# Detecting Gender Biasness via Gender Differential Item Functioning Analysis on Integrated Meaningful Hybrid E-learning Instrument

ROSSENI DIN*[1], M FAISAL KAMARULZAMAN[2], HELMI NORMAN[1], PARILAH M SHAH[1]
AIDAH A KARIM[1], NUR AYU JOHAR[1] & VERAWATI[1]

Faculty of Education[1]
Universiti Kebangsaan Malaysia
43600 Bangi, Selangor
Malaysia

Faculty of Economy and Business[2]
Universiti Kebangsaan Malaysia
43600 Bangi, Selangor
MALAYSIA

*rosseni@ukm.my, rosseni@yahoo.com

*Abstract:* - Facilitation of meaningful learning in e-learning curriculum and environments at higher education level is essential to ensure students incorporate concepts and information delivered to them. In ensuring that meaningful learning is disseminated among students, a learner's level of meaningful e-learning can be measured using a valid and reliable measuring instrument. For this reason, we have developed a measuring instrument to measure meaningful e-learning for students in Malaysian higher educational institutions called the Integrated Meaningful e-Training instrument. The focus of the paper is aimed to determine whether there are differences in meaningful e-learning scores between male and female students in Malaysia as well as identify whether there is gender bias in the items of the meaningful e-learning instrument. This method of study is survey design. SPSS and WINSTEP software were used for data analysis. The findings showed that there was only one item that still has gender differential item functioning and one item that needs attention, which are item 12 and 22 from section B of the instrument. Thus, this instrument is fair to measure meaningful learning either for male or female students.

*Key-Words:* - GDIF, Meaningful Learning, E-Learning, Meaningful e-Training, Hybrid Learning

## 1 Introduction

Meaningful learning has been the aim for any teaching and learning practice. Meaningful in this study means any training delivered with active, constructive, collaborative, authentic and intentional learning strategy via conventional or alternative method (Fig. 1). Fig. 2 shows the original meaningful learning attributes. Various conventional methods such as cooperative learning, experiential learning, problem-based learning, project-based learning and problem-oriented project-based learning can be employed to attain meaningful learning. However, as time becomes an issue, most trainers resort to lecture-based training. When training is restraint to predominantly lecture method, meaningful learning may not be the main intention of training any longer. This is essentially more pertinent for learners with auditory learning style preference only. According to various studies such as [1, 2, 3, 4] and many others, learners with auditory preferences constitute only one third of the population or less. As such, an alternative method (Fig. 3) is needed to accommodate other learners with differentiated learning style preferences in order for them to experience a more meaningful learning experience.

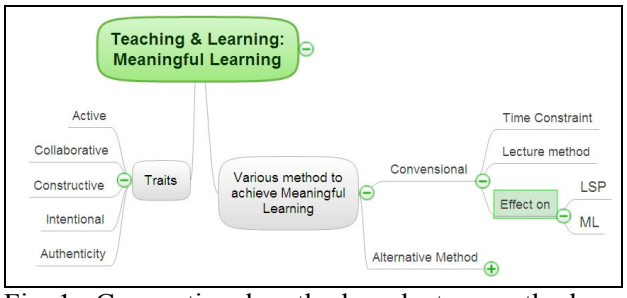


Fig. 1: Conventional methods vs lecture method

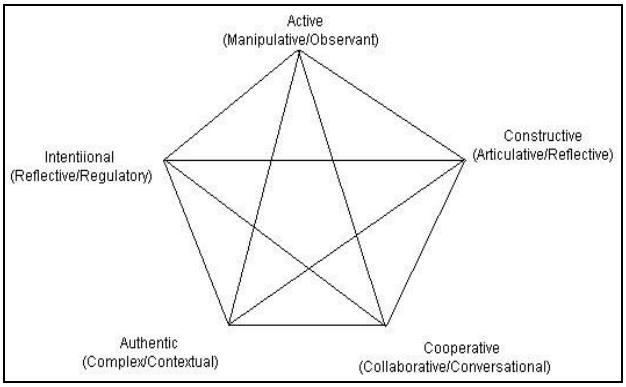

Fig. 2: Five attributes of meaningful learning [5]

## 2 Problem Statement

Existing e-learning models, framework or guideline for hybrid method include the criteria needed for a superior hybrid system such as a superior structure, superior delivery, superior content, superior service and superior outcome [5, 6, 7, 8, 9, 10, 11, 12]. However, most of the hybrid systems are still limited to acting solely as online repositories. This scenario and the system's lack of personalization to cater learners with diverse learning style preferences to achieve meaningful learning has become the main research problem for this study. However, for the purpose of this paper, we focused on testing whether the instrument used to evaluate meaningful e-learning (MeT) is suitable, reliable and valid to be used on both males and female students. MeT is a section in the integrated meaningful e-training (I-MeT) instrument which consists of measures for hybrid learning and learning style to evaluate meaningful hybrid e-learning delivered using a knowledge management system.

With the advent of knowledge-economy, embracing the concept of knowledge management (KM) for lifelong learning (LLL) as the foundation of a learning society, takes priority. This is because people will have to continuously update their knowledge and skills to maintain a competitive edge in the global economy [13]. The Malaysian Qualification Framework (MQF) provides the structure for actualizing LLL because it facilitates learners in selecting a learning pathway that is most appropriate for them [13, 14]. Thus, a response was made to create an academic culture capable of producing learners with qualities ranging from competencies in soft skills, intellectual qualities and affective attributes, in addition to the typical technical and professional skills [15].

To successfully create the much desired academic culture, the Committee of Deputy Vice Chancellors and Rectors of Malaysian Higher Learning Institutes [15] had drawn up four strategies: (i) having competent and professional academicians, (ii) providing conducive facilities, (iii) implementing an updated, relevant curriculum with various delivery methods, and (iv) making initiatives to improve and monitor key performance indicators. No framework or model have yet been provided to implement the third strategy although some work have been done to materialize the first through fourth strategies by the Centre for Academic Advancement, Universiti Kebangsaan Malaysia (UKM) and other centres for professional development of various institutes of higher learning in Malaysia. The second strategy has been continuously implemented, maintained and upgraded by the university, wherever and whenever needed. As for the third strategy, all academicians involved will have to do their part as a means to achieve the shared vision of the university; that is to create an academic culture comparable to international standards at the same time, able to nurture a holistic development of the learner.

It is widely accepted that ICT infrastructure enables e-Training. The technology may save university administrators costs and add a measure of convenience for learners, but educators may reason that if e-training programs do not produce workers who are capable of higher order thinking and reasoning to solve intricate and authentic problems in the workplace, then the programs are not worth much [6]. In the strategic planning process to implement a new e-training program or enhance existing ones, the focus should therefore not be primarily on how technology can be used to achieve educational goals, but also on the human aspects of teaching and learning.

Various studies have been conducted in relation to: (i) e-learning, (ii) e-training, and (iii) meaningful learning and learning styles [6, 16, 17, 18, 19, 20, 21, 22, 23, 24]. All too often, though, researchers are faced with questions inter-relating these three variables. How does learning style preference affect meaningful learning? Does blending conventional learning with technology facilitate one to achieve meaningful learning? These series of issues have both practical and theoretical importance. Yet, none of the conventional multivariate techniques such as multiple regression enable us to address all these questions with one comprehensive technique. The overall research examines the technique of structural equation modelling (SEM), an extension of several multivariate techniques, most predominantly factor analysis and multiple regression analysis. This technique will enable the researcher to assess both measurement properties and test the key theoretical relationships in one technique. Before doing SEM however, item analysis and test against biasness of the instrument need to be done. This paper discusses about the gender bias aspects of the instrument.

With a valid, unreliable and non-gender biased instrument, the overall study focused on developing a model for meaningful e-training using the hybrid method to cater to learners with differentiated learning style preferences, especially those with kinaesthetic, tactual and group preferences. This is due to the fact that this group of learners has been receiving less focus in view of the fact that the design of most instructional media is inclined to cater to learners with visual, audio and individual preferences. Many literatures supported the fact that many instructional media supported learning for learners with visual, audio and individual preferences and for those with different levels of ICT ability [1, 2, 3, 4, 21, 22, 23, 24]. However, not many instruments go back to investigate if these media produces meaningful e-learning.

The data analysis techniques in this paper only focuses on the results of the instrument developed in the study to evaluate hybrid e-learning as to whether or not if there exist any gender biasness in terms of meaningful learning. The presentation of findings will first observe the aspects of the mean and standard deviation. Further analysis techniques using Rasch Model was subsequently performed to identify the gender differential item functioning (GDIF) in the instrument. This step is important to improve the quality of the items in the next version of the instrument so that it will be able to avoid gender biasness.

## 3 Research Methodology

Various attempts have been carried out to improve

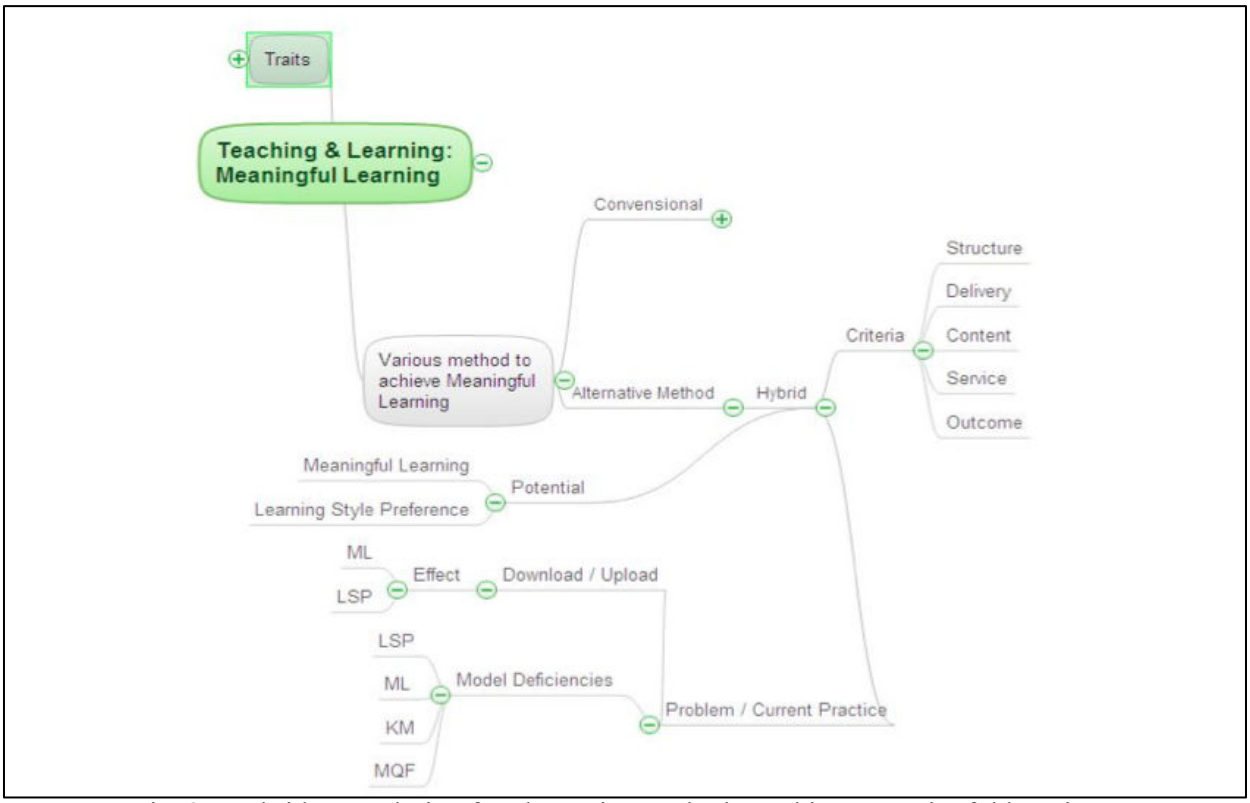


Fig. 3: Hybrid as a solution for alternative method to achieve meaningful learning

the quality of the Meaningful Hybrid E-Training (MeT) instrument. Numerous analysis using a variety of techniques and methods based on either Classical Test Theory (CTT), Rasch Measurement Theory or Item Response Theory (IRT) have been done. The fundamental difference of these theories lies in the aspect of which is to be measured. IRT and Rasch measurement does not focus only on the person but it also includes the item. Both theories are classified as modern test theory. These theories, modern and classic test theories are equally important in the measurement world because if we aim to come out with a highly reliable and valid instrument, we will have to start examining the items comprehensively from both angles. This study used two viewpoints to answer the research questions which are :

(1) *Do gender differences exist in the MeT instrument used in Malaysia*? Classical Test Theory (CTT) approach in this study will attempt to answer the question of whether there is any significant difference in mean score between male and female in terms of meaningful learning after undergoing hybrid e-training.

(2) *Which items are still bias against male or female*? Subsequently, using a similar approach to Item Response Theory (IRT) namely the Rasch Model, the researchers will attempt to answer the second question.

## 4 Results and Discussion

This study aims to identify whether there are significant differences in mean scores between male and female. In addition, this study also aims to identify whether there is gender bias (GDIF) in the MeT instrument. Fig. 4 - Fig. 8 shows the demographic profile of respondents who answered the MeT test.

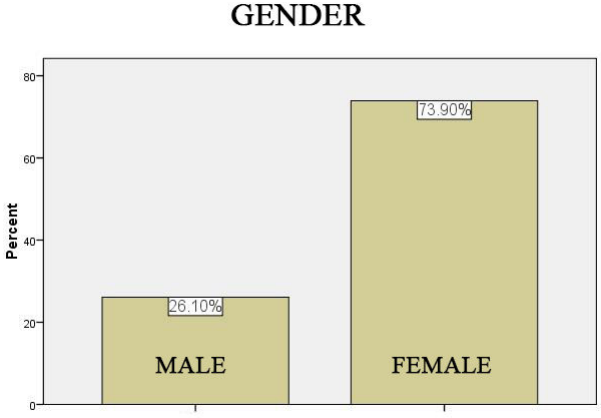


Fig. 4: Respondent Profile: Gender

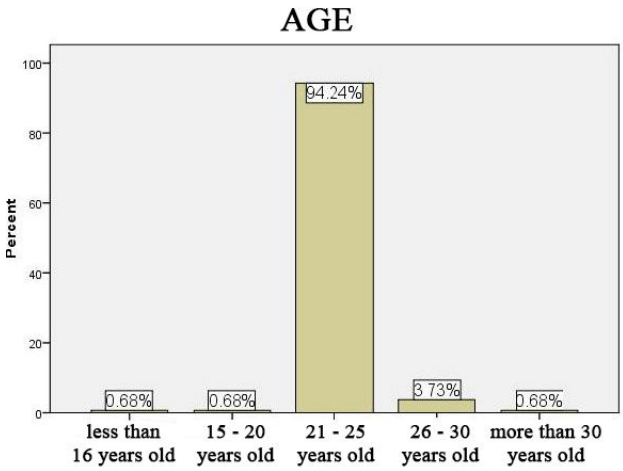


Fig. 5: Respondent Profile: Age

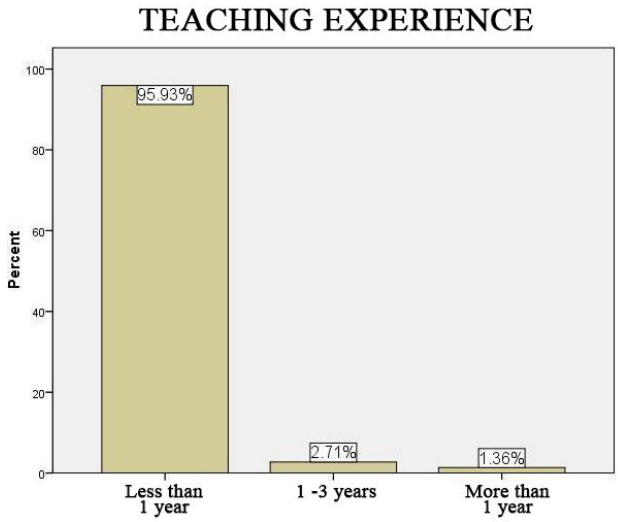


Fig. 6: Respondent Profile: Experience

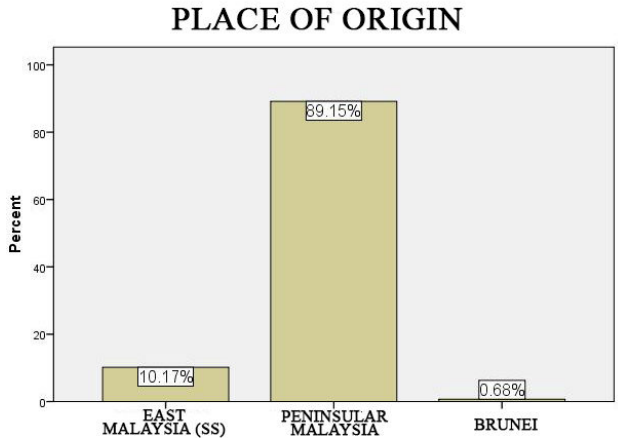


Fig. 7: Respondent Profile: Study Program

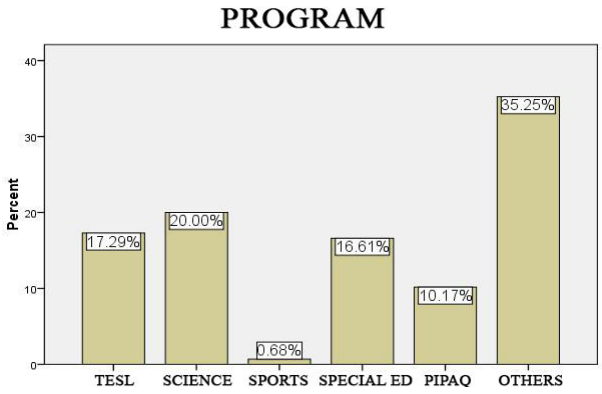


Fig. 8: Respondent Profile: Program of Study

An independent-sample t-test was conducted to compare the perceived achievement of meaningful learning between male and female students. As referred to Table 1, the mean scores of male (M=46.68, SD=5.11) and female (M=46.96, SD=4.82) are almost the same. Next, we refer to the Levene's *F*-test for equality of variances, which equals 0.14 and is statistically significant at the 0.905 level. This indicates that two samples randomly drawn from populations with similar variances would generate an *F*-test with a value of 0.14 for 905 times out of 1000 trials. Therefore, the null hypothesis that assumes the variances of the two populations from which the samples were drawn are equal, and the *t*-test of assuming equal variances was accepted; t (293)=.-.444, p = 0.657 as shown in Table 2. These results suggest that there is no significant difference between the two groups (will show significance if p is less than .05). Specifically, our results suggest that gender have the same perception about perceived achievement of meaningful learning throughout their study.

Table 1: Group Statistics

| | Gender | Number of Sample | Mean | Std. Dev | Std. Error Mean |
|---|---|---|---|---|---|
| Total Score | Male | 77 | 46.68 | 5.11 | .58 |
| | Female | 218 | 46.96 | 4.82 | .33 |
| | **Total** | **295** | | | |

Table 2: Independent Samples Test

| | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | |
|---|---|---|---|---|---|---|---|
| | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference |
| Equal variances assumed | .014 | .905 | -.444 | 293 | .657 | -.28798 | .64883 |
| Equal variances not assumed | | | -.432 | 126.87 | .667 | -.28798 | .66730 |

Subsequent gender analysis will be presented from the standpoint of the instrument to determine whether there is gender bias in the constructs used in the MeT instrument. A total of 295 respondent data were used to perform the GDIF analysis [24, 25].

Item analysis to determine if gender bias exists in the MeT instrument was done using version 3.64.2 Winstep software. To determine whether there is GDIF or not, three indicators were used [26], namely:

(i) t value of < -2.0 or > 2.0
(ii) DIF contrast value of < -0.5 or > 0.5
(iii) p (Probability) value < 0.05 atau > -0.05

The three indicators were examined accordingly. Each item needs to meet those three conditions to be considered bias and be dropped from the instrument. However, if the item meets only one of the conditions, it should not be dropped but instead it should be separated and fixed. Based on those characteristics, GDIF for the MeT constructs can be determined from Fig. 9 – Fig. 13.

Fig. 9 shows a good pattern both for male and female. There is not much gap or distance between both lines. Item B02 is item number 2 from section B of the I-MeT instrument which is a rubric item with two choices of answer (i) – "*Little of my time is spent gainfully engaged with experts outside the institution and (ii) I often involved in activities with experts outside the institution*). This item exhibit a little distance between the two lines suggesting the item is more difficult for females (Red line 2) to answer and more convenient for male (Blue line 1) to answer. However, the visible distance is very small and not much difference can be measured, thus it is save to conclude that no gender biasness exist in items for the first construct (**Cooperation**).

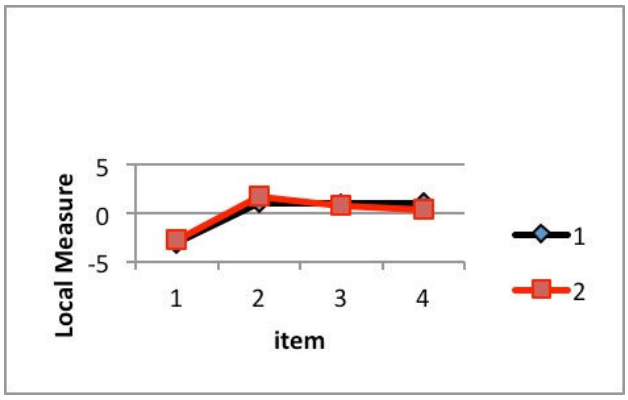


Fig. 9: GDIF for **Cooperation** Construct

Same as the first construct in Fig. 9, all five items in the second construct (**Activity**) as shown in Fig. 10 does not show any sign of biasness where even the hardest item B06 seems to be easier for male students, but the values are not strong enough to make it a bias item. Item B06 is item number 6 from section B of the I-MeT instrument which is a rubric item with three choices of answer, (i) *I rarely think or write about my activities and reflections,* (ii) *I often stop and think about the activities in which I am engaged,* (iii) *I write to share my observations about my activities.* Therefore, we would conclude that all five items are fair for both male and female students.

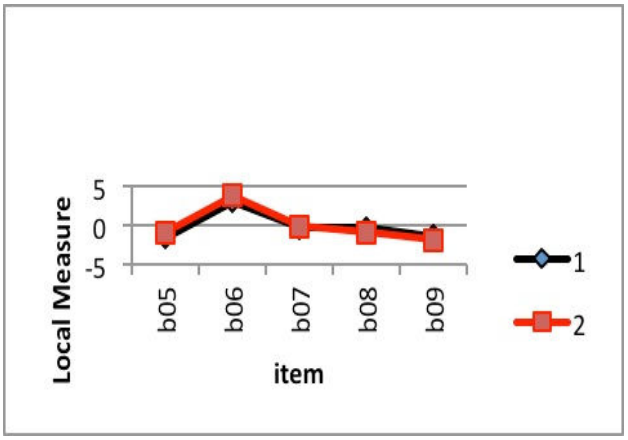

Fig. 10: GDIF for **Activity** Construct.

For the third construct **Authenticity**, when referred to Fig. 11, all items seem to have good values and fair for both male and female students. The t value of 1.94, if rounded statistically will be 2.00. This value is the cut-off point for GDIF. DIF contrast value is above 0.5 which is 0.65. From the graph, the lines look far enough between male and female students. Going back to the item in question, it appears that item B012 is about recognizing problem. Item B012 is a rubric item with three choices, (i) *Learners are not expected to be problem finders, but are instead expected to be able to solve well-structured tasks,* (ii) *Learners are expected to refine given task as well as solve it, and* (iii) *Learners develop skill and proficiency after identifying, defining and solving various task.* In this matters, females in general have been known to be able to identify problem easier than their male counter parts [27]. This does make sense because male tend to think in simpler terms and does not easily recognize small details or issues as a problem.

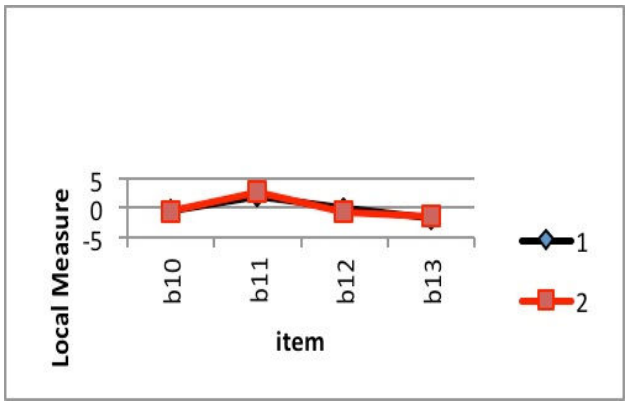

Fig. 11: GDIF for **Authenticity** construct

For the fourth construct **Constructivity**, when referred to Fig. 12, all items measuring the construct have good values and fair for both male and female students. The construct try to gather information about how much struggle learners have to put up in order to become an expert and solve problems. Apparently there is also not much difference where male and female both have to struggle to solve problems particularly involving e-learning. This is not surprising as Malaysia has revealed herself as a developing country where male and female alike must struggle to obtain equal rights and opportunities to become experts in various fields.

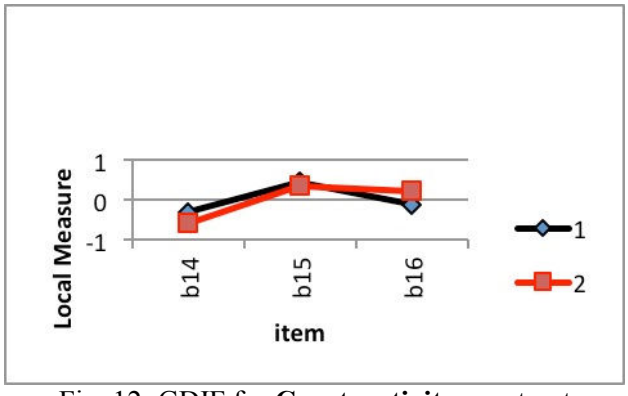

Fig. 12: GDIF for **Constructivity** construct

For the last construct (**Intentionality**) as can be referred to in Fig. 13, item B22 meets all the 3 criteria to be given the verdict of gender bias. This item is easier for female as opposed to the male counterpart. Item B022 is a rubric item with three choices, (i) *The use of technology seems unrelated to thinking,* (ii) *The use of technology contributes to thinking, (iii) The use of technology makes a powerful contribution to the thinking process.*

This conclusion is drawn from the three evidences where first, the t value is 2.18 which is > 2.00. Secondly, the DIF contrast of 0.71 is way above 0.5. Lastly, it is evidence by the p value which is < 0.03. Thus this item is gender bias and should be dropped from the instrument.

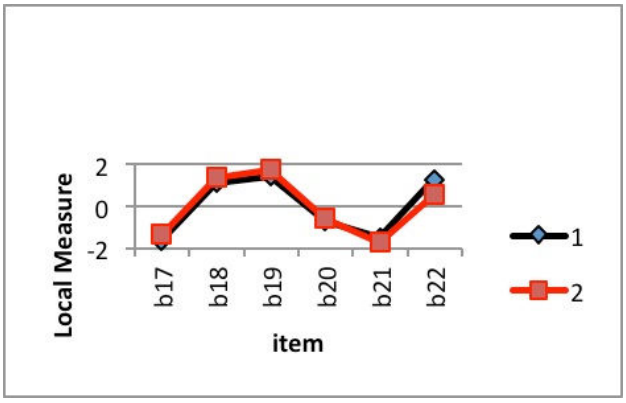


Fig. 13: GDIF for **Intentionality** construct

There were several interesting issues in regards to the research findings. First, it was found that there was no significant difference found between males and females in terms of gaining meaningful learning when experiencing technology training. This shows that there is no difference in the ability of males and females to learn meaningfully using technology unlike what have been taken for granted by our society who believed that males expected to learn better about machines, tools and how things work. There is a clear pattern from this study that shows although females are usually better in mathematics and scientific study programs, and are never really comfortable around machines and technology, they are still capable of achieving meaningful learning when required to use technology in their learning experience.

Secondly, females were found to be able to identify problems easier than males. This may be so due to their awareness towards details when males sometimes focus on the overall picture instead of the details. More often than not, males are more likely to approach things in a more direct and "short-cut" approach which in turn may cause them to bypass some important details.

Both Malaysian males and females are resilient in struggling to be an expert in their field. This is a criteria needed for citizens in a developing country to move forward [28, 29, 30]. Another interesting phenomenon to be discussed is the biasness that exists in the item that asks about problems in relation to technology use to support critical and critical thinking. Surprisingly, the item was easy for females to answer as compared to males which are contrary to what the society believes and findings from previous studies [31, 32]. Again, this is a welcoming phenomenon for a developing country such as Malaysia to move forward as a developed country.

## 4 Conclusion

Based on the findings of this study, it can be concluded that MeT is a valid and reliable instrument since only one out of 22 items were found to be gender bias and another one only need to be improved in terms of sentence structure so that it could be more easily understandable by the respondents. In short, a concluding table is presented in Table 3 summarizing the results of GDIF analysis done in this study.

Table 3: GDIF Summary

| Construct | Original Items | Item GDIF | Remaining Items |
|---|---|---|---|
| Cooperation | b01, b02, b03, b04 | 0 | 'b01, b02, b03, b04 |
| Activity | b05, b06, b07, b08, b09 | 0 | b05, b06, b07, b08,b09 |
| Authenticity | b10, b11, b12, b013 | 0 | 'b10, b11, b12, b013 |
| Construction | 'b14, b15, b16 | 0 | 'b14,b15, b16 |
| Intentionality | b17, b18, b19, b20, b21, b22 | 1 | 'b17,b18, b19,b20, b21 |
| **Total** | **22** | **1** | **21** |

## 5 Acknowledgement

In accomplishing this report, we would like to convey our greatest appreciation to Professor M Sahari Nordin, Dean of Research and Innovation Centre, International Islamic University Malaysia who had generously given us his time and insights on details of data analysis using the Principal Component Analysis, Associate Professor Dr. Norlide Abu Kassim who taught us about Factor Analysis and Rasch, Dr. Tunku Badariah Tunku Ahmad for her help in editing and reviewing the

research report and Pn. Kemboja from the Language Center of the university for reviewing and editing the language of the earlier versions of the instrument. The first author would also like to thank both her PhD supervisors, Associate Professor Dr. M Shanudin Zakaria from FTSM, UKM and Prof. Dr. Khairul Anwar Mastor from PPU, UKM in which this work have derived from. I would also like to thank Prof Mohamed Amin Embi and Associate Prof. Dr. Norazah Nordin as leaders of the E-Learning OUP Research Group, Prof. Ruzy Suliza as Head of Social Science PTS Research Group and Assoc. Prof Dr. Saemah Rahman as the leader for Innovasion OUP Research Group. Last but not least, we would like to thank the government of Malaysia and the Center for Research and Innovation Management at the University Kebangsaan Malaysia for the university grant [UKM-GUP-TMK-08-03-308] and [PTS-2012-17]GGPM-CMNB-026-2010] given to support the study.

*References:*

[1] H.M. El-Bakry, A. A., Saleh, T. T. Asfour, & N. Mastorakis, A new adaptive e-learning model based on learner's styles, Proceedings of the 13th WSEAS International Conference on Mathematical and Computational Methods in Science and Engineering, 2011

[2] E. Özpolat, & G.B. Akar, Automatic detection of learning styles for an e-learning system, Computers & Education, Vol. 53, No.2:2009, pp. 355-367.

[3] E. Popescu, Adaptation provisioning with respect to learning styles in a Web-based educational system: an experimental study, Journal of Computer Assisted Learning, Vol 26, No.4:2010, pp. 243-257.

[4] C., Ipbuker, Learning styles and teaching models in engineering education, Proceeding of the 6th WSEAS International Conference on Engineering Education. 2009.

[5] C.J. MacDonald, & M.A. Gabriel, Toward a partnership model for web based learning, *The Internet and Higher Education: A Quarterly Review of Innovations in Post-Secondary* Education, Vol. 1, No.3,1998, pp.203-216.

[6] M.J. Rosenberg, E-*learning: Strategies for delivering knowledge in the digital age*, McGraw Hill Inc., 2001.

[7] C.J. MacDonald, E. Stodel, L. Farres, K. Breithaupt, & M.A. Gabriel, The demand driven learning model: A framework for web-based learning. *The Internet and Higher Education,* Vol. 1, No.4, 2001, pp.9-30.

[8] C.J. MacDonald, K. Breithaupt, E. Stodel, L. Farres, & M.A. Gabriel, Evaluation of web-based educational programs: a pilot study of the demand-driven learning model. *International Journal of Testing,* Vol. 2, No. 1, 2002, pp. 35 - 61.

[9] K. Breithaupt, & C.J. MacDonald, Quality standards for e-learning: Cross validation study of the Demand-Driven Learning Model (DDLM), *Testing International*, Vol. 13, No. 1, 2003, pp. 8-12.

[10] C. J. MacDonald, **&** T.L. Thompson, Structure, content, delivery, service and outcomes: Quality e-learning in higher education. *International Review of Research in Open and Distance Learning,* Vol. 6, No. *2.*

[11] C.J. MacDonald, D. Archibald, E. Stodel, P. Hall, Knowledge translation of interprofessional collaborative patient-centred practice: The working together project experience, *McGill Journal of Education.* Vol. 43, No. 3, 2008, pp. 283-307.

[12] S. Liaw, H. Huang, & G.C. Huang, An activity-theoretical approach to investigate learners' factors towards e-learning systems. *Computers in Human Behavior*, Vol. 23, No. 4: 2007, pp. 1906-1920.

[13] S.H.S.H. Shahabudin, *The development of a Malaysian Qualifications Framework (MQF),* Ad-hoc Inter Agency Meeting, Ministry of Education, 6 November 2003.

[14] S.H.S.H. Shahabudin, The *development of a Malaysian Qualifications Framework (MQF),* Ministry of Higher Education, Malaysia, 2004.

[15] Committee of Deputy Vice Chancellors and Rectors of Malaysian Higher Learning Institutes, *Strategi dan Piawaian Pengajaran dan Pembelajaran IPTA*, 2006.

[16] M. Verkroost, L. Meijerink, H. Lintsen, & W.Veen, Finding a balance in dimensions of blended learning, International *Journal on e-learning*, Vol. 7, No. 3, 2008, pp. 499-522.

[17] T. Anderson, & F. Elloumi, *Theory and practice of online learning,* Athabasca: Burbules & Callister, 2004.

[18] G. Salmon, *e-Moderating: the key to teaching and learning online*, Kogan Page, 2000.

[19] M. Scardamalia, & C. Bereiter, Computer support for knowledge-building communities, *Journal of Learning Sciences,* Vol. 3, No. 3, 1993, pp.265-84.

[20] J. Reid, Perceptual learning style preference questionnaire, [Online] http://lookingahead. hienle.com/filing/l-styles.htm, Accessed on December 12th, 2006.

[21] V.H.K., Hung, M. Keppel, M.S.Y. Jong, and Learners as producers: Using project based learning to enhance meaningful learning through digital video production, Proceedings *of ASCILITE 2004,* 2004.

[22] R., Din, H., Norman, A.A. Karim, R. Rahmat, & M.F. Kamarulzaman, F. Hybrid E-Training Assessment Tool for Higher Education, 2012.

[23] R. Din, M.S. Zakaria, K.A. Mastor, N. Abdul Razak, M.A. Embi, & S.R. Ariffin, Meaningful hybrid e-training model via POPEYE orientation, *WSEAS International Journal of Education and Information Technologies*, Vol. 3, No. 1, 2009.

[24] F.J. Abad, R. Colom, I. Rebollo, & S.Escorial, Sex differential item functioning in the raven's advanced progressive matrices: Evidence for bias, *Pergamon Journal: Personality and Individual Differences* 36; 2004, pp. 1459–1470.

[25] J. Durand, & S. Park, *A study of gender and academic major-based differential item functioning (DIF) in KEPT Mexico*, 2006.

[26] M. Tennant, & J. Pallant, *SPSS Survival.* 3rd Ed., Open University Press, 2007.

[27] H. Hamzah, S.R. Ariffin, & R.M. Yassin, Girl's overachievement in the mathematics national examination: Can type of items be the cause?, *Journal of Science and Mathematics in Southeast Asia*, Vol. 29, No. 2, 2006, pp. 1-20.

[28] J. Hyde, How large are cognitive gender differences? A meta-analysis, *American Psychologist Journal*, Vol. 36, 1981, pp.892–901.

[29] J. Guiller, A. Ross, & A. Durndell, *The role of gender in a peer-based critical thinking task*, Research Developments in Learning Technologies, 2005.

[30] M.A. Miller, & N.S. Malcolm, Critical thinking in the nursing curriculum, *Nursing & Health Care Journal*, Vol. 11, No. 2, 1990, pp. 66-73.

[31] B. Gadzella, D. Ginther, & G. Bryant, Prediction of performance in an academic course by scores on measures of learning style and critical thinking, *Psychological Report Vol.* 81, No. 1990, pp. 595-602.

[32] T. Elizabeth, What Is Critical Thinking, *Journal of Excellence in Higher Education,* University of Phoenix, 2005.