren, K. A. (2007). The renaming of mental retardation: Understanding the change to the term intellectual disability. *Intellectual and Developmental Disabilities*, *45*, 116-124.

Silverman, I., Choi, J., Mackewn, A., Fisher, M., Moro, J., & Olshanksy, E. (2000). Evolved mechanisms underlying wayfinding: further studies on the hunter-gatherer theory of spatial sex differences. *Evolution and Human Behavior, 21,* 201-213

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (2005). *Vineland-II adaptive behavior scales*. San Antonio: Psychological Corporation.

Wechsler, D. (2008). *Wechsler adult intelligence scale–Fourth Edition (WAIS–IV)*. San Antonio, TX: NCS Pearson.

Woods, C. M., Cai, L., & Wang, M. (2013). The Langer-Improved Wald Test for DIF Testing With Multiple Groups Evaluation and Comparison to Two-Group IRT. *Educational and Psychological Measurement*, *73*, 532-547.

Table 1

LDSQ items and proportion of endorsement in males and females

Item	Male		Female	
	N	Proportion Endorsed	N	Proportion Endorsed
Time	133	.65	82	.60
Read	180	.42	118	.47
Write	171	.36	118	.46
Literacy (Read + Write)	211	.50	132	.58
Independent Living	199	.23	132	.34
Employment	197	.21	132	.22
Previous Contact	191	.32	129	.33
School	189	.24	131	.29

Table 2

Item parameter estimates for male and female groups

Item	Male						Female					
	a	*s.e.*	*c*	*s.e.*	*b*	*s.e.*	*a*	*s.e.*	*c*	*s.e.*	*b*	*s.e.*
Independent	1.13	0.36	-1.51	0.26	1.34	0.35	1.39	0.39	-1.31	0.37	0.94	0.22
Job	0.90	0.32	-1.49	0.23	1.66	0.51	0.43	0.20	-1.44	0.25	3.35	1.42
Prev	1.31	0.38	-0.95	0.23	0.72	0.20	1.47	0.43	-1.42	0.41	0.97	0.22
School	2.50	1.15	-2.19	0.77	0.88	0.17	2.50	0.97	-2.54	0.82	1.02	0.22
Literacy	0.96	0.31	-0.35	0.18	0.36	0.21	1.53	0.47	0.06	0.29	-0.04	0.19

Note: a = slope; b= threshold; c = intercept. IRTPRO estimates a and c parameters, with values for b derived from these estimates with the relation; $a\Theta+c$ or $a(\Theta-b)$.

Table 3

Item fit for males and females

	Male			Female		
Item	χ^2	*df*	**Probability**	χ^2	*df*	**Probability**
Independent	2.40	3	.50	3.30	3	.35
Job	1.48	3	.69	4.15	3	.25
Prev	2.11	3	.55	5.88	3	.12
School	2.17	3	.54	3.00	3	.39
Literacy	1.47	3	.69	1.03	2	.60

Table 4

DIF results for LDSQ in males and females from the Wald-2 and Wald-1 tests

	Wald-2 All Items Anchor						**Wald-1 School Anchor**					
	Total χ^2 (df)	**p**	**χ^2_a (df)**	**p**	**$\chi^2_{c\|a}$ (df)**	**p**	**Total χ^2 (df)**	**p**	**χ^2_a (df)**	**p**	**$\chi^2_{c\|a}$ (df)**	**p**
Independent	1.2(2)	0.553	0.2(1)	0.624	0.9(1)	0.332	2.0(2)	0.374	0.1(1)	0.761	1.9(1)	0.171
Job	1.7(2)	0.418	1.5(1)	0.221	0.2(1)	0.623	1.2(2)	0.553	1.0(1)	0.308	0.1(1)	0.704
Previous Experience	1.2(2)	0.547	0.1(1)	0.780	1.1(1)	0.289	0.2(2)	0.911	0.0(1)	0.857	0.2(1)	0.695
School	0.4(2)	0.814	0.0(1)	0.998	0.4(1)	0.522	-	-	-	-	-	-
Literacy	2.8(2)	0.243	1.0(1)	0.315	1.8(1)	0.177	2.9(2)	0.232	0.3(1)	0.584	2.6(1)	0.106

Figure Captions

Figure 1

Distribution of FSIQ scores in males

Figure 2

Distribution of FSIQ scores in females