p-ISSN 1693-4849
e-ISSN 2549-2306

Jurnal SERAMBI ILMU

Journal of Scientific Information and Educational Creativity

VOLUME 22 NOMOR 1 EDISI MARET 2021

Diterbitkan Oleh
FKIP Uviversitas Serambi Mekkah Banda Aceh

Jurnal Serambi Ilmu	Volume 22	Nomor 1	Hal. 1 - 149	Banda Aceh March 2021

Computer Adaptive Test Development To Assess Students' Psychology

Phan Thi Thu Nguyet,[1] and Muslem Daud[2]

[1]Phan Thi Thu Nguyet is a Lecturer of National University of Social Sciences and Humanities, Ho Chi Minh City, Vietnam
Email: phannguyet@hcmussh.edu.vn
[2]Muslem Daud is a Lecturer of Universitas Serambi Mekkah, Banda Aceh, Indonesia.
Email: muslem.daud@serambimekkah.ac.id

Abstract
Psychology stability plays great role in students' learning, and stress becomes a significantly serious issue among university students. An efficient tools is needed to understand it more. The aim of present study is to develop a Computerized Adaptive Testing (CAT) as pioneer project in South East Asia, and starting in Vietnam. In this vein, an item bank of 68 items has been constructed, which is based on Likert Polytomous Scales through five sub-domains: behavior, academic performance, family, lecturer and finance. It has assessed 2,085 students (704 males and 1,381 females). Multidimensional Random Coefficients Multinomial Logit (MRCML) Model is applied to develop Multidimensional Stress Scales and Computerized Adaptive Testing procedure. The result findings indicate that Multidimensional Random Coefficients Multinomial Logit (MRCML) can be used to develop new scale with psychometric properties. Indicated by various fit criteria MNSQ, standard errors, Z (t-test) implemented in software ConQuest. The sub-domain has a good reliability (from .857 to .798). As respect to CATs, a simulated experiment based on the empirical data is applied to evaluate the performance of the proposed computerized adaptive testing. The standard error of the estimated stress proficiencies are reported in this study. The 68 items stress data appropriate fit the Multidimensional model applied.

Keywords: stress assessment, psychology test, computerized adaptive test

INTRODUCTION

Large-scale assessment is usually used to improve the educational process in order to monitor student achievement (Colwell: 2013). To do this, traditional methode was run to collect data and information so that the program for improvement can be developed proportionally and strategically. In traditional assessment, paper pencil, writing test and check lists have been a basic ways for data collection. Therefore, if the data collection is for large-scale assessment such as national assessment, this would take a larger effort for administrative and paper work, paper test distribution and human resource mobilization different part of a country (Kuo, Daud & Yang: 2015). With technology, this traditional mode has been gradually left.

Nowadays, with the invention of alternative method by using technology, project manager has been helped to support his/her work in data and information collection far

easier. To do that, computerized assessment delivery system could evaluate how student performance. Besides, this advance in technology makes it possible to capture more complex performances. Computerized adaptive testing (CAT) is test administered by a computer (Linden & Pashley: 2000). The first administered College Board exam occurred in 1901, also in 1926 the first Scholastic Aptitude Test (SAT) was given, then the Test of English as a Foreign Language (TOEFL).

Adaptive tests are consisted of items selected from a collection of items, known as an item bank. Item bank assumed the availability of large collection of items stored with their characteristics in a computer (Daud, Kuo, Yusrizal, Mok: 2018). CAT starts with an item of average difficulty, if the student succeeds on response, a slightly more challenging item is presented next or if the student responses an item wrongly, the computer will select a subsequently easier item. This process continues until the computer has enough information to produce a reliable test score.

CAT was designed for cognitive testing and it is an ongoing studies. Various CAT procedures for an online-assessment continually (Daud, Kuo, & Mok: 2015). Many CATs are on going used and improvement, for example in mathematics (Yang, Kuo, and Liao: 2011), and in Biology (Kuo, *et al*: 2015). CAT had now been developed for depression (Fliege, Becker, Walter, Rose, Bjorner, & Klapp: 2009), and anxiety (Walter, Becker, Bjorner, Fliege, Klapp & Rose, : 2007). This research is purpose to development a Computerized Adaptive Testing (CAT) to measure the mentioned stress of Vietnam University Students, based on Multidimensional Random Coefficients Multinomial Logit (MRCML) Model perspective.

Psychological Aspect in Learning

Many aspects can interfere students' psychology stability and these would become reasons on their lower achievements. Those factors can be internal and also external of the students' (The First Psychology: 2017, Cabib, Campus and Colelli: 2012). Internal factor mostly come from inner of an individual becomes reflected in a behavior of individual students (Hill. & Wallace: 2011). Student who cannot coup with his matter can affect his psychology stability. The worse is when a students bring out his instability in form of angriness and could lead to bad habits such as drinking alcohol (Kumar & Bhukar, 2013: Brady & Sonne:1999), smoking, and other shorter solution to reduce his instability (The First Psychology: 2017). In fact, even though this might not apply to all South Asian Countries, drinking alcohols become one of significant factor that bring negative impact on students' life and education. Therefore this internal psychology factor has been put as one variable to be discussed in this research.

External factors are factors impact on students from outside students. External factors that focuses in this research are on daily students' life are from academic burden, lecturer's relationship, financial management and family support while his/her study accomplishment (Balasheva & Pretrova, 2016; Kumar & Bhukar, 2013).

There are a lot of academic factor to impact students psychology stability (Rafiq & Shah: 2015). Examination and perform in front of class also two burdens that would make students psychologically feel stressed. This might be experienced by middle to lower achievement's student. While students with higher achievement usually become role models in class that standing in front of class is not an obstacle for them. However, number of these model students most of the time from ranked 1st to 5th. Thus, for those who less prepared to class evaluation in form of examination, midterm test, and

summative test are potentially to feel stressed. This means his/ her psychology would be unstable (Balasheva & Pretrova, 2016).

Meanwhile, relationship with lecturer also play a great role stability on how students get familiar with lesson taught and homework requested for example. When a relationship is poor, then students might get into a interpersonal relationship particularly to the one who need his/her keep runs well (Rafiq & Shah: 2015; Kumar & Bhukar, 2013). However, not all students have this expected situation that to have well maintained relationship and flowed communication with their university teachers. If so, students might have a risk for their psychology stability. Moreover, when a conflict takes place and students have a deadlock communication with a lecturer.

Other important external factors is financial management (Heckman, Lim and Montalto: 2014; Midgley, Attree, Wheeler, and Lee: 2014). This factor can become obstacles not only to students from lower income family but also to those from middle to wealthy family. Financial management is a skill on how amount of fund spent on a stable spending all over students' life. Imagine when a student cannot manage this, his/her money can be over before expected time such as in one month for example. In fact, most students in different universities in Vietnam, Indonesia, and other South Asian countries are come from rural area. They left their family and come to cities to continue their higher education after graduation from various high schools in their towns. This means, they financial support depend on fund supply from their families. Then, they need to manage their money properly and keep flows their life until the end of financial support's period before other money transfer come from their families.

Financial management would be little bit different when these students have ability to coupe with academic burden, and perform successfully in exams, have a reliable relationship with tutors, and time management effectively. Then they can look for financial income with different light work, suitable for academic life. In this situation, then their situation would be better. If not, they still need to keep rely on their family.

Families' expectation is usually become other burden when this student cannot fulfill it. In fact, it is rationale when a father requests a good mark from their sons/ daughters because they pay for their study and assure their financial support. For student who registered their studies suitable with their academic background his/her stress would be lighter than students have chosen department wrongly. Then, changing or moving to appropriate department also take time and would extend study duration when other fellows are ready for graduation.

Other issue in this external factor is homesick to be with families. In this case many students move their study to a university closer to their hometown. This would be a solution but still time consuming and waste of energy. Then, how those students cope with this situation depends to their self-esteem management to cope with stress (Hill. & Wallace: 2011).

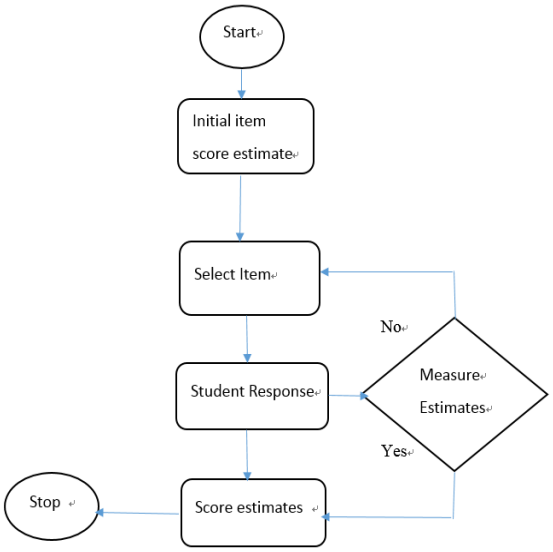
Computerized Adaptive Testing

In Computer Adaptive Testing, an item is selected, and offered to test taker to start the test. This process is a behind screen of the computer activity, while test taker can only an item of question for example appears on the computer screen (Daud, et al.: 2015). When test taker cannot answer the offered item/ question, then computer calibatively selects easier item and offers to test-taker. In fact, Computerized adaptive testing that based on Item Response Theory (IRT) has been used in a family of mathematical functions and evaluation of cognitive and higher abilities tests.

There are four major components of an Adaptive test:

- Calibrated item bank
- Item selection
- Trait estimation (Scoring)
- Stopping rule

This procedure is illustrated in Figure 1.



Flow diagram of an adaptive test.

As figure 1 shows the test begins with an initial estimate of respondent's score form an item offered. The initial score is used to select the most information item ether harder or easier item to be administered to test taker. Test taker/ student responses to different domains (subject and sub-subject tested) are used to estimate his/her competencies until the computer determines stopping criterion and then defined is met. Measurement precision is calculated as a confidence interval after each response. This determination and stop criteria is set by CAT developer.

Multidimensional Item Response Theory

Multidimensional Random Coefficients Multinomial Logit (MRCML) model is used in this research. The MRCML model has been used most widely in recent CAT development particularly large scale assessment but still has result in many domains tested. Adam, Wilson, & Wang: 1997 said main reason that MRCML is able to measure the partial credit scoring for a test with multidimensional domains.

The adaptive algorithm was based on the formal of the MRCML model formulation:

$$P(x;\xi|\theta)=\frac{\exp(x'(B\theta+A\xi))}{\sum_{Z\in\Omega}\exp(z'(B\theta+A\xi))}. \qquad (1)$$

Where θ is the vector of each dimension, ξ is the vector of items parameter, Ω is the set of all possible response patterns. A is scoring matrix B, and Z denotes a pattern from the full set of response, while x denotes the response pattern of interest. The response pattern, x, is comprised of vector for each item with one element in the vector for each item category.

$x=\{x_1 x_2 ..., x_I\}=\{x_{11}, x_{12} ..., x_{1m1}, x_{21}, x_{22,}.., x_{\mathrm{Im}\,I}\}$ for m_i = number of categories for item i, and I = number of items.

This formulation the item parameters, ξ, are considered known and conditioned θ. The Multidimensional RCML model allows us to construct response probabilities and proficiency estimates across multiple dimensions of knowledge, behavior, or attitude (Wang, Kuo, & Chao: 2010). Multidimensional RCML implemented in software ConQuest that advantage is that parameter estimates for very large sample and long test can be obtained (Wu, Adam, & Wilson: 1998).

Scale Development

An item bank has been constructed of dichotomous items. A common version used a scale with two categories by yes or no (correct/incorrect), scored 0 to 1, but with polytomous items as in this current research has five categories (never, rarely, sometime, often, almost always), score with 0, 1, 2, 3, and 4 (Never=0, Rarely=1, Sometime=2, Often=3, Almost always=4).

The degree of difficulties performed specified functions. The highest score implies the high risk at stress issue. All items are put on the same scale because they estimate simultaneously, the item scores are manipulated mathematically (Adam, Wilson, & Wang: 1997).

The ordered Likert's format for polytomous scale is easier to understand and used in this study purpose. Besides, Likert's polytomous scale can represent the probabilities of stopping from a lowest category to highest categories (Linarce: 2000). The item selection for polytomous data presents multidimensional test is more challenge than for unidimensional dichotomous items test. When there is no item clear difficulty but rather a number of them, one for each inter-category threshold. Accordingly, item selection for polytomous items in this study considered the Multidimensional Partial Credit Scales (Masters: 1982).

Measures

To select items for the computerized adaptive assessment in this study was administered drawing from an item bank of 68 items, which were the most information for the individual taking the CAT. The item bank had been developed by given to

n=2,085 students (704 males and 1381 females), participants had mean age of 1.72 years (SD= 2.00). There were freshmen 708 (34%), sophomores 560 (31.2%), Juniors 556 (26.7%) and Senior 171 (8.2%), included science 840 (40.4%) and social science 1242 (59.6%) majors, nine departments who are studying at four universities in Ho Chi Minh City, Vietnam.

Sixty eight items showing the best item properties were selected to build the Stress-CAT item bank. The five sub-domains item bank covers: behavior, academic, family, lecturer and finance. Item banks below contains information on the wording of each item, the concept it measures, and measurement characteristics

SUB-DOMAINS TO BE DEVELOPED/CALIBRATED

Item bank/ Sub-domain	Exemplar context
Behavior (19 items)	Drink alcohol, angry, smoke..,
Academic (11 items)	Essay, exam, in font of class..,
Family (10 items)	High expected, homesick..,
Lecturer (17 items)	Poor relationship, conflict, boring..,
Financial (11 items)	Pay bill, foods,

Evaluation Of Computerized Adaptive Testing (Cat) Performance

A. Scale Reliability

The Multidimensional item responses was checked for its reliability coefficient, calculated to estimate the internal consistency of each of the Stress-CAT scale. The consistency of the scale is important to accurately shown its reliability whenever used.

The internal consistency reliability estimates of the scale ranged from .857 to .798 respectively. After all, in making decisions of the various stresses reliability of each domain was experienced. Then, reliability of this scale has been properly fulfilled.

B. Item Information Category Probability Curves

The item information function shows latent trait level where the item has more precision and reliability. In Multidimensional IRT, information function plots show the varying precision of the theta θ estimate across the trait continuum.

Almost all item characteristics curves (ICC) showed a well fit of the items. The two items below with the most extreme standard Multidimensional model in are Item 8 (MNSQ=0.85) and Item 29 (MNSQ=1.30)

When items have more than two possible outcomes, they need more information than the item difficulty. Mok and Zhu (2014) said that it is needed to know more how much of the construct (behavior, knowledge or attitude) items to achieve each possible score on every item. These would contribute to overall estimate of test takers' achievement.

As the Figure 1 below illustrates the Item 8 characteristic curves (ICC) and maximum information function for Item 8. Each curved line in the upper plots represents the model's prediction of the probability of choosing each of item response categories for various degrees of nervous and anxiety impact.

Anxiety

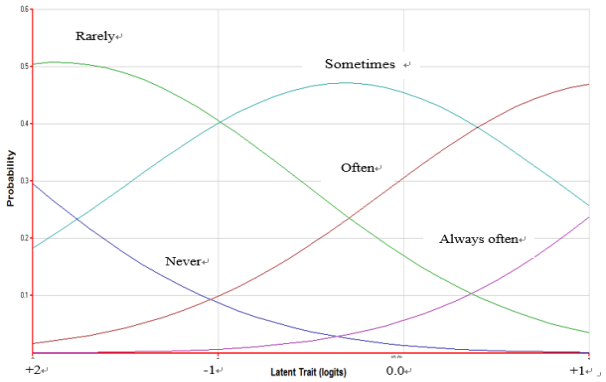


Figure 1. Item characteristics and maximum information functions

Mathematically expressed the Greek letter theta θ (Latent trait) axis reflects anxiety impact MIRT estimates. Thus, the plots show that most of respondents endorse category 1 'Rarely' located at the negative. The curves for category 2 'Sometimes' is interested at 1 'Rarely' and 3 'often' location. Subsequent step in difficulty are interested in much the same way. The highest of the Item information function at its maximum reflects the discrimination of the Estimates stress.
Get the deadline for assignment

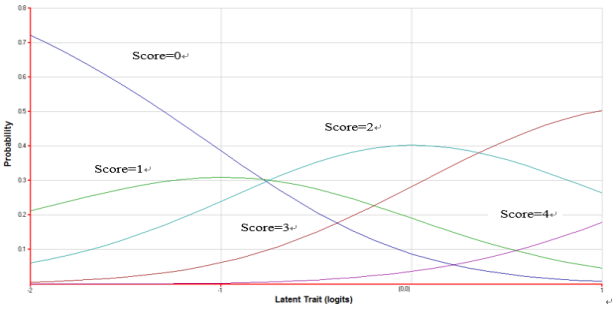


Figure 2. Five categories score ICC

The step difficulty parameters represent the values on the θ (Latent trait) scale at which consecutive category response curves intersect. The relative order of the step difficulty parameters (i.e., intersections) indicates that going from 0 to 1 (Step 1) or 2 to 3 (Step 3) is relatively easy for students, while going from 1 to 2 (Step 2) is moderately difficult, and going from 3 to 4 (Step 4) is more difficult. Furthermore, the effect of the reversal (i.e., Step 2 being more difficulty than (Step 3) can also be seen in the lower probability of receiving a score of 2 relative to the other categories. This is one of the most extreme standard Multidimensional models in is Item 29 (MNSQ=1.30)

C. Weight Map of Latent Distribution and Response Model Parameter Estimates

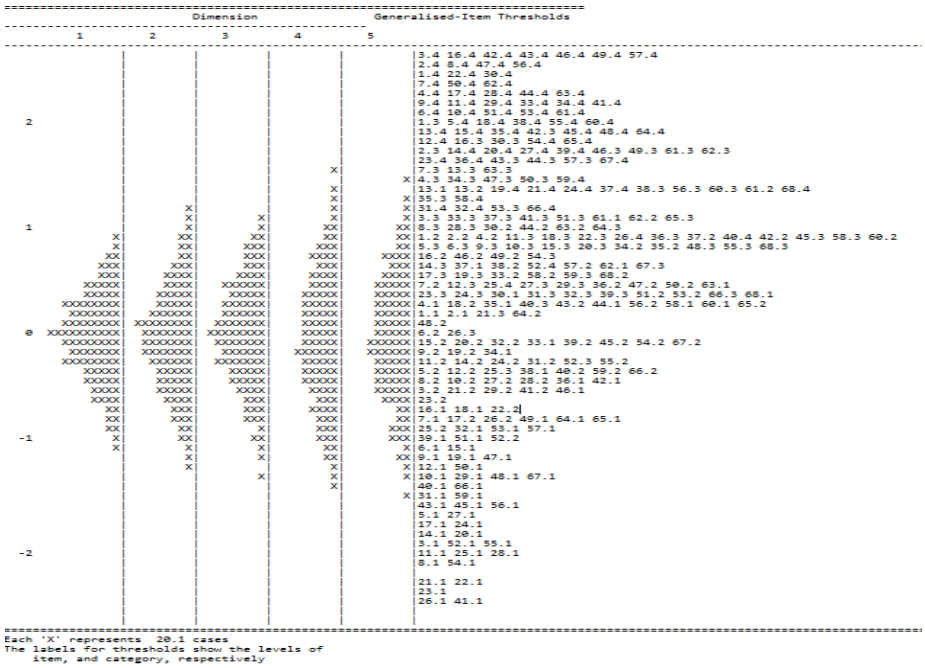


Figure 3. Distribution map for students and items

In Figure 3 shows normal distribution of the data presented. It represents data starting from the position of the item category, can be seen the students and items map which have been measured with the logit scale. The location of students (students column) on the left side and the item stress calibration on the right side from the top to the bottom, with each 'X' representing distribution that has highest score at the top, the middle score in the middle and lowest score appears near the bottom. Therefore, Multidimensional Partial Credit model showed the relationship between the scores obtained with 68 item test and those measures assigned to the students.

D. Simulation study for Stress-Computerized Assessment precision

In regard to the Multidimensional of Stress items, and performance of CAT systems, a simulation study was conducted. This simulation was purposed to see if the system developed can run accurately and present acceptable Standard Error (SE) as follows:

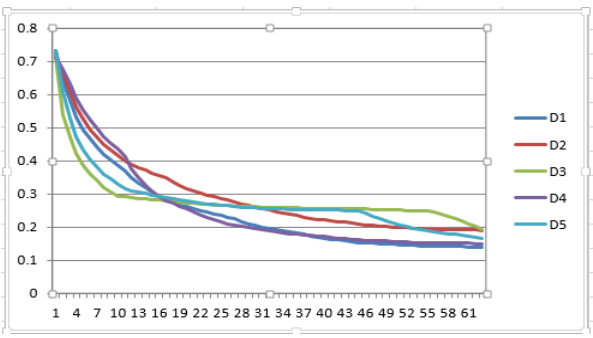


Figure 4. The stopping rule criteria (SE<0.30)

The implementation of algorithms for Multidimensional polytomous data, a new estimation of θ as well as standard error estimation (SE) is values. The stopping rule criteria SE<0.30 for illustrative the purpose as showed in figure 4. When Standard Error is smaller that 0.30, this developed CAT system has passed the minimum requirement for SE. The stopping rule used was the pre-specified level of assessment precision. Then, this system has fulfilled the requirement under approved SE rate.

CONCLUSION

This study on CAT Psychology particularly Stress Multidimensional Computerized Adaptive Testing assessment has been initiated. The process of development has been carried out as required for CAT system development with data collection, data calibration, data simulation, and finds the proper SE for the system. The system also demonstrated its measurement precision from data collected.

Objectivity in Multidimensional Random Coefficients Multinomial Logit (MRCML) model to magnitude of a measure estimated is affected in important of applied Computerized Adaptive Assessment. In fact, there are 68 items in the study that score as polytomous variable, iindicated by various fit criteria MNSQ, standard errors, Z (t-test). The item stress a good item fit to model. Fitting five dimension Multidimensional Partial Credit model (MPCM) between-item multidimensionality, the dimensions being the content areas.

The amount of item is accepted. The sub-domain has a good reliability (from .857 to .798). Students have given the opportunity to respond and to share with educator their stress issues with a variety of item types and various item information are known to differ in the type of stress they gather. Based on the simulated experiment, this Computerized Adaptive Assessment has tested the items thoroughly in all five sub-domains. Simulation study has been carried out and SE test has been conducted. With its Standard Error is smaller that 0.30, this developed CAT system has passed the minimum requirement for SE.

This work is the first pioneer research that to applying a Computerized Adaptive Assessment on Psychology of Vietnamese university students. By using this CAT system, it is expected to know test results immediately at the end of the test. It is also expected that this system is equipped with feedback to students who are in stressed situation with their different problems of inappropriate behavior, academic burden, and improper relationship with lecturer, financial problem and missing their families. Computerized Adaptive Assessment proven that test shorter and give a higher level of precision.

A similar study for other South Asian countries would be a recommendation with its modification wherever needed. This would enrich researches in the area and their comparison, as well as to look for their solutions.

ACKNOWLEDGEMENT

This research was supported by Vietnamese university students who were willing and voluntary participants. The authors are also grateful to K-Lab team of National Taichung University of Education, Taiwan.

REFERENCES

Adam, R.J., Wilson, M. & Wang, W.C., (1997) "The Multidimensional Random Coefficients Multinomial Logit Model." *Applied Psychological Measurement*. Vol. 21 No.1, pp. 1-12, 1997.

Balasheva, I.V. & Pretrova, N. F, 2016, Psychological stability of a personality and capability of tolerant interaction as divers manifestations of tolerance, *International Journal of Environmental and Science Education,* Vol. 11 (10), p.3367-3384

Brady, K. T. & Sonne, S.C, 1999, The Role of stress in alcohol use, alcoholism treatment, and relapse, *Alcohol Research and Heakth*, Vol. 23 (4), p. 263-271

Cabib, S., Campus, P., and Colelli, V., 2012, Learning to cope with stress: psychobiological mechanism of stress resilience, *Reviews in the Neurosciences*, Vol. 23 (5-6), p. 659-672

Colwell, N.M. (2013) "Test Anxiety, Computer-Adaptive Testing and the Common Core," *Journal of Education and Training Studies.* Vol. 1, No. 2, October 2013

Daud, M., Kuo, B.C., Yusrizal, Y., and Mok, M.C.M. (2018) *Computerised Adaptive Testing: Application and Implications for Teaching and Learning in Indonesia,* in Routledge International Handbook of Schools and Schooling in Asia, England, Routledge Inc.

Daud, M., Kuo, B.C., & Mok M.C.M. (2015). The Construction of an online curriculum-based multidimensional computerized adaptive testing, In *International Multidisciplinary Conference* (pp.322-328). Banda Aceh. Universitas Serambi Mekkah.

First Psychology, 2017, The Stress Factor: Your guide to stress, UK

Fliege, H., Becker, J., Walter, O.B, Rose, M., Bjorner, J.B., & Klapp., B.F.,,(2009) "Evaluation of a computer-adaptive test for the assessment of depression (D-CAT) in clinical application." *International Journal of Methods in Psychiatric Research. Int. J. Methods Psychiatr.* Vol. 18, No 1, pp. 23-36, February 2009

Hill, V.Z. & Wallace, M.T., 2011, Self-esteem instability and psychological Adjustment, *Self and Identity,* Vol 11, p. 317-342

Kuo. B.C, Daud, M., & Yang, C.W. (2015); Multidimensional Computerized Adaptive Testing for Indonesia Junior High School Biology. *Eurasia Journal of Mathematics, Science & Technology Education, 11*(5), 1105-1118

Linarce, J.M., (2000), "Understanding Rasch measurement: Optimizing category effectiveness." *Journal of Applied Measurement.* Vol. 10, pp.85-106

Linden. W.J, & Pashley., P.J., (2000). Item selection and Ability Estimation in Adaptive Testing. In W.J. van d er Linden & C.W. Glas, "Computerized Adaptive Testing: Theory and Practice. T*he Netherlands: Kluwer Academic Pulishers, 2000, ch.1,* pp1.

Masters., G.N, (1982) "A Rasch model for Partial credit scoring." *Psychometrika. Pp.* 149-174. 1982

Midgley,. T., Attree, L., Wheeler, T., and Lee, H.V., (2014), Defining and Measuring the External Stress Factors that Lead to Conflict in the Context of the Post-2015 Agenda, Savetheworld, UK

Mok, M. M. C. & Zhu, J. (2014).Cluster analysis of attitudes towards feedback and their mathematics achievement: a study of Hong Kong primary students. *Asia Pacific Journal of Educational Development, 3*(2), 1-13. DOI: 10.6228/APJED.03.02.01

Kumar, S. & Bhukar, J.P., 2013, Stress level and coping strategies of college students, *Journal of Physical Education and Sports Management,* Vol. 4 (1), p. 5-11

Rafiq, R. & Shah, P.A., 2015, Teachers' role and stress relationship: A Review, *The International Journal of Human Resource Management and Industrial Research*, Vol 2, (1), 8-17

Walter, O.B., Becker, J., Bjorner, J.B, Fliege, H., Klapp, B.F. & Rose, M., (2007), "Development and evaluation of a computer adaptive test for Anxiety (Anxiety-CAT)". *Quality of Life Research*. Vol. *16*, pp. 143-155, 2007.

Wang, H.P., Kuo, B.C., & Chao, R.C (2010) "A Multidimensional Computerized Adaptive Testing System for Enhancing the Chinese as Second Language Proficiency Test" *Selected Topics In Education And Educational Technology*. 2010.

Wu, M.L, Adam, R.J, & Wilson, M.R. (1998) "ACER ConQuest: generalized item response modelling software." *Mebourne, Australia: Australian Council for Educational Research.*

Yang., C.W., Kuo, B.C, & Liao, C.H. (2011) "A HO-IRT Based diagnostic assessment system with constructed response items." *TOJET: The Turkish Online Journal of Educational Technology*. Vol.10, No 4, pp46. October, 2011.

Your Step to Heckman, S., Lim, H. and Montalto, S., 2014, Factors related to financial stress among college students, *Journal of Financial Therapy*, Vol. 5 (1), 1, p. 19-39