The Response Patterns of the Career Interest Instrument Based on Holland's Theory

Farida Agus Setiawati, Yulia Ayriza, Endah Retnowati, Rizki Nor Amelia
Universitas Negeri Yogyakarta

This study aims to identify: patterns of responses, the item parameters, and the possibility of gender bias in the career interest instrument developed by the authors based on the Holland's theory. The sample of this study was 576 elementary students in Daerah Istimewa Yogyakarta who were recruited using the cluster random sampling method. Two parameters were employed to analyze the response patterns using BILOG program. The results were: (1) three items have inappropriate response patterns to the model; (2) all items of the career interest instrument showed good item parameter criteria; and (3) ten items were identified containing Differential Item Functioning (DIF) in relation to gender bias as shown by the Item Characteristic Curve (ICC). The implications of this study are this instrument can be used in assesing career interest of students and the information of biased items may be considered in the selection of careers for male and female students, including in scoring and interpretation.

Keywords: career interest, Holland's theory, item bias, item response theory, response pattern

Studi ini bertujuan mengidentifikasi: pola respons instrumen minat, parameter butir instrumen minat yang dikembangkan berdasarkan teori Holland, dan kemungkinan adanya bias gender butir-butir pada instrumen minat. Subjek studi ($N = 576$) siswa SD di DI Yogyakarta diperoleh melalui *cluster random sampling*. Instrumennya adalah instrumen minat yang disusun berdasarkan Teori Holland. Pola respons butir dianalisis menggunakan model 2 Parameter Logistik dengan bantuan Program BILOG. Hasil menunjukkan: (1) terdapat tiga butir yang memiliki pola respons yang tidak cocok dengan model; (2) semua butir dalam instrumen minat yang dikembangkan tersebut memiliki kriteria parameter butir yang baik; dan (3) terdapat sepuluh butir yang teridentifikasi memuat DIF berdasarkan gender berdasarkan hasil analisis menggunakan *Item Characteristic Curve* (ICC). Implikasi studi ini adalah didapatkannya instrumen minat karier yang dapat digunakan untuk mendeteksi minat karier siswa dan adanya informasi bias butir yang dapat dijadikan bahan pertimbangan dalam pemilihan karier pada siswa pria dan wanita. Informasi tersebut juga dapat digunakan untuk pertimbangan dalam pemberian skor dan interpretasinya.

Kata kunci: minat karier, teori Holland, bias butir, teori respons butir, pola respons

Success and happiness in accomplishing a particular activity is often considered related to the self-interest in performing the activity. A person who is interested in doing research will be happy to write scientific papers, and someone who is interested in arts will put great efforts to make their art performance impressing. In the workplace, a person who does not like to be a teacher is more likely to feel unhappy in her/his career as a teacher, than the one who is interested in teaching. Therefore, interest may be proposed as a critical factor to consider if an individual needs to make a decision, such as to receive or refuse certain careers or activities. This is supported by Holland (1997) who asserts that a person should easily perform in her/his career if she/he understands themselves, recognizes their own vocational ability and build interest based on those understandings.

Career comprehension is a process which may start from early childhood (Hartung, Porfeli, & Vondracek, 2008; Super, Savickas, & Super, 1996). Children may learn to recognize various careers around them, such as occupations held by adults in their family. Children's perceptions about those occupations might in-

Correspondence concerning this article should be addressed to Farida Agus Setiawati, Department of Psychology, Yogyakarta State University, Jalan Colombo No. 1, Caturtunggal, Yogyakarta, Indonesia 55281. E-mail: farida_as@uny.ac.id; faridaagus@yahoo.co.id

fluence their positive or negative attitudes toward such occupations. Positive attitudes – when further accompanied with opportunity to find more information about that occupation – may trigger interest on that specific career. On the contrary, negative attitudes toward a particular occupation may cause lack of interest on it. When children were not introduced to certain career paths, they may be uninterested simply because they do not have the knowledge about those occupations. When environment nurtures children's positive attitudes to an occupation, they tend to grow interest to take the career path more naturally.

The definition of interest is frequently associated with career, vocation, and occupation (McDaniels & Gysbers, 1992). The National Association of Vocational Guidance (as cited in Gies, 1990) defines career as a job performed by an individual. Arthur, Hall, and Lawrence (1989) viewed carrier as a person's work experience over a lifetime, so that all working people have a career. This is in accord with Department of Education and Science (1989) that defines career as various job roles which are held by an individual throughout his/her lifetime. Further, interest is defined as an individual's preference to observe certain objects (Savickas, 1999). Strong, Jr. (as cited in Savickas, 1999) defines interest as the feeling of like or dislike. Therefore, career interest can be understood as the feeling of like/dislike toward a job one has in his/her lifetime. The feeling of like might make a person pay attention to certain jobs in his/her life, while feeling of dislike makes a person ignore and avoid the job.

Holland is one of the pioneers in measuring career interest. He theorizes that the construct of career interest includes Realistic, Investigative, Artistic, Social, Enterprising, and Conventional (RIASEC) dimensions. Realistic is the types of career that do not require verbal and social skills, but rather, application-oriented. Investigative is the careers that required abstract and creative reasoning, and the ability to think logically and analyze problems. Artistic is the careers that involve feeling, sensing, imagination, and harmony. Social is the careers that related to helping others. These careers involve social interactions with others, responsibility and humanity. Enterprising is the careers that involve verbal ability to influence others. Conventional is the careers that require verbal and numerical abilities, practical-orientation, and a focus on routine and systematization.

In 1994, Holland created his career instrument measurement which is called the Self-Directed Search (SDS). Reardon and Lenz (1999) said that SDS provides information regarding a person's readiness for career decision making, and also the intensity and the duration of career interventions that might be effective in solving career problems. However, previous studies showed that SDS might be more appropriate in detecting adults' career interest rather than children's.

In 2006, Sverko and Babarovic tested the concurrent validity of an instrument that was developed based on Holland's theory which consists of 228-items. It measures a person's profile of RIASEC using four types of item: activities, competencies, occupations, and self-estimates. The results showed that the instrument has a good reliability and concurrent validity, which mean that Holland's RIASEC scores can be used to predict the educational program suitable for students' need. A similar study was conducted by Ayriza, Setiawati, and Triyanto (2016) who developed instruments to detect children's knowledge and interest on several careers which are appropriate for the Indonesian context. The construct validity and confirmatory factor analysis were used to investigate the instrument. The result showed that the instrument confirmed Holland's theory.

Factor Analysis is one of the methods to explore instrument characteristic as suggested by the classical theory of item analysis. This analysis put more emphasis on the whole item characteristics than on each item's characteristics. According to psychometric theory, the weakness of classical theory is that the instrument characteristic is depended on the measured subject samples. Psychometrical characteristics obtained from analyzing samples in one area might be different with another area. The weakness of classical theory is then amended by Item Response Theory (IRT).

IRT enables each item or question to gain response patterns. Scoring process using IRT approach is based on a model which relates a person's ability or theta (θ) to the probability to answer correctly on each item (Baker & Seock, 2004; Demars, 2010; Embretson & Raise, 2000; Hambleton & Swaminathan, 1985). This theory assumes that instrument characteristic is not bounded to group or sample characteristics, but to each individual or item. Consequently, the modern theory that can estimate fault on each individual and item is more favorable (Hambleton & Zaal, 1991).

The use of IRT in the analysis of an instrument that measures career interest has been reviewed by Turner, Betz, Edwards, and Borgen (2010). They analyzed the psychometric properties of career self-efficacy instrument based on the six themes in Holland's theory using item response theory. The results of their research suggests that only 7% of the items has poor index discrimination. Separately, Betz and Turner (2011)

extended the research in the online environment where the adaptive testing and the scoring process were based on IRT. In summary, IRT and adaptive testing have promising implications for developing a career assessment.

A slightly different analysis was conducted in a study by Oliveira, Taveira, Cadime, and Porfeli (2016). In this study, IRT analyzes was carried out more specifically using the Rasch model. The results showed that all items in the Career Exploratory Outcome Expectations Scale (CEOES) were declared fit into the Rasch models and have a significant positive contribution in the measurement of career decision-making self-efficacy. The Rasch model was also employed by Athanasou (2001) who tested the 24-item questionnaire among 2,709 high schoolers in Australia. This questionnaire measured vocational interest typologies based on Holland's theory. The results showed that only four items were not fitting into the Rasch model.

Item analysis using IRT method can also detect response bias in each item. Wetzel, Hell, and Passler's (2012) study identified 11 items which were biased in relation to gender: five items favoring male students and six items favoring female students. In another research detecting gender bias, Adedoyin (2010) used IRT 3-PL to analyze item characteristics curves of male and female groups. The findings showed that five items (item numbers 2, 3, 12, 15, and 31) were gender-biased.

Item bias in a test is unfair and inconsistent; it usually caused by contamination of external factors (Osterlind, 1983). An item is considered unbiased when the probability of correct answers on the item is not significantly different across test takers who have similar ability from the same population regardless their group membership (Camili & Shepard, 1994). Holland and Thayer (as cited in Camili & Shepard, 1994) called item bias as Differential Item Functioning (DIF) and test bias as Differential Test Functioning (DTF).

There are two types of group in detecting DIF: the focal group, where item biases are examined; and the reference group, which is used as a comparison. DIF measures whether or not the focal group is disadvantaged or advantaged as compared to the reference group (Naga, 2012). The focal group can be based on age, gender, race or ethnic, culture, disability, or language. In terms of gender, for example, females can be determined as a focal group while the reference group is males, and vice versa (Budiono, 2009).

DIF detection methods are categorized into two major groups: based on classical test theory and based on modern test theory. Methods which are based on classical test theory include: transformed item difficulty method, item discrimination method, chi-square method, log-linear method, Mantel-Haenszel method, standardization method, and logistic regression method. Methods which are based on modern test theory encompass: Item Characteristic Curve, Lord's chi-square method, Raju's extent method, and likelihood ratio test. These two major methods have their own strengths and weaknesses; studies showed that no method is superior to the other. Therefore, this study uses IRT approach in order to obtain maximum detection result. DIF detection method used was based on modern test theory, namely, Item Characteristic Curve (ICC).

Based on the explanation above, this study aims to demonstrate theoretical benefits in assessing the response patterns of career interest instruments, especially in assessing interest instruments using IRT. By analyzing response pattern of each item, it becomes possible to decide which items are good and which items are not good. The analysis of this study also identifies the item biases, which may provide information for the development of psychology in the measurement of career interest among Indonesian students.

In Indonesia, analysis of item bias have been studied by Budiono (2009), Rahayu (2010), Retnawati (2013), and Sudaryono (2012). They have conducted research to investigate response bias on cognitive instruments, especially National Examination instrument. In the current study, the researcher investigates response patterns of a non-cognitive instrument, that is, interest instrument based on Holland's theory. The items were developed based on various occupations, which could be responded differently by male and female students. Therefore, this career interest instrument might generate gender bias on each item. The research questions of the current study are: (1) How does the response pattern of career interest instrument based on Holland's theory?; (2) How does the items parameter of career interest instrument based on Holland's theory?; and (3) Are there any gender biases on the instruments of career interest based on Holland's theory?

Method

Participants

A secondary data, derived from Ayriza's research in 2015 were used in this study, comprising 576 students of elementary schools in the early grades (first,

second, and third grade) in the special region of Yogyakarta, Indonesia. The participants were chosen using a cluster random sampling method. In each of the five districts in Yogyakarta, five elementary schools were randomly chosen.

The Instrument

The instrument in this study is the career interest instrument in which the dimensions were based on Holland's theory: realistic, investigation, artistic, social, enterprising, and conventional. The instrument was constructed by Ayriza, Setiawati, and Triyanto in 2016. Each dimension consists of 10 items or jobs. Thus, there were 60 jobs in the instruments. The specifications and items can be found in Table 1 (see Appendix A). The students/participants responded by giving a check mark on the items that were more interesting than the other. All items were scored 1 or 0, where 1 is the score for selected items and 0 is the score for unselected items. The score of the instrument was interpreted as participant's career interest orientation or their profile of career interest.

All six dimensions in this multidimensional instrument were analyzed separately. The result of *Confirmatory Factor Analysis* (CFA) showed that all items in each dimension formed one factor. A dimension is considered fit the model if the probability *Chi Square* is more than 0.05 ($\chi^2 > 0.05$) and *Root Mean Square Error of Approximation* (RMSEA) < 0.08; where the RMSEA is the average size of the expected difference per degree of freedom *(df)* in the population. Using these two criteria, all dimensions of RIASEC were declared fit the latent variable, that is, the Holland's theory. The summary of CFA is displayed in Table 2, and the conceptual diagram of each dimension in Appendix B.

Data Analysis

The aim of this study is to analyze psychometric characteristics of the career interest instrument using items response theory (IRT). IRT basically analyzes items not the instrument; therefore the analyses were conducted on each item in all dimensions. Each dimension was analyzed separately.

The response patterns of the items were analyzed using BILOG program (Du Toit, 2003). Response patterns of the career interest instrument were analyzed using chi square statistics, χ^2. This analysis was conducted by comparing item's chi square value with the critical value of chi square distribution according to items' degree of freedom and significance level α. An item is considered unfit with the model if the value of χ^2 item $\geq \chi^2$ distribution of critical value. In contrast, an item is considered fit with the model if the value of χ^2 item is smaller or equal with χ^2 distribution value; it is considered fit with the model if χ^2 probability $\geq .01$. Significance level (*α*) = .01 is the default value in BILOG program with the degree of freedom which has been determined by that program (Mislevy & Bock, 1990).

Characteristics of item parameters were analyzed using the BILOG 2-PL program. Model 2-PL was chosen because there were two item parameters to be examined, namely: index of discrimination (a) and index of endorsement (b). According to Embretson and Reise (2000), the degree of difficulty in noncognitive instruments is the probability of endorsement or probability of support. In addition, the use of 2-PL model was based on the samples size in this study (576 students). This is in accordance with Thorpe and Favia (2012) who stated that some authorities suggested that 100 respondents are sufficienct for the 1PL or Rasch model with a dichotomously-scored test, and some suggested that as many as 200 are sufficient for more complex 2PL model, but others recommended at least 500. The probability of item bias was detected by employing modern approach, that is, by comparing Item Characteristic Curve of the two observed groups. ICC analysis was conducted using BILOG model 2-PL, Excel program, and SPSS.

Results

Response Patterns of Career Interest Instrument

Response patterns in this study are the items' response patterns or the item fit. Previous studies involving dichotomous items suggested that one way to detect the item fit is using comparative statistical test. Comparison method that has been widely used is the chi-square. Orlando and Thissen (as cited in Sinharay, 2003) used this method because it is more intuitive and acceptable than those using estimated abilities method because the approach deals with observed counts (Sinharay, 2003).

Chi-square statistic is also known as the goodness of fit (GOF). This statistical method describes the degree in which a set of observation fits with the conceptual model. GOF method informs the discrepancy between values observed in the data and the

Table 2
Result of Confirmatory Factor Analysis

Source of Statistic	Dimension					
	R	**I**	**A**	**S**	**E**	**C**
χ^2	34.03	35.30	16.04	41.49	34.97	39.44
df	26	26	15	29	27	32
Significance (p)	0.13436	0.10535	0.26074	0.06235	0.13960	0.17136
RMSEA	0.023	0.025	0.019	0.027	0.023	0.020
Result	Fit	Fit	Fit	Fit	Fit	Fit

values expected in the conceptual model (Olivares, 2013). Based on that suggestion, item fit in this study is analyzed using BILOG program. This program employs likelihood ratio chi-square statistical test (which will be addressed here as chi square) to test the model fit. An item is considered having fit response to the conceptual model when $\chi^2 \geq 0.01$. Appendix C shows the result of item fit for each dimension: 57 items are fit, because they have χ^2 probability > 0.01. The 14th, 22nd and 41st item are not fit, because their χ^2 probability < 0.01. The 14th item is agricultural engineer, 22nd is barber, and the 41st is manager of a company.

Items Parameters of Career Interest Instrument Based on Holland's Theory

There are two item parameters which are likely to correlate with characteristic analysis of the career in-terest instrument based on Holland's theory. These two item parameters are index of endorsement (*b*) and index of discrimination (*a*). Item parameter estimation was conducted by employing Item Response Theory Model 2-Parameter Logistic (IRT 2-PL).

As displayed in Appendix C, the analysis shows that all items have good index of endorsement and discrimination. The mean of the index of endorsement (*b*) is - 0.975, while the mean of the index of discrimination (*a*) is 1.399, both are considered good items because they are on the range determined. According to Haladyna (2004), good *b* index ranged from -3 to +3; while, good *a* index ranged from 0 to positive indefinite (+ ~) (Baker, 2001). Good index of discrimination indicates that the items function in differentiating students' with high and low interest. All the item parameter results proved that career interest instrument based on Holland's theory meet the requirement of a good instrument.

The Probability of Gender Bias in Career Interest Instrument Items

Modern DIF detection used in this study employs Item Characteristic Curve (ICC). This method identifies an item which contain DIF if the item characteristic curves on subgroup are different and thus, the curves does not co-inside. Lord (1980) argues that an item will indicate DIF if the item characteristic curves of two groups are different. Specifically, Naga (2012) adds that if item characteristic curves coincide, the item does not contain DIF. By conducting variance testing on the probability of correct answer in each ability or *P(θ)*, an item is declared containing DIF (favoring certain group) if the probability of significance $< \alpha$. Coefficient of probability of all items can be found in Appendix C. Ten items contain gender bias, because they have probability less than .05. There are item number 6, 15, 25, 27, 28, 38, 49, 51, and 58. All biased items are displayed in Table 3. Artistic dimension has the most biased items, while enterprising dimension do not have biased item. ICC graph shows which group is benefitted from the biased items (Appendix D). The advantaged group has larger probability to answer correctly than other groups on similar ability or theta.

Discussion

Item Response Theory (IRT) provides procedures for scoring tests both essays and multiple-choice items, where each response pattern is associated with some modal or expected a posteriori estimate of trait level (Thissen, Pommerich, Billeaud, & William, 1995). That argument demonstrates the importance of study about the response patterns itself. Response pattern is very important because it indicates response accuracy of each student (test taker) on question item. Consistencies of responses pattern is called i-

Table 3
Career, Gender, and Dimension of the Biased Items

Gender	Items	Career	Dimension
	6	Administration Staff	Convensional
	15	Singer	Artistic
Woman	27	Dancer	Artistic
	28	Judge	Social
	58	School Counselors	Social
	8	Pilot	Investigation
	25	Ship Captain	Realistic
Man	38	Astronout	Investigation
	49	Athlete	Realistic
	51	Sculptors	Artistic

tem fit. Item fit has the potential to indicate errors that occurred in the calibration phase of test development. For instance, if a program consistently underestimates the discrimination parameter for highly discriminating items, an item-fit statistic should identify this problem (Reise, 1990). Item response pattern is considered good (fit) if its probability value $\chi^2 \geq .01$. Based on that value, three items from 60 item response pattern were proven unfit with the model. Unfit items indicate that the items have questionable validity, because they do not accurately represent how participants respond to test items (Reise).

Unlike the ideal model, these items can also indicate answer inexpediency (inconsistent thinking) based on their ability. Deviations from the ideal model reflect patterns of responses which are unexpected given a person's level of ability (Sumintono & Widhiarso, 2015). Furthermore, Meijer (1996) explains that there are eight causes for improper or bad responses. They are:

1. cheating, that is, when the test questions are difficult, but test takers can answer many questions correctly;
2. lucky guess, that is, when test takers unpredictably can give correct response on difficult items;
3. test takers' confusion or anxiousness;
4. slowness, that is, when test takers' never completed all items;
5. language, that is, when test takers have low ability in understanding questions or constructions;
6. random prediction, that is, when test takers respond to items without thinking;
7. over-creativity, that is, when test takers interpret items unusually or creatively; and
8. carelessness, that is, when test takers are less thorough in giving response on answer sheet.

Ferrando (2015) adds that the response patterns that does not fit is most likely caused by lack of interest among the respondents (do not have motivation), so they exhibit random and inconsistent responses pattern.

Among the 60 items analyzed, there are ten items or careers which positively contain DIF. The career orientation as a administration staff, singer, judge, and school counselor were more attractive for female students than male students. In contrast, the career as a pilot, ship captain, astronaut, athlete, and sculptors were more attractive for male students than female students. These differences were caused by a contextual factor, namely, the gender factor. This factor can prevent an individual to achieve their career goals.

Previous studies have shown the relationship between gender and career choices. Gottfredson (1996) found that gender stereotype influences children in deciding their career choice because they adjust their career choice to their gendered self-image. Children, age 6 to 8, are able to give negative responses to certain careers such as heavy equipment operator, which is frequently rejected by female students; and secretary or nurse, which is frequently rejected by male students. Gottfredson suggested parents, teachers, and the community to guide their children, otherwise, children's early self-rejection of career choices cannot be altered again, and it becomes permanent self-rejection.

In another study which examined the relationship between gender and career choice, Miller and Stanford (1987) observe students of grade one to grade five. The results showed that students generally have career selection which confirms gender stereotypes, with male students have more career choices than female students. Park et al. (as cited in Ceci & Williams, 2010) adds that among males and females of comparably

outstanding mathematical aptitude, females are more likely to also have outstanding verbal ability. This gives them more career choices than males who see their strength only in math, because females can consider math-oriented fields as well as law, social sciences, humanities, and medicine.

Traditionally, career orientations can be classified into person-oriented or thing-oriented (P/Person-T/Thing) careers. Careers that deal primarily with the creation and manipulation of human-made artifacts, such as engineering and mechanics, were considered thing-oriented careers. These are usually technology-focused fields that are considered masculine because technology has traditionally and culturally been constructed as masculine. Careers centered on interpersonal interactions are considered person-oriented careers and generally taken up by women (Ngambeki et al., 2012). Within this classification, female children tend to be interested in careers such as nurse, teacher, or social worker; while, male children tend to be interested in careers involving engineering, science, or technology (Lippa, 1998).

Correspondingly, the meta-analysis study by Su, Rounds, and Armstrong's (2000) explored reasons for the little representation of women in the STEM (science, technology, engineering and maths). The results proved that men prefer working with things and women prefer working with people. Further, men were more interested in realistic and investigation careers, and women were more interested in artistic and social careers.

Sex differences favoring men were also found in more specific measures of career orientation related to engineering, science, and mathematics. Another study showed that male children in grade three and six tend to be interested in thing-oriented careers while fe-male children in similar grades tend to be interested in person-oriented careers (Graziano et al., 2012). That study also indicates that female children show greater flexibility in selecting person-oriented careers.

The results of the current study indicate that there is gender bias in students' career selection. Consequently, a gender might be less capable or less skilled in careers that have been dominated by the opposite gender. This gender bias is caused by differing perceptions about the jobs by male and female students. These biased perceptions were influenced by parental, community, and cultural factors. The implication of the current study for educators is that they need to acknowledge and consider gender biases in career selection when discussing careers with students.

For test developers, there are several implications of the current study. In term of scoring, if the scoring analysis used classical theory, where the index of endorsement or difficulty was ignored, the interpretations of the instrument usually use category or cutting point. The test developer can make different norm categories for men and women. For researchers who are interested to develop career interest scale, items in the instrument might be constructed differrently for men and women.

If the administration of test used items response theory, the estimation of level score theta (ability or trait) may need different estimations for men and women. The scoring of the theta is usually general, not separated based on gender.

Limitations and Suggestions

There are several limitations in this study. Firstly, among many career options, only ten careers were chosen to represent each of the RIASEC dimensions as referred to Holland's Theory. Therefore, the recommendation for further studies is to provide more career options in order to obtain more comprehensive results of career interest. Secondly, the use of ICC method as a method of bias detection has some advantages and disadvantages. It is possible that using another bias detection method may generate different results. Thirdly, this study used gender variables for reference groups and focus groups as the basis for the analysis of the item bias. Future study is suggested to use other variables such as age, race or ethnicity, culture, or language groups. In addition, further study might investigate career response patterns among older participants, such as adolescents or adults.

Conclusion

Based on the findings and discussions above, it can be concluded that most items in this instrument have good item parameter so that it can be used for practical use as well as for further studies. However, items that do not fit the model need to be amended before they can be used for further studies as well as for practical use. In term of gender, ten items were identified as containing Differential Item Functioning (DIF) as shown by the Item Characteristic Curve (ICC). This indicates that there is a relationship between career interest and gender, in which female children tend to be interested in person-oriented careers, such as teachers, nurses, or social workers;

while male children tend to be interested in thing-oriented careers, such as science and technology.

References

Adedoyin, O. O. (2010). Using IRT approach to detect gender biased items in public examinations: a case study from the Botswana junior certificate examination in Mathematics. *Educational Research and Reviews, 5*(7), 385-399. Retrieved from http://academic journals.org/ERR2

Athanasou, J. A. (2001). Analysis of response to vocational interest items: a study of australian high school students. *Journal of Career Assessment, 9*(1), 61-79. http://dx.doi.org/10.1177/106907270 100900105

Arthur, M. B., Hall, D. T., & Lawrence, B. S. (1989). Generating new directions in career theory: The case for a transdisciplinary approach. In M. B Arthur, D. T. Hall, & B. S. Lawrence (Eds.), *Handbook of career theory* (pp. 7-25). Cambridge, UK: Cambridge University Press.

Ayriza, Y., Setiawati. F. A., & Triyanto, A. (2016). *Career interest and knowledge of lower grade students of primary school.* Paper presented at the International Conference of Computer, Environment, Social Science, Engineering and Technology (ICEST 2016). Medan, Indonesia, May 23-25th.

Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). Wisconsin: ERIC Clearinghouse on Assessment and Evaluation. Retrieved from http://echo.edres.org/8080.

Baker, F. B., & Seock, H. K. (2004). *Item response theory parameter estimation techniques.* New York: Marcel Dekker.

Betz, N. E., & Turner, B. M. (2011). Using item response theory and adaptive testing in online career assessment. *Journal of Career Assessment, 19*(3), 274-286. http://dx.doi.org/10.1177/10690727103 95534

Budiono. (2009). The accuracy of mantel-haenszel, sibtest, and regression methods in differential item function detection. *Jurnal Penelitian dan Evaluasi Pendidikan, 12*(1), 1-20.

Camili, G., & Shepard, L. A. (1994). *Methods for identifying bias test items.* Thousand Oaks, CA: Sage Publication.

Ceci, S. J., & Williams, W. M. (2010). Sex differences in math-intensive fields. *Current Directions in Psychological Science, 19*(5), 275-279.

Demars, C. (2010). *Item response theory, understanding statistics measurement.* New York: Oxford University Press.

Department of Education and Science. (1989). The boundaryless career: A competency-based perspective. *Journal of Organizational Behavior, 15,* 307-324.

Du Toit, M. (2003). *IRT from ssi:bilog-mg, multilog, parscale, testfact.* Lincolnwood: Scientific Software International.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychology.* Mahwah, NJ: Lawrence Elbaum Associates Publisher.

Ferrando, P. J. (2015). Assessing person fit in typical response measures. In S. P. Reise, & Revicki (Eds.), *Handbook of item response theory modelling: application to typical performance assessment* (pp. 128-155). New York: Routledge, Taylor & Francis Group.

Gies, V. (1990). Developing a personal career counseling theory: An overview of the theories of Donald Super and David Tiedman. *Guidance & Counseling, 6*(1), 1-5.

Gottfredson, L. S. (1996). A theory of circumspection and compromise. In D. Brown, & L. Brooks (Eds.), *Career, choice and development: applying contemporary theories to practice* (3rd ed., pp. 179-232). San Francisco, CA: Jossey-Bass Publishers.

Graziano, W. G., Habashi, M. M., Evangelou, D., & Ngambeki (2012). Orientations and motivations: Are you a "people person," a "thing person," or both? *Motivation and Emotion, 36,* 465-477. http://dx.doi.org/10.1007/s11031-011-9273-2

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and application.* Boston: Kluwer Academic Publishers.

Hambleton, R. K., & Zaal, J. N. (1991). *Advance in educational and psychological testing.* Boston: Kluwer Academic Publishers.

Hartung, P. J., Porfeli, E. J., & Vondracek, F. W. (2008). Career adaptability in childhood. *The Career Development Quarterly, 57*(1), 63-74. http://dx.doi.org/10.1002/j.2161-0045.2008.tb00166.x

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL: Psychological Assessment Resources.

Lippa, R. (1998). Gender-related individual differences and the structure of vocational interests: The importance of the people-things dimension. *Jour-*

nal of Personality and social Psychology, 74(4). 996-1009. http://dx.doi.org/10.1037/0022-3514.74.4.996

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

McDaniels, C., & Gysbers, N. C. (1992). Counseling for career development theories, resources, and practice. San Francisco, CA: Jossey-Bass Publishers.

Meijer, R. R. (1996). Person-fit research: An introduction. Applied Measurement in Education, 9(1), 3-8.

Miller, M. J., & Stanford, J. T. (1987). Early occupational restriction: An examination of elementary school children's expression of vocational preferences. Journal of Employment Counseling, 24, 115-121.

Mislevy, R. J,. & Bock, R. D. (1990). Bilog 3: Item analysis and test scoring with binary logistic models (2nd Ed.). Mooresville: Scientific Software Inc.

Naga, D. S. (2012). Teori sekor pada pengukuran mental (Edisi kedua). Jakarta: Nagarani Citarayasa.

Ngambeki, I., Habashi, M. M., Evangelou, D., Graziano., W. G., Sakka, D., & Corapci, F. (2012). Using profiles of person-thing orientation to examine the underrepesentation of women in engineering in three cultural context. International Journal of Engineering Education, 28(3), 621-632.

Olivares, A. M. (2013). Why should we assess the goodness–of fit of irt models? Measurement, 11, 127-137. http://dx.doi.org/10.1080/15366367.2013.841511

Oliveira, I. M., Taveira, M. C., Cadime, I., & Porfeli, E. J. (2016). Psychometric properties of a career exploratory outcome expectations measure. Journal of Career Assessment, 24(2), 380-396. http://dx.doi.org/10.1177/1069072715580577

Osterlind, S. J. (1983). Test item bias. Beverly Hills: CA: Sage Publication Inc.

Rahayu, W. (2010). Metode linking dan butir false positive pada pendeteksian DIF berdasarkan teori respons butir. Jurnal Penelitian dan Evaluasi Pendidikan, 14(1), 20-36.

Reardon, R. C., & Lenz, J., G. (1999). Holland's Theory and Career Assessment. Journal of Vocational Behavior, 55, 102-113.

Reise, S. P. (1990). A comparison of item and person fit methods of assessing model data fit in irt. Applied Psychological Measurement, 14(2), 127-137.

Retnawati, H. (2013). Pendeteksian keberfungsian butir pembeda dengan indeks volume sederhana berdasarkan teori respons butir multidimensi. Jurnal

Penelitian dan Evaluasi Pendidikan, 17(2), 275-286.

Savickas, M. L. (1999). The psychology of interests. In M. L. Savickas, & A. R. Spokane, (Eds.), Vocational interest: Meaning, measurement, and counseling use (pp. 19-56). Palo Alto, CA: Davies-Black.

Sinharay, S. (2003). Bayessian item fit analysis for dichotomous item response theory models (Research Report No. 03-34). Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/Media/Research/pdf/RR-03-34.pdf

Su, R., Rounds, J., & Armstrong, P. I. (2000). Men and things, women and people: A meta-analysis of sex difference in interests. Psychological Bulletin, 135(6), 859-884. http://dx.doi.org/10.1037/a0017364

Sudaryono (2012). Kajian metode deteksi differential item functioning (dif) butir soal ujian nasional dengan teori tes klasik. Jurnal Pendidikan dan Kebudayaan, 18(2), 132-144.

Super, D. E., Savickas, M. L., & Super, C. M. (1996). The life-span, life-space approach to careers. In D. Brown & L. Brooks (Eds.), Career, choice and development (3rd Ed.). San Fancisco: Jossey-Bass Publishers.

Sumintono, B., & Widhiarso, W. (2015). Aplikasi permodelan rasch pada assessment pendidikan. Cimahi: Trim Komunikata.

Sverko, I., & Babarovic, T. (2006). The validity of Holland's theory in croatia. Journal of Career Assessment, 14(4), 190-507. http://dx.doi.org/10.1177/1069072706288940

Thissen, D., Pommerich, M., Billeaud, K., & Williams, V. S. L. (1995). Item response theory for scores on tests including polytomous items with ordered responses. Applied Psychological Measurement, 19(1), 39-49. http://dx.doi.org/10.1177/014662169501900105

Thorpe, G. L., & Favia, A. (2012). Data analysis using item response theory methodology: An introduction to selected programs and applications (Paper No.20). Orono, Maine: The University of Maine. Retrieved from http://digitalcommons.library.umaine.edu/cgi/viewcontent.cgi?article=1019&context=psy_facpub

Turner, B. M., Betz, N. E., Edwards, M. C., & Borgen, F. H. (2010). Psychometric examination of an inventory of self-efficacy for the Holland Vocational Themes using item response theory. Measurement and Evaluation in Counseling and Development, 43(3), 188-198. http://dx.doi.org/10.1177/0748175610384810

Wetzel, E., Hell, B., & Passler K. (2012). Comparison of different test construction strategies in the development of a gender fair interest inventory using verbs. *Journal of Career Assessment, 20*(1), 88-104. http://dx.doi.org/10.1177/1069072711417166

(Appendix follows)

Appendix A

Instrument Specifications of Career Interest Based on Holland's Theory

Career Orientation	Number of Item	Career
Realistic	1	Tailor
	7	Driver
	13	Breeder
	19	Farmer
	25	Ship Captain
	31	Cobbler
	37	Painter
	43	Chef
	49	Athlete
	55	Mechanic
Investigation	2	Surgeon
	8	Pilot
	14	Agricultural Engineer
	20	Television Mechanic
	26	Doctor
	32	Civil Engineer
	38	Astronaut
	44	Laboratory Staff
	50	Dentist
	56	Veterinarian
Artistic	3	Decorator
	9	Musician
	15	Singer
	21	Painter
	27	Dancer
	33	Model
	39	Actor
	45	Author
	51	Sculptor
	57	Cartoon Illustrator
Social	4	Head Master
	10	Nurse
	16	Hair Dresser
	22	Barber
	28	Judge
	34	Teacher
	40	Community Leader

(Appendix continues)

Career Orientation	Number of Item	Career
	46	Scout Teacher
	52	Firefighters
	58	School Counselor
Enterprising	5	Shopkeeper
	11	Traders
	17	Peddler
	23	Master of Ceremony
	29	Police
	35	Tour guide
	41	Manager in a company
	47	Member of the Parliament
	53	Salesperson
	59	Lawyer
Conventional	6	Administration Staff
	12	Bank employee
	18	Office Administration
	24	Journalist
	30	Secretary
	36	Finance Officer
	42	Treasurer
	48	Librarian
	54	Cashier
	60	Tax Officer

(Appendix continues)

Appendix B

The Conceptual Diagram and the Result Of Confirmatory Factor Analysis The Dimensions of Instrument of Career Interest

Dimension of Realistic
Standardized Solution

Chi-Square=34.03, df=26, P-value=0.13436, RMSEA=0.023
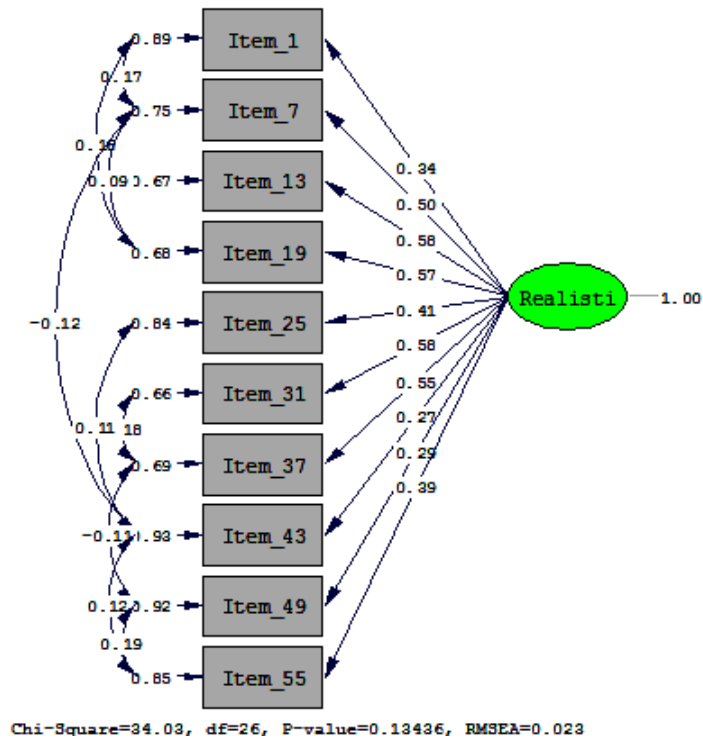
(Appendix continues)

Dimension of Investigative
Standardized Solution
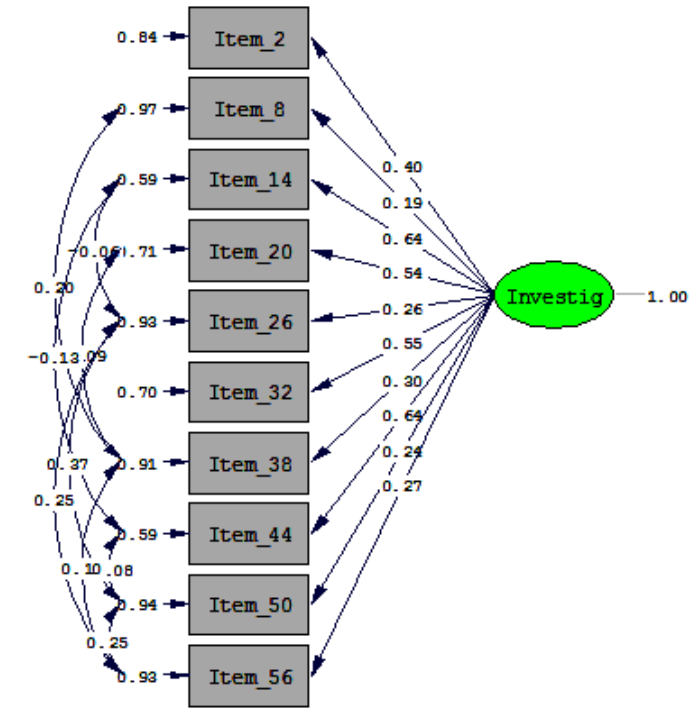
Chi-Square=35.30, df=26, P-value=0.10535, RMSEA=0.025

Dimension of Artistic
Standardized Solution
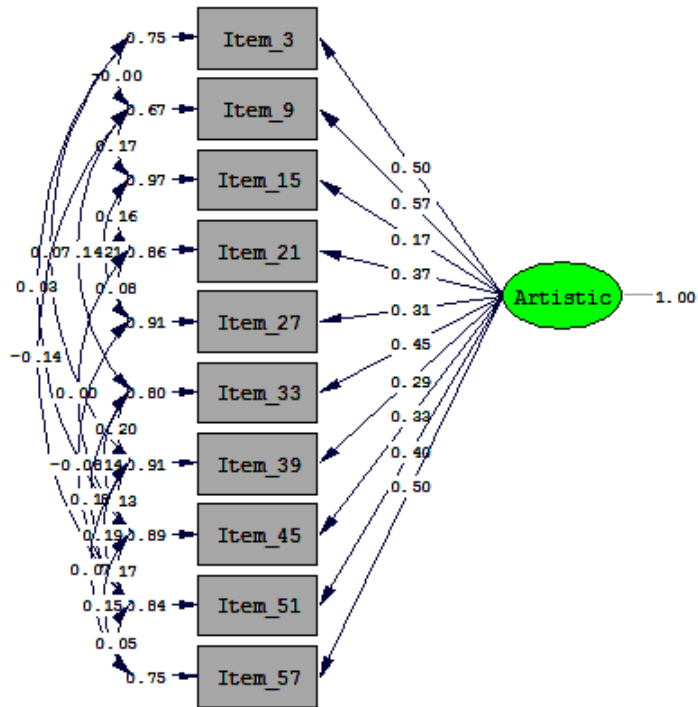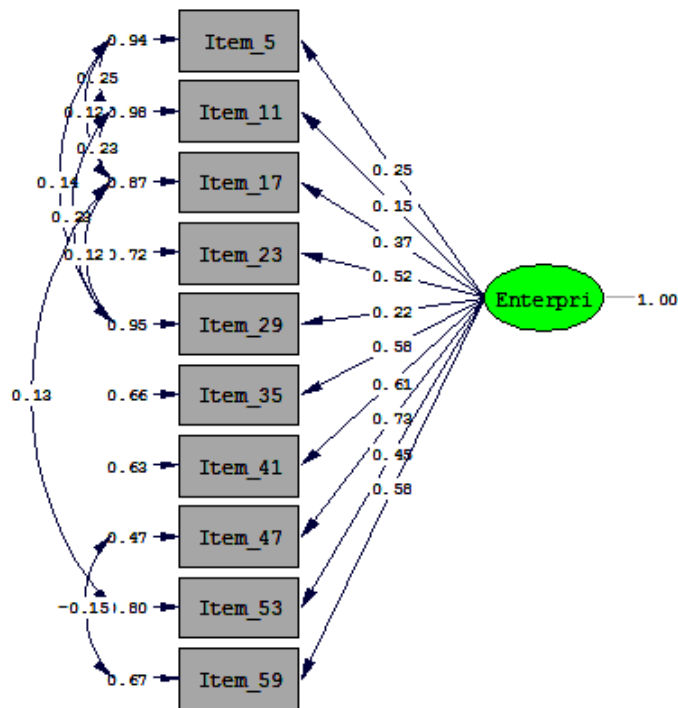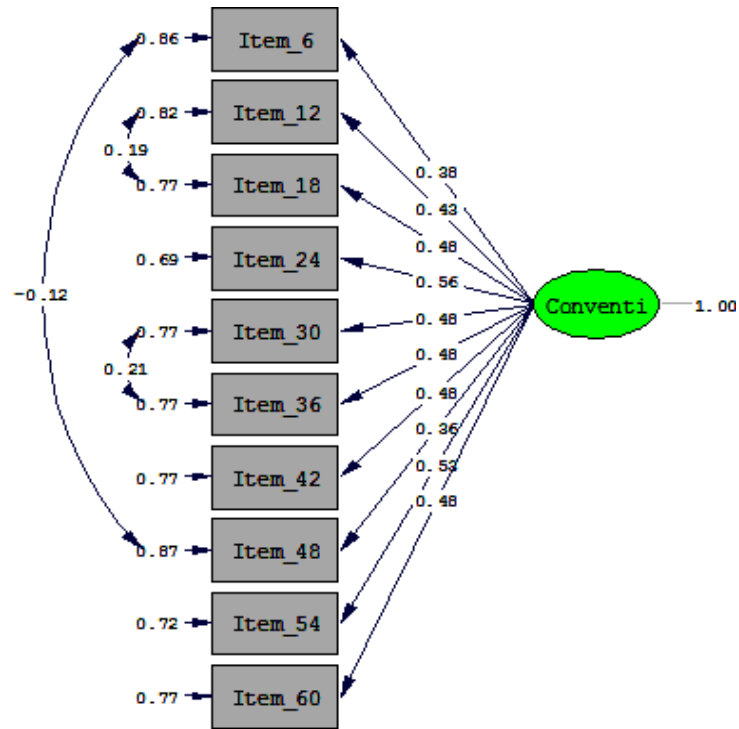

Chi-Square=18.04, df=15, P-value=0.26074, RMSEA=0.019

(Appendix continues)

Dimension of Social
Standardized Solution
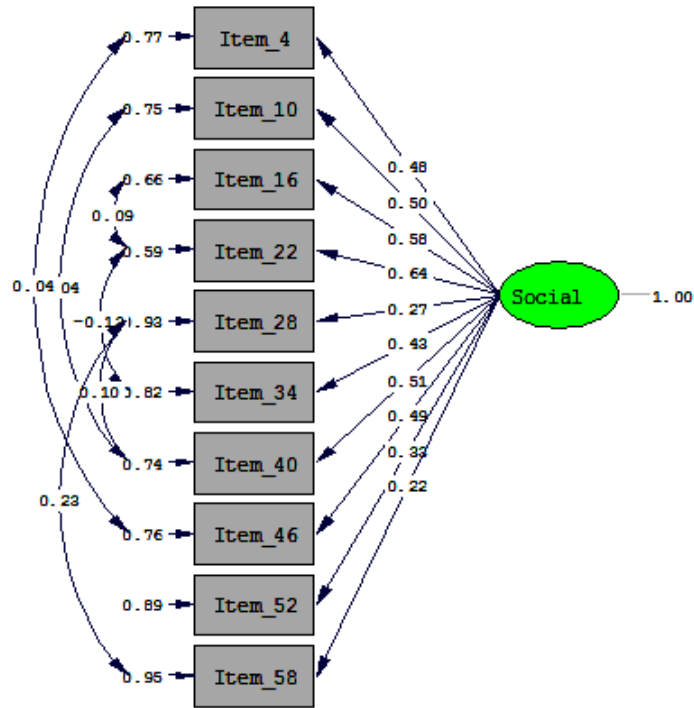
Chi-Square=41.49, df=29, P-value=0.06235, RMSEA=0.027

Dimension of Enterprising
Standardized Solution

Chi-Square=34.97, df=27, P-value=0.13960, RMSEA=0.023

(Appendix continues)

Dimension of Conventional
Standardized Solution

Chi-Square=39.44, df=32, P-value=0.17136, RMSEA=0.020

(Appendix continues)

Appendix C

Fit Model of Response Patterns of Items

Career Orientation	Items	*p-value* of Chi-Square	Index of Discrimination *(a)*	Index of Endorsement *(b)*	*p-value* of ICC Differentiation
Realistic	1	0.8457	1.155	-1.685	0.052
	7	0.6743	1.546	-1.245	0.432
	13	0.8793	1.436	-0.441	0.462
	19	0.8930	1.886	-1.037	0.939
	25	0.4994	1.064	-1.092	0.016**
	31	0.9047	1.901	-0.372	0.307
	37	0.7320	1.667	-0.661	0.635
	43	0.7957	0.791	-2.005	0.492
	49	0.3254	0.662	-0.457	0.011**
	55	0.9612	1.068	0.428	0.187
Investigation	2	0.4592	0.992	0.760	0.588
	8	0.3039	0.763	-2.374	0.000**
	14	0.0087*	1.611	0.657	0.140
	20	0.0496	1.360	0.396	0.263
	26	0.3408	1.642	-1.803	0.150
	32	0.7134	1.476	0.426	0.132
	38	0.1879	0.888	-1.584	0.007**
	44	0.3267	1.384	0.414	0.186
	50	0.8422	1.102	-1.567	0.150
	56	0.8276	1.149	-1.145	0.975
Artistic	3	0.2282	1.137	0.208	0.757
	9	0.3559	1.331	-1.235	0.645
	15	0.1506	1.229	-1.889	0.003**
	21	0.3645	1.370	-1.689	0.580
	27	0.7320	0.851	-1.461	0.000**
	33	0.5507	1.480	0.729	0.534
	39	0.7424	1.271	0.176	0.446
	45	0.1041	1.415	-0.552	0.288
	51	0.7950	1.349	0.796	0.000**
	57	0.4658	1.363	-0.758	0.212
Social	4	0.5896	1.745	-1.988	0.914
	10	0.0400	1.423	-0.932	0.592
	16	0.5915	2.095	-0.939	0.998
	22	0.0000*	2.302	-0.779	0.273
	28	0.3295	0.945	0.700	0.001**
	34	0.7354	1.351	-2.510	0.344

(Appendix continues)

40	0.5660	1.620	-1.242	0.269
46	0.0620	1.408	-1.392	0.661
52	0.2707	1.028	-2.097	0.064
58	0.0370	1.155	-1.685	0.004**
5	0.3318	1.546	-1.245	0.452
11	0.7388	1.436	-0.441	0.379
17	0.8688	1.886	-1.037	0.320
23	0.9899	1.064	-1.092	0.885
29	0.7853	1.901	-0.372	0.232
Enterprising 35	0.3963	1.667	-0.661	0.378
41	0.0068*	0.791	-2.005	0.456
47	0.1584	0.662	-0.457	0.439
53	0.0759	1.068	0.428	0.670
59	0.7529	0.992	0.760	0.878
6	0.8737	0.763	-2.374	0.000**
12	0.8389	1.611	0.657	0.902
18	0.5655	1.360	0.396	0.981
24	0.2110	1.642	-1.803	0.410
30	0.8457	1.476	0.426	0.946
Conventional 36	0.3869	0.888	-1.584	0.378
42	0.6151	1.384	0.414	0.808
48	0.4543	1.102	-1.567	0.778
54	0.6439	1.149	-1.145	0.375
60	0.9862	1.137	0.208	0.275

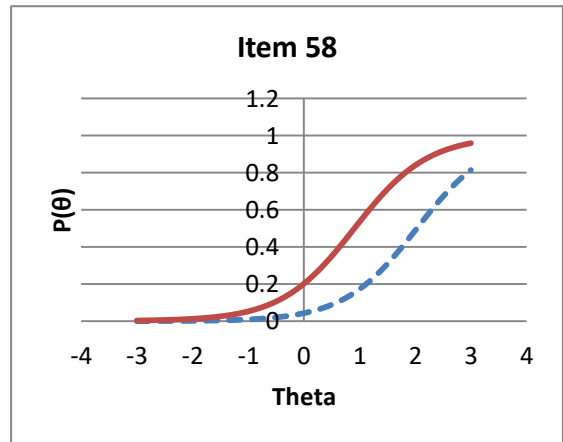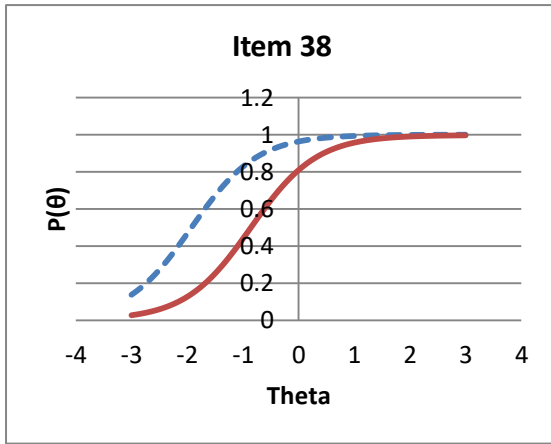*Note.* * Unfit model; ** Biased item

(Appendix continues)

Appendix D

ICC Graph of Items Containing DIF

Item 6

Item 8

Item 9

Item 25

Item 27

Item 28
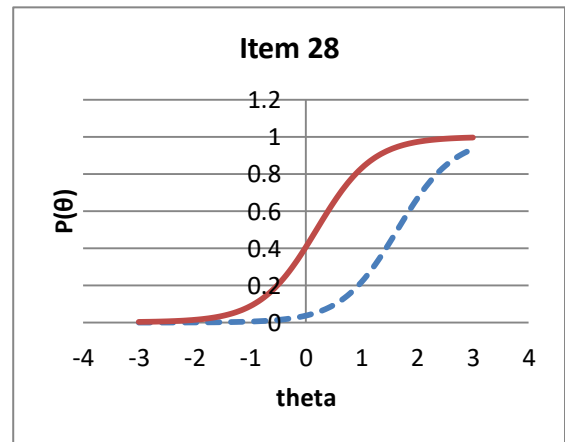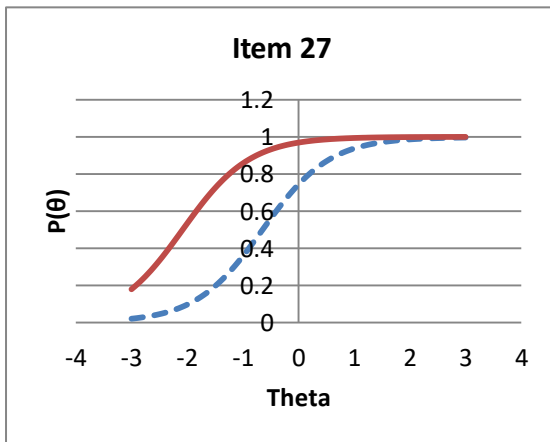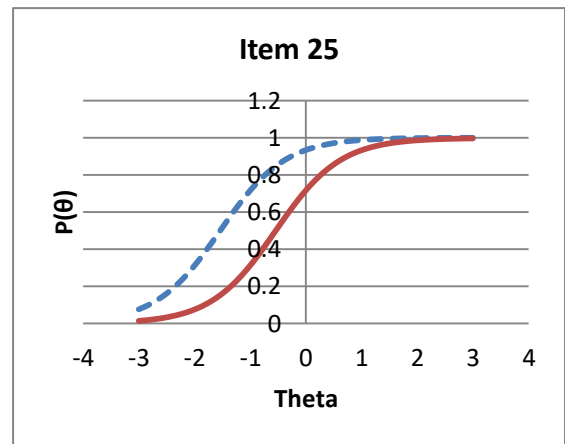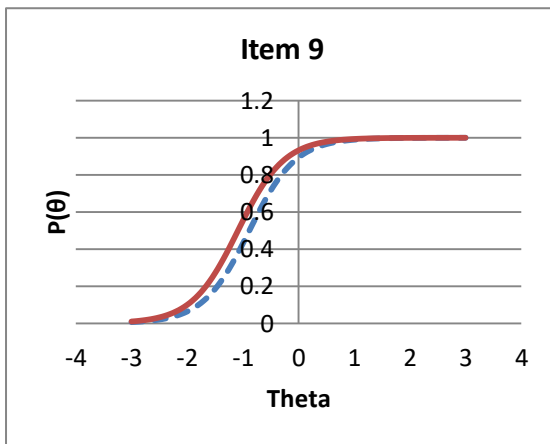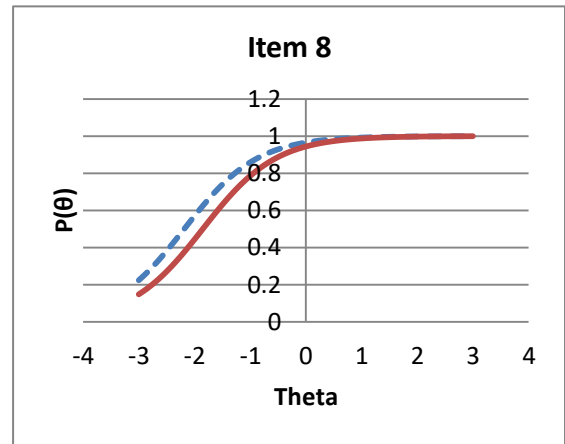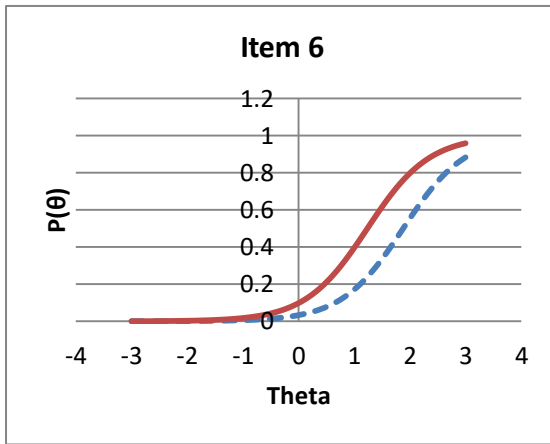
(Appendix continues)

Note. _____ = Women; - - - - - = Men