

¹ Gender Differential Item Functioning of Test Items of 2015 Joint
² Senior Secondary II Mathematics Promotion Examination in
³ Ondo State, Nigeria

4 Oloda Festus Sunday Smart (Ph.D)

Received: 15 September 2021 Accepted: 30 September 2021 Published: 15 October 2021

7 Abstract

8 The research is aimed at finding out whether the multiple-choice mathematics items of Ondo
9 State Joint Senior Secondary II Promotion Examination (OSJSSPE) administered in 2015
10 function differentially in terms of the gender (male and female) of the examinees and also
11 investigate the general performance of students in the OSJSSPE multipleâ??choice
12 mathematics items administered in 2015. The study employed an ex-post facto research
13 design. A sample of 3,135 examinees was selected from a population of 52,922 examinees who
14 sat for the examination using two-stage random sampling techniques. One research question
15 was raised, and one hypothesis formulated and tested for significance at $p<0.05$ level. The
16 analyses revealed that the general performance of the students in the OSJSSPE multipleâ??choice
17 mathematics items administered in 2015 was high. However, the examinees of equal
18 ability from both male and female students had a different probability of answering some
19 items correctly; thus, the multiple-choice mathematics items of OSJSSPE administered in
20 2015 functioned differentially on male and female examinees.

Index terms— differential item functioning, item biased, joint senior secondary ii promotion examination, latent trait, mathematics.
??2017)

, mathematics is a discipline that has various areas of studies which relate with other discipline or subjects such as Basic Science, Basic technology and others which plays important role in the security, sustainability and technological development of any nation. In other words, mathematics is the linchpin in the task of technological development of any nation. Ale & Adetula (2010) highlights the intricate link of mathematics to science, and technology and opined that without the knowledge of mathematics, there would be no science, and without science, there will be no technology and without technology, there is no modern society.

Despite the importance accorded mathematic, it has been observed that students still perform poorly at both internal and external examinations. The poor performance of students at the secondary school level has been a concern to the public in Nigeria. This is usually noticed when the yearly West African Senior School Certificate Examination (WASSCE) and Senior School Certificate Examination (SSCE) conducted by National Examination Council (NECO) results are released. This is evident in the performance of Nigerian candidates in general mathematics in the West African Secondary School Certificate Examination (WASSCE) from 2006 to 2011, which revealed that the percentage of students that passed mathematics at credit level was between 39% and 47%, except for a little improvement of 57.27% in 2008.

40 One of the factors attributed to the poor performance of students in mathematics achievement tests in schools
41 as observed in literature is gender. There is the general belief that male students seem to perform better than
42 female students in mathematics and mathematics-related subjects. According to Smith & Walker (1988) and
43 Popoola & Ajani (2011), male students perform better than female students in mathematics. Etukudo (2002)
44 opined that boys generally perform better than girls even though they are put into the same classroom situation.

Gender Differential Item Functioning of Test Items of 2015 Joint Senior Secondary II Mathematics Promotion Examination in Ondo State, Nigeria

45 However, Cronbach (1977) was of the opinion that boys and girls do not differ much on ability measures. This
46 is not unconnected with the fact that opportunities for development are much more the same until the school
47 leaving age.

48 Multiple-choice is one of the various types of objective tests that can be used to measure learning outcomes; the
49 score of the multiple-choice item is independent of the subjective influence of the marker or the examiner. Thus,
50 the individual examiner doing the scoring is not required to make the judgment; the score is consistent regardless
51 of the prejudice of the examiner. According to ??arret (2001), multiple-choice tests are generally biased towards
52 males, while the female students experience more difficulties with questions involving numerical, spatial, or high
53 reasoning skills. It is generally believed that multiple-choice tests are prone to guessing. Thus, guesswork in
54 most multiple-choice achievement tests has become the order of the day in most institutions. Ojerinde ??1985)
55 reported that in Nigeria, most students who do not have a flare for mathematics would result in guesswork at
56 the secondary and tertiary levels. This is not unconnected with students testing their luck and thus gets out of
57 the examination hall as quickly as possible. However, no statistics exist in this regard as to why students resort
58 to guesswork, but if critically viewed within the test theories, it is assumed to be due to some psychological and
59 situational factors.

60 Also, Lee, as cited in Adebule (2013), observed that questions always arise concerning whether high average
61 test scores by certain groups are due to actual achievement differences, bias in test, or a combination of both.
62 Conversely, the favored groups are the advantaged group during promotion and admission or selection into
63 science-based courses in the high institutions while, the disadvantaged groups, on the other hand, are disallowed
64 due to some factors tagged extraneous and irrelevant variables that interfere with the measurement of the
65 underlying psychological construct being measured. These factors relate to the group like gender, socio-economic
66 status, location has significant influence on the examinees' response to the item. This implies that the test is
67 multidimensional or measures more than one trait.

68 Consequently, Smith (1985) referred to the above as measurement disturbance which was classified by Smith
69 (1985) into three categories: 1. Disturbances as a result of characteristics of the person that is independent of
70 the items such as fatigue, boredom, illness, cheating, among others. 2. Disturbances as a result of interaction
71 between the characteristics of the person and the properties of the item, such as item content, item type, guessing,
72 item bias, among others. 3. Disturbances as a result of the properties of the items which are independent of the
73 characteristics of the person, such as the person's ability and item difficulty.

74 Many research findings have shown that there are differences in the academic performance of separate groups
75 since the inception of testing. An item is biased if it discriminates between members of different groups who have
76 the same ability on what is being measured. Put differently, members of different groups who have the same
77 trait level differ in their score on the item. In his contribution, Plake, as cited in Adebule (2009), defines bias in
78 tests as a situation when items in an achievement tests are found to favour one group over another for reasons
79 not explainable by differences in achievement level between groups. Generally, bias in test items is regarded
80 as a systematic error in measurement. Item bias is the degree to which items that comprise a measurement
81 scale are systematically related to various exogenous variables (e.g., age, gender, location, race, and so on) after
82 conditioning on the latent variable of interest. Items that show bias in any measuring instrument may affect the
83 properties of the measuring instrument; for instance, test items that are for a group of equal ability should not
84 have statistical differences between the groups. However, if differences exist between the groups, then the validity
85 of the test items is threatened.

86 A test is gender-biased if men and women with comparable ability levels tend to obtain different scores.
87 Thus, we can say that the test contained items that measure different traits for male and female examinees with
88 comparable abilities. Test fairness is very important in test development; a test is said to be fair if systematic
89 errors (biases) are not present. A systematic error occurs when the construct measured by a test contains some
90 irrelevant elements that can threaten the validity of the test. A test should enable all examinees to have an equal
91 chance to demonstrate personal skills and knowledge vital to the purpose of the test. Items that show bias in any
92 measuring instrument may affect the properties of the measuring instrument. Thus, score generated from a test
93 that contains items that are biased against one group or the other or test result from unfair testing procedures
94 cannot be used to make a valid quality decisions in education.

95 Item response theory (IRT), also known as the latent trait theory or strong true score theory, is a family of
96 latent trait models that are used to establish psychometric properties of items. According to Ojerinde, Popoola,
97 Ojo, & Onyeneho (2012). Item response theory connotes and theoretically assumes that there exist a relatively
98 common trait or characteristic that can be used to determine an individual's ability to succeed with a particular
99 task. Such tasks may be in terms of the individual's response by thinking (cognitive), feeling (affective), and
100 acting (psychomotor). This theory is considered to be one of the most important developments in psychological
101 testing in recent times. Item response theory (IRT) model was designed to solve the problems of classical test
102 theory (CTT); it has been the only theory that gives the valid measurement in terms of test construction and
103 interpretation for assessment of test taker's ability in psychometric analysis. Many of the recent development in
104 testing have their origin in the concept of IRT, such as tailored testing, adaptive computer testing, and improved
105 equating of test forms. According to Pine (1977), Item bias based on the Item response theory (IRT) concept
106 is classified as unbiased if all individuals having the same underlying trait (ability) have an equal probability of
107 getting the item correct regardless of subgroup membership.

108 Differential item functioning (DIF), also referred to as measurement bias, exists when persons with different
109 group membership but identical overall test scores have different probabilities of solving a test item correctly or
110 giving a certain response on a Differential Item Functioning (DIF) occurs whenever people of the same ability
111 level but different groups have different probabilities of endorsing an item. The focus of DIF analysis is on
112 differences in performance between groups that are matched concerning ability, knowledge or skill of interest.
113 ??ee (1990) opined that to investigate bias at the item level, developers of large-scale assessments usually conduct
114 a differential item functioning (DIF) analysis. However, not all cases of DIF necessarily have to be interpreted as
115 item bias that will jeopardize the fairness of the test. Instead, DIF is a necessary but not sufficient condition for
116 item bias. Thus, if DIF is not apparent for an item, then no item bias is present. However, if DIF is present, then
117 its presence is not a sufficient condition to declare the item bias; rather one would have to apply a follow-up item
118 bias analysis (e.g. content analysis, empirical evaluation) to determine the presence of item bias. For instance, if
119 in a mathematics test, students of equal ability level from urban areas display a higher probability of answering an
120 item correctly more than students from rural areas of equal ability level because the content in the test is biased
121 against students from rural areas, then we say the item exhibit DIF and should be considered for modification
122 or removal from the test items.

123 The Ondo State Joint Senior Secondary II Promotion Examination (OSJSSPE) was introduced as an
124 intervention measure to reduce the poor performances of students in public examinations. It is only those students
125 that passed the OSJSSPE both in public and private secondary schools that the state government would pay
126 their registration fees and also allow them to sit for the West African Senior School Certificate Examination
127 (WASSCE) and the Senior School Certificate Examination (SSCE) conducted by NECO in the Ondo State. The
128 senior secondary II students that take these examinations are expected to have been exposed to the same course
129 content at the same time frame within the same number of periods. Thus, they are supposed to be of equal
130 probability of success irrespective of gender, location, and type of school.

131 The 2014 Ondo State Joint Senior Secondary II Promotion Examination (OSJSSPE) has undergone criticism
132 from stakeholders in the education sector due to the failure rate, which was as bad as having less than half of
133 the senior secondary II students in some schools promoted. Some critics believed that the state government,
134 through the state functionaries in the ministry of education, influenced the result of the examination of the
135 public schools to reduce the number of students in the Senior Secondary School III (SSS3), which will, in turn
136 reduce the total amount to be paid to WAEC or NECO for registration by the state government through the
137 ministry of education. The study investigated the differential item functioning (DIF) of all items in mathematics
138 multiple-choice items of the 2015 Ondo State Joint Senior Secondary II Promotion Examination concerning the
139 gender (male and female). The study was guided by the research question: What is the general performance of
140 the students in the OSJSSPE multiple-choice mathematics items administered in 2015? Also, a Null hypothesis
141 was postulated to guide the study; that is, the Ondo state Joint Senior Secondary II Promotion Examination
142 (OSJSSPE) multiple-choice mathematics items administered in 2015 will not function differentially between males
143 and female examinees.

144 1 II.

145 2 Research Method

146 The study adopted an ex-post-facto research design. For the ex-post facto design, the researcher started with
147 the observation of the dependent variable and then studied the independent variables in retrospect for their
148 possible relation to an effect on the dependent variable(s). This design is therefore relevant to this study because
149 the researcher does not have direct control over the independent variables since the manifestations had directly
150 occurred and the analysis would be performed on existing data. The population for the study consisted of
151 52,922 of male and female students in the senior secondary II that responded to the 50 multiplechoice items
152 in mathematics of Ondo State Joint Senior Secondary II Promotion Examination (OSJSSPE) administered in
153 2015. The total sample for the study consisted of 3,135 senior secondary II students that responded to the
154 50 multiple-choice items in mathematics of OSJSSPE administered in 2015 as contained in the Optical Mark
155 Recorder (OMR) sheets from twenty-four selected senior secondary schools in Ondo State, Nigeria, using two-
156 stage sampling techniques. In the first stage, two local government areas (LGA) in each of the three senatorial
157 districts of Ondo State were selected using the purposive sampling technique. In stage two, two public schools
158 (one each from rural and urban areas) and two private schools (one each from rural and urban areas); thus, 12
159 public schools and 12 private schools were selected using a stratified sampling technique. A total of 24 schools
160 were used for the study. The instruments used for the study were the responses of all the sampled students to
161 the 50 multiple-choice items in mathematics of the OSJSSPE administered in 2015 in the selected schools as
162 contained in the Optical Mark Recorder (OMR) sheets. The items were already subjected to the processes of
163 validation and standardization by the examination Department of Ondo State Ministry of Education. Thus, they
164 were already valid and reliable instruments.

165 The 50 multiple-choice items in mathematics which the students responded to and were used for this study
166 was constructed, conducted, and administered by the Examination Department of the Ministry of Education in
167 Ondo State. The OMR sheets comprised of section A and section B. Section A contains the demographic data
168 of the respondents, while section B consisted of all items whose differential item functioning was determined.

8 RECOMMENDATIONS

169 Descriptive statistics like percentages was used to answer the research question, while inferential statistics like
170 the Welch t-test was used to test the hypothesis at a 0.05 level of significance.

171 3 III.

172 4 Results

173 The results of the analysis are presented below. What is the general performance of students in the OSJSSPE
174 multiple-choice mathematics items for 2015? Table 1 shows that 1686 (53.8%) of the students who sat for
175 OSJSSPE had Distinction in Mathematics, 579 (18.5) had credit, 192 (6.1%) passed, while 678 (21.6%) failed.
176 This implies that the General performance of the students in the OSJSSPE multiple-choice mathematics items
177 for 2015 is high. The performance of the students in the OSJSSPE multiplechoice mathematics items for 2015 is
178 further depicted in the figure below. A cursory look at Table 2 above shows that 28 (56%) items were flagged as
179 DIF against both male and female examinees since their P value were less than 0.05. 13 items were flagged as DIF
180 items against male examinees. The flagged items against male examinees are 8, ??6, ??7, ??0, ??2, ??5, ??3, ??4,
181 ??5, ??7, ??2, ??7 and 50. This is because their p -values are less than 0.05 in each case. Similarly, 15 items were
182 flagged as DIF items against female examinees because their p-values are less than 0.05 in each case. The items
183 that flagged as DIF items against female examinees are as follows: items 4,9,13,14,15,19,21,26,36,38,40,43,44,45
184 and 46 respectively. The study showed that 28 items listed above are statistically functioning differentially
185 between males and females at the significance level of 0.05. The null hypothesis is rejected. This implies that
186 the OSJSSPE multiple-Choice Mathematics items administered in 2015 function differentially between males and
187 females examinees.

188 IV.

189 5 Discussion

190 The result showed that the general performance of the students in the OSJSSPE multiple-choice mathematics
191 items administered in 2015 was high. The findings also revealed that 28 (56%) of the OSJSSPE multiple-choice
192 mathematics items administered in 2015 displayed DIF based on the gender of the examinees using a statistically
193 significant level of P-value of 0.05 ($P < 0.05$) (Linacre, 2010a). 13 items favoured male examinees while 15 items
194 favoured the female examinees. The null hypothesis is rejected. This implies that the OSJSSPE multiple-choice
195 mathematics items administered in 2015 functioned differentially between male and female examinees. This
196 finding conforms with the finding of Adedoyin (2010) that mathematics items were gender-biased. Also, the
197 result agrees with the findings of Banabas (2012) that mathematics questions used by WAEC contained items
198 that measure different things for male and female examinees with the same mathematics ability. Abedalaziz
199 (2010) and Madu (2012) also reported that mathematics items exhibit differential item functioning in favour
200 of male examinees. However, the study contradicts Adebule (2013) that mathematics items did not function
201 differentially based on the gender of the examinees.

202 V.

203 6 Conclusion

204 The study investigated items that exhibit differential item functioning in the 2015 Joint Senior Secondary II
205 Mathematics Promotion Examination in Ondo State, Nigeria. Based on the findings, it was concluded that the
206 general performance of the students in the OSJSSPE multiple-choice mathematics items administered in 2015
207 was high. Also, examinees of equal ability from both male and female students had a different probability of
208 answering some items correctly; thus, the multiple-choice mathematics items of OSJSSPE administered in 2015
209 functioned differentially on male and female examinees.

210 7 VI.

211 8 Recommendations

212 Based on the findings, the following recommendations are made: 1) Examination bodies, Test experts, and
213 people charged with developing, validating, and administering tests need to carry out differential item functioning
214 analysis for all items before administration of a test.
215 2) There is a need for teachers, officials of the examination department in the ministry of Education to be
216 trained by experts on item writing. This would acquit them with the processes of finding the psychometric
217 properties and the detection of DIF of each item, which will, in turn, improve the quality of students' assessment
218 in our schools. 3) Detection of DIF has been investigated using OSJSSPE multiple-choice mathematics items in
219 this study. There is a need for replication using other cognitive subjects in OSJSSPE.

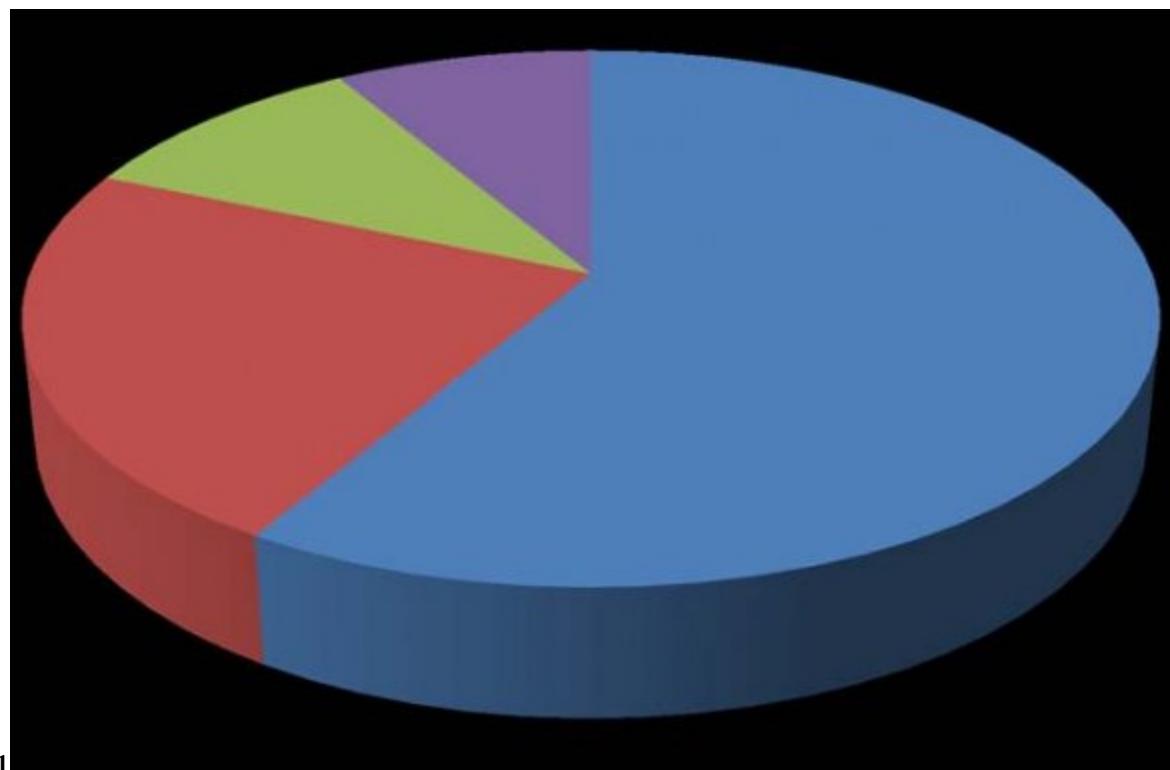


Figure 1: Figure 1 :

Introduction

The importance of mathematics in the manpower and technological development of a nation cannot be overemphasized. According to Oloda and Fakinlede

Figure 2:

1

Academic Performance	N	2015	%
Fail (0 -39)	678	21.6	
Pass (40 -49)	192	6.1	
Credit (50 -69)	579	18.5	
Distinction (70 -100)	1686	53.8	
Total	3135	100.0	

Figure 3: Table 1 :

8 RECOMMENDATIONS

2

Item Numb	Item Name	Joint S E	Gender DIF Contras	male = 1 female = 2		Bias Against	Decision
				t	Rasch df	-Welch Prob	
er 1	I0001	.11	.19	.19	INF	.0729	No DIF
2	I0002	.10	-.03	-.03	INF	.7534	No DIF
3	I0003	.11	.00	.00	INF	1.000	Female No DIF
4	I0004	.10	-.29	-.29	INF	.0045	DIF
5	I0005	.10	.00	.00	INF	1.000	No DIF
6	I0006	.10	.00	.00	INF	1.000	Male No DIF
7	I0007	.10	.15	.15	INF	.1605	Female No DIF
8	I0008	.10	.27	.27	INF	.0077	DIF
9 10	I0009 I0010	.09 .09	-.46 .00	-.46 .00	INF	.0000 1.000	Female DIF No
11 12	I0011 I0012	.10 .10	.00 -.18 -	.00 -.18 -	INF	1.000 .0703	Female DIF No
13 14	I0013 I0014	.09 .10	.48 -.21	-.48 -.21	INF	.0000 .0340	Female DIF No
					INF		Male DIF DIF
					INF		Male DIF
15 16	I0015 I0016	.10 .09	-.25 .21	-.25 .21	INF	.0117 .0177	Female DIF DIF
17 18	I0017 I0018	.09 .10	.36 -.06 -	.36 -.06 -	INF	.0000 .5396	Male DIF No
19	I0019	.09	.54	-.54	INF	.0000	Female DIF DIF
					INF		Male
20 21	I0020 I0021	.10 .10	.23 -.25	.23 -.25	INF	.0235 .0087	Male DIF DIF
22	I0022	.09	.59	.59	INF	.0000	Female DIF
					INF		
23	I0023	.10	.00	.00	INF	1.000	No DIF
24	I0024	.10	-.12	-.12	INF	.2269	No DIF
25	I0025	.09	.32	.32	INF	.0004	DIF
26	I0026	.09	-.76	-.76	INF	.0000	DIF
27 28	I0027 I0028	.10 .10	-.06 .12	-.06 .12	INF	.5236 .2319	Male No DIF
29 30	I0029 I0030	.09 .10	.07 -.18	.07 -.18	INF	.4365 .0642	Male No DIF
31 32	I0031 I0032	.10 .09	.00 .14	.00 .14	INF	1.000 .1286	Male No DIF
33	I0033	.10	.44	.44	INF	.0000	Female No DIF
					INF		Male No DIF
					INF		Female No DIF
					INF		DIF
34 35	I0034 I0035	.10 .09	.42 .92	.42 .92	INF	.0000 .0000	Female DIF DIF
					INF		
36 37	I0036 I0037	.10 .09	-.22 .36	-.22 .36	INF	.0268 .0001	Male DIF DIF
38 39	I0038 I0039	.10 .09	-.27 .00	-.27 .00	INF	.0062 1.000	Female DIF No
40 41	I0040 I0041	.09 .10	-.20 .12	-.20 .12	INF	.0257 .2148	Female DIF DIF
42	I0042	.10	.52	.52	INF	.0000	Female No DIF
					INF		Female DIF
					INF		Male
43 44	I0043 I0044	.09 .09	-.57 -.22	-.57 -.22	INF	.0000 .0131	Male DIF DIF
					INF		
45	I0045	.10	-.84	-.84	INF	.0000	DIF
46	I0046	.10	-.49	-.49	INF	.0000	DIF
47	I0047	.09	.62	.62	INF	.0000	DIF
48	I0048	.10	-.02	-.02	INF	.8247	No DIF
49	I0049	.09	-.14	-.14	INF	.1065	No DIF
50	I0050	.10	.61	.61	INF	.0000	DIF

220 .1 Conflict of Interest Statement

221 The author declares no conflicts of interest regarding the article entitled "Gender Differential Item Functioning
222 of Test Items of ??015

223 [Adebule ()] 'A study of differential item functioning in Ekiti State Unified Mathematics Examination for Senior
224 Secondary Schools'. S O Adebule . *Journal of Education and Practice* 2013. (17) p. 4.

225 [Abedalaziz ()] 'Agender-related differential item functioning of Mathematics test items'. N Abedalaziz .
226 *International Journal of Educational and Psychological Assessment* 2010. 5 (1) p. 101.

227 [Cronbach ()] L Cronbach . *Educational Psychology*, (New York) 1977. Harcourt Brace Jovanovichs.

228 [Berret ()] 'Differential item functioning: A case study of first year economics students'. S Berret . *International
229 Education Journal* 2001. 2 (3) .

230 [Adebule ()] 'Identification of biased items in school certificate mathematics achievement test'. S O Adebule .
231 *International Journal of Research in Education* 2009. 1 (2) .

232 [Ale and O Adetula ()] 'The National Mathematical Centre and the mathematics improvement project in nation
233 building'. S Ale , L O & Adetula . *NMC Journal of Mathematical Sciences Education* 2010. (1) p. 1.

234 [Adedoyin ()] 'Using IRT approach to detect gender biased items in public examinations: a case study from
235 Bostwana Junior certificate examination in mathematics'. O O Adedoyin . *Educational Research and Review
236 Academic Journal* 2010. (7) p. 5.