# Developing of Bloomian HOTS Physics Test: Content and Construct Validation of The PhysTeBloHOTS

# Developing of Bloomian HOTS Physics Test: Content and Construct Validation of The PhysTeBloHOTS

**Edi Istiyono[1*], Wipsar Sunu Brams Dwandaru[2], Muthmainah[3]**

[1,2] Department of Mathematics Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta
[3] Graduate School, Universitas Negeri Yogyakarta

[*]Corresponding author e-mail: edi_istiyono@uny.ac.id

**Abstract**. Higher order thinking skills (HOTS) of the students in physics learning need to be assessed. There are several kinds of HOTS, one of which is Bloomian HOTS. The Bloomian HOTS Physics test for Senior High School Students called PhysTeBloHOTS has been developed to assess students' HOTS. The purpose of this study is to prove validation of the PhysTeBloHOTS. There were two steps of validation: proving the content validity and construct validity. The content validity analysis was conducted using Aiken's V and the analysis of construct validation was conducted using exploratory factor analysis (EFA). The PhysTeBloHOTS consisted of four test sets: A, B, C, and D. The each of test set consists of 40 items with 8 anchor items. For the test trial, PhysTeBloHOTS were administrated to 301 Senior High School Students in Yogyakarta. The result of the study shows that PhysTeBloHOTS is valid according to the content and construct validity. The PhysTeBloHOTS items are valid according to the expert judgments with Aiken's V index ranged from 0.83 to 0.97, respectively. The result of EFA and CFA show that PhysTeBloHOTS consist of three aspects: analysis, evaluate, and create.

## 1. Introduction

Assessment is a very important process in learning, especially in physics. Jensen, et.al state that assessment informs what the students expect from the learning process [1]. This statement is corroborated by Kusairi who underlines the importance of assessment done by teachers in learning process, assessment activity can help teachers to understand the strengths and weaknesses possessed by the students during studying time [2]. The more quality is shown in the assessment process, the better is the understanding ofstudents' strengths and weaknesses. As it is still rare that teachers do such assessment to measure students' high order thinking skill, it is important to design a test for that level.

The assessment instrument used by teachers to score students' learning result is a cognitive assessment instrument. Assessment category used is in the cognitive area as revised in Bloom's Taxonomy (Bloomian) which is divided into Lower Order Thinking Skills (LOTS) and Higher Order Thinking Skills (HOTS). So far the most frequently used is Bloomian Lower Order Thinking Skills (LOTS) that includes test assessment using questions in knowledge level C1 (remember), C2 (understand), and C3 (apply), which means the students are not used to answering questions on levelC4(analyze),C5 (evaluate), and C6 (create) [3]. The questions in level C4, C5, and C6 are still rarely encountered. Thus, it is important to develop a test assessment up to the level of Bloomian HOTS.

Teachers' incompetence to measure the students' higher order thinking becomes the reason why it is very important to develop a test to measure those particular skills. The investigation done by Jensen et al, argues that there are many teachers who fail to give a statement on their students' knowledge and thinking skills, instead, teachers only give statement on the students' memorizing aspect that is included in the Lower Order Thinking Skills [1]. Thus it is important to make questions that really measure the higher order thinking skills. Madhuri, Kantamreddi & Goteti argues that active learning that promotes HOTS plays an important role in education system [4].

Instruments used in assessment are usually test and non-test instruments. In this research, test instrument is used to measure the students' skills. Test instruments include multiple choice and essay questions, each has strengths and weaknesses. Multiple choice questions are the most frequent form of test in the assessment process in high school level, while in general most of multiple choice questions are seldom able to measure the students' higher order tinking skills. This becomes a problem as the semester tests and national tests are served in the form of multiple choice [5]. To measure the students' higher order thinking skills, it is better if the question used is reasoning multiple choice, as the development of the regular multiple choice questions. Thus, reasoning multiple choice has to be developed. Cullinane argues that attaching reasons in the second tier of two-tier multiple-choice question can be used to increase the higher order of thinking skill and it can be used to see the students' ability in reasoning [6]. Winarti et al Thus, in choosing the answer, the students have to think about the reasoning of their answers, and it is a straight forward thinking process to decide the proper reasoning so that it is a proper exercise to higher order thinking skills [7].

The frequently used test execution and scoring procedure are step-by-step scoring and item-based scoring, whose result is then summarized. This scoring procedure is used for regular multiple choice questions. This scoring model is not reliable because each question's unique difficulty index is not considered. Alternative approach that can be used is the approach of item response theory for polytomous scoring. This scoring is adapted to the reasoning multiple choice questions. One of the models frequently used in polytomous item scoring is Partial Credit Model (PCM) [8]. PCM is developed to analyze test items that need several steps to do, in which each item follows a partial credit pattern, thus if an individual has a higher skill it is expected that he or she would get a higher score than those who have lower skill. PCM application in modified multiple choice as an alternative model in physics learning assessment is considered effective and fair [9].

The objective to introduce HOTS as the form of the revision of assessment system started in the 21st Century. Even, many developed countries have applied HOTS-based assessment in their education system. Besides, the development of HOTS assessment instrument, according to Hassan, Rosli & Zakaria also can bring benefits, such as ; (1) to increase awareness that there is a benchmark to achieve by students, (2) HOTS guides the students to seek solution and invention (analyze, evaluate, and create), (3) to introduce the students to scientific research, (4) to increase understanding of the subject, especially physics, that is proven to be more effectively taught with HOTS, and (5) to increase the student's competence in investigating and exploring ideas in learning, especially in physics [10].

Not only assessment development, technology development also requires a more effective and efficient assessment process. In the 21st century, various computer-based media are invented to complete classic paper-based test process and helped the advancing of assessment process [11]. Computer-based test is designed using adaptive system in which the next questions given are adapted to the ability level of the students shown by the performance when answering the previous questions. Right now the popular media using adaptive system in test scoring is Computerized Adaptive Test (CAT). Kantrowitz, Dawson, &Fetzer argues that CAT is a part of technology development that helps the teachers as well as the students to improve the learning process, especially a more organized assessment method that adapts to the students' learning ability [12].

The research done by Veeravag, Muthusamy, & Marimuthu shows that the level of thinking as stated by Bloom taxonomy influences the work of the students [13]. HOTS influences the students' ability to answer questions of the higher level. Midyarto, et.al stated that the assessment in students' cognitive

competence can encourage and enable the teacher to monitor the students' competence achievement [14]. Thus it is easier to indicate what to teach and what to prepare for the students' further learning.

Cognitive area in popular Revised Bloom's Taxonomy is referred to as Higher Order Thinking Skills, or what is usually abbreviated as HOTS. The students' higher order thinking skills include the ability to analyze, to evaluate, and to create, in accordance to the levels of C4, C5, and C6 revised Bloom Taxonomy. In this study, the discussed HOTS is HOTS in physics, in which the students are expected to be able to answer Physics questions by analyzing, evaluating, and creating. Physics is usually regarded as a difficult subject by students. Learning the subject needs logical thinking competence, critical thinking, creativity, and objectivity. Drilling the students with the above stated thinking habits will familiarize them to think in higher order thinking skills. Several reasons stated above support the objective of the study, which is to develop the Bloomian HOTS instrument assessment in the form of reasoning multiple choice with the scoring based on PCM to measure students' higher order thinking skills using CAT.

## 2. Methods

This research is research and development (R & D) which aims to develop of Bloomian HOTS Physics Test (PhyTeBloHOTS). The research procedure is carried out through three stages, namely: (a) planning, (b) try out, and (c) measurement and interpretation. The purpose of this study is to prove validation of the PhyTeBloHOTS. The draft of the instruments have been validated to get an expert judgment. Then the instruments tested to 300 first year students in three senior high schools in Yogyakarta. The product trial design in this study included content validation steps and construct validation steps, which were carried out to measure the validity of each item developed.

The content validation is carried out by expert judgement by filling in the validation sheet. The sheet was composed with interval scale of 1 to 4. The data analysis of the validation questionnaire was done in one by the following steps:

1) The first step was to find Aiken's V index using formula [15]:

$$V = \frac{\sum S}{[n(c-1)]} \tag{1}$$

Note :
s = r - lo, lo = lowest validity score,
c = highest validity score,
r = score given by rater

2) The second step is to convert Aiken's V index of every questionnaire item into qualitative data with V index range between 1 to 0. Validation result is declared valid if Aiken's V index results > 0.37.

Construct validity states the extent to which the performance of the test is consistent with the construct in certain theoretical considerations. This study also investigates the construct validity of the instrumen to verify the consistency instrument with the construct of higher order thinking skills on physics subjects. The construct validity can use factor analysis procedures. The exploratory factor analysis (EFA) is a method of factor analysis to identify the relationship between manifest variables or indicator variables in constructing a construct. Besides EFA, confirmatory factor analysis (CFA) is a method of factor analysis to test the indicators that have been grouped according to their latent variables (constructs). The latent variables are consistent in their constructs. The standardized loading factor (SLF) resulting from confirmatory factor analysis (CFA) is expressed as a result of construct validity.

## 3. Result and Discussion

This development research results a product of 62 test items, each is a question fit for Bloomian HOTS with the material of elasticity and Hooke law, static fluid, temperature and calorie, and optical

instruments with 10 anchor items. The test (PhysTeBloHOTS) are divided into two packages, which are 36 questions in Package A and 36 questions in Package B.

*3.1. Content Validation*

In this step, the expert validates the test instrument that was developed. The test developed was PhysTeBloHOTS. The validation score of the PhysTeBloHOTS based on material aspect, construction aspect, language aspect, and appearance aspect. Each test item is scored based on the four scoring aspects. Based on the response and scoring of two experts of physic education scoring and physics expert. Each test item is scored based on the four scoring aspect. Based on the response and scoring of two physics education expert and physics expert, then it resulted Aiken's V index of 62 items from 0.71 to 0.78 shown completely by Tabel 1, so that the entire question items developed is considered valid according to expert and can be used for the further step of trial, which is an empirical trial. This is in accordance to the interpretation steps done by Kowsalya et al, 2012), all of the items are valid if it is in the range of 0.37-1.00.

**Table 1.** Aiken's V of Set A and Set B tems

| ***Set A*** | | | | ***Set B*** | | | |
|---|---|---|---|---|---|---|---|
| *Item* | Aiken's V | *Item* | Aiken's V | *item* | Aiken's V | *Item* | Aiken's V |
| Item 1 | 0,78 | Item 19 | 0,72 | item 1*) | 0,72 | Item 19 | 0,72 |
| Item 2 | 0,73 | Item 20 | 0,72 | Item 2*) | 0,72 | Item 20 | 0,72 |
| Item 3 | 0,73 | Item 21 | 0,74 | Item 3*) | 0,72 | Item 21 | 0,74 |
| Item 4 | 0,73 | Item 22 | 0,72 | Item 4*) | 0,72 | Item 22 | 0,73 |
| Item 5 | 0,75 | Item 23 | 0,72 | Item 5*) | 0,74 | Item 23 | 0,73 |
| Item 6 | 0,72 | Item 24 | 0,72 | Item 6*) | 0,72 | Item 24 | 0,72 |
| Item 7 | 0,72 | Item 25 | 0,72 | Item 7*) | 0,72 | Item 25 | 0,72 |
| Item 8 | 0,72 | Item 26 | 0,72 | Item 8*) | 0,73 | Item 26 | 0,73 |
| Item 9 | 0,72 | Item 27*) | 0,72 | Item 9*) | 0,73 | Item 27 | 0,71 |
| Item 10 | 0,72 | Item 28*) | 0,72 | Item 10*) | 0,73 | Item 28 | 0,73 |
| Item 11 | 0,73 | Item 29*) | 0,72 | Item 11 | 0,74 | Item 29 | 0,73 |
| Item 12 | 0,72 | Item 30*) | 0,72 | Item 12 | 0,74 | Item 30 | 0,73 |
| Item 13 | 0,72 | Item 31*) | 0,74 | Item 13 | 0,74 | Item 31 | 0,73 |
| Item 14 | 0,72 | Item 32*) | 0,72 | Item 14 | 0,75 | Item 32 | 0,74 |
| Item 15 | 0,73 | Item 33*) | 0,72 | Item 15 | 0,73 | Item 33 | 0,73 |
| Item 16 | 0,72 | Item 34*) | 0,73 | Item 16 | 0,73 | Item 34 | 0,73 |
| Item 17 | 0,72 | Item 35*) | 0,73 | Item 17 | 0,73 | Item 35 | 0,73 |
| Item 18 | 0,72 | Item 36*) | 0,73 | Item 18 | 0,72 | Item 36 | 0,77 |

### *3.2. Construct Validation*

After going through the expert validation process, the PhysTeBloHOTS was packed into two question packages, i.e. package A and package B, each containing 36 questions with 10 anchor items that would be used in the construct validation stage.

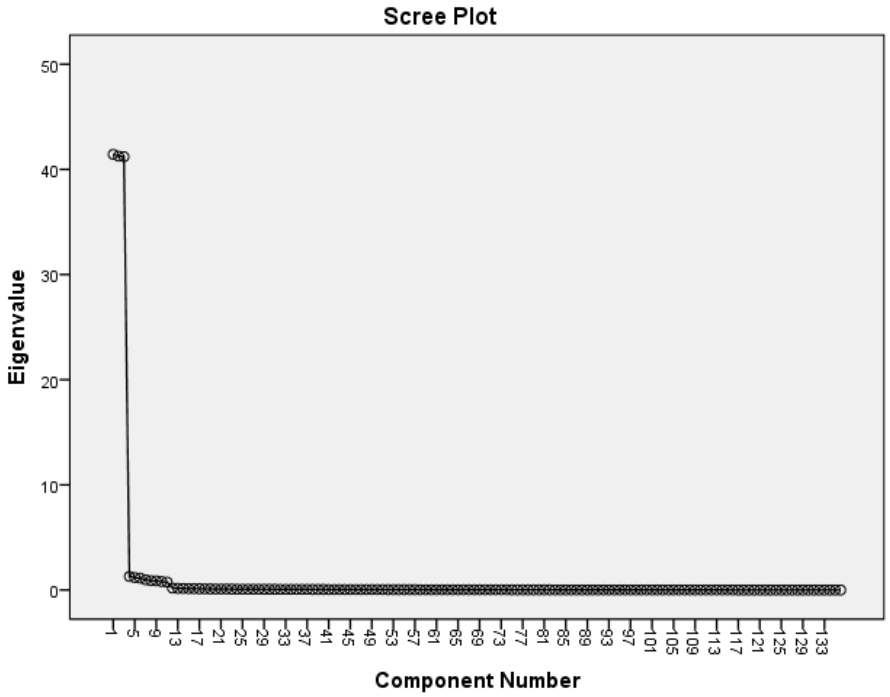


**Figure 1.** Scree Plot of Component Number

**Table 2**. Total Variance Explained

**Total Variance Explained**

| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | | Rotation Sums of Squared Loadings | | |
|---|---|---|---|---|---|---|---|---|---|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 41.425 | 30.460 | 30.460 | 41.425 | 30.460 | 30.460 | 41.411 | 30.450 | 30.450 |
| 2 | 41.252 | 30.332 | 60.792 | 41.252 | 30.332 | 60.792 | 41.252 | 30.332 | 60.782 |
| 3 | 41.200 | 30.294 | 91.086 | 41.200 | 30.294 | 91.086 | 41.214 | 30.305 | 91.086 |
| 4 | 1.293 | .951 | 92.037 | | | | | | |
| 5 | 1.186 | .872 | 92.909 | | | | | | |

Base on Tabel 2, the results of factor analysis using EFA, PhysTeBloHOTS is formed 3 factors. The percentage of variance explained is 88,057%. The first factor explains the variance of 43,234%, this indicates that the instrument meets the dimensional assumptions. This is in accordance with Reckase (Smits, Cujpers & van Straten, 2011) that if the output of factor analysis produced by the first factor is able to explain the variance of more than 20%, then the dimensional assumption has been fulfilled. Grouping items into factors formed based on the value of loading factors from each item. The critical value of loading factors with a sample size of 300 is 0.35. Based on the guideline for determining the factor loading significance presented by SOLO Power Analysis, BMDP Statistical Software, Inc. 1993 [16].

Table 3. Guidelines for Identifying Factor Loading Values Based on Sample Size

| **The factor loading value is considered significant** | **Required of Sample size** |
|---|---|
| 0,3 | 350 |
| 0,35 | 250 |
| 0,4 | 200 |
| 0,45 | 150 |
| 0,5 | 120 |
| 0,55 | 100 |
| 0,6 | 85 |
| 0,65 | 70 |
| 0,7 | 60 |
| 0,75 | 50 |

Table 4. Factor Grouping

| | **Factors** | | |
|---|---|---|---|
| | **1** | **2** | **3** |
| **Number of Items** | 9A, 10A, 11A, 12A, 13A, 14A, 15A, 16A, 17A, 18A, 19A, 20A, 21A, 22A, 18B, 19B, 20B, 21B, 22B, 23B, 24B, 25B, 26B, 27B, 28B. | 1B, 2B, 3B, 4B, 5B, 6B, 7B, 8B, 9B, 10B, 11B, 12B, 13B, 14B, 23A, 24A, 25A, 26A, 27A, 28A, 29A, 30A. | 1A, 2A, 4A, 6A, 8A, 15B, 31A, 32A, 33A, 34A, 35A, 36A, 37A, 38A, 39A, 40A, 16B, 17B, 29B, 30B, 31B, 32B. |
| **Name of Factor** | Analyze | Evaluate | Create |

CFA is similar to EFA, but it is quitely different. At CFA, first the researcher must specify the number of factors and which factor of each variable before analyze it. Researchers can get a better comprehension of the measurement quality if they combined construct validity with CFA (Hair et al., 2006). The result of confirmatory factor analysis of Bloomian HOTS Physics Test shows in Figure 2.

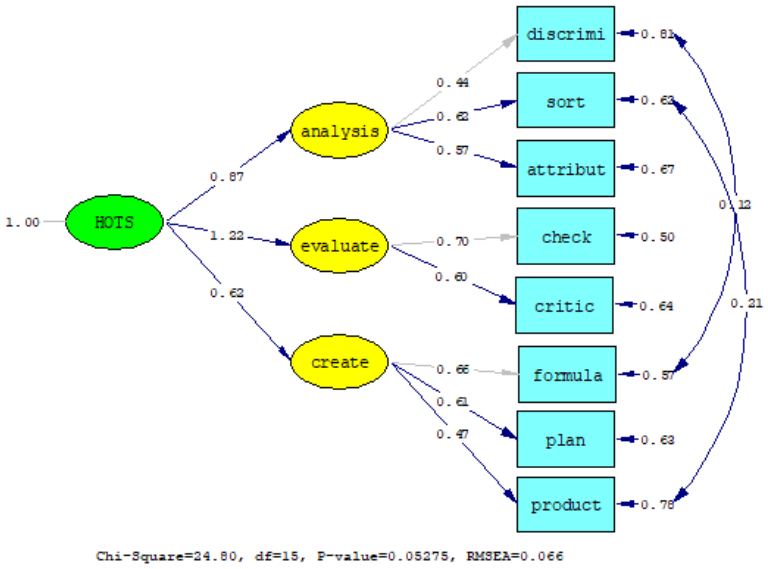


**Figure 2.** Model of the PhysTeBloHOTS

The first step to analyzing CFA was directed to assessing the goodness of fit (GOF) between the data and the model. In Figure 1, the Chi Square obtained = 24.80, df = 15, p. value = 0.05275 and RMSEA = 0,066. It is indicates that the model has a fit. The results and criteria of the model-fit can be seen in the Table 5.

.

**Table 5.** The Results of Model Fit

| GOF Measure | Degree Target of Fit | Estimation Results | Degree of Fit |
|---|---|---|---|
| $\chi^2$ | $\chi^2 \geq 24.99579$ | 24.80 | Marginal fit |
| *p* value | $p \geq 0.05$ | 0.05275 | Good fit |
| RMSEA | RMSEA $\leq 0.08$ | 0.066 | Good fit |
| NFI | NFI $\geq 0.90$ | 0.94 | Good fit |
| CFI | CFI $\geq 0.92$ | 0.97 | Good fit |
| GFI | GFI $\geq 0.90$ | 0.96 | Good fit |

Tabel 5 shows that the fit between the data and the model was in general good, the next step was looking the construct validity of the model [17]. Construct validity refers to a theoretical view to explain constructs in a psychological theory [18]. The results of the second order CFA for the value of t-value and Standarized Loading Factor (SLF) to evaluating the construct are presented in Table 6

.

**Table 6.** The Results of Second Order CFA

| Variabel | Aspect | Sub aspect | Second Order CFA | |
|---|---|---|---|---|
| | | | t-value | SLF |
| HOTS | Analysis | Discriminate | - | 0.44 |
| | | Sort | 4.37 | 0.62 |
| | | Attribution | 4.21 | 0.57 |
| | Evaluate | Check | - | 0.70 |
| | | Criticize | 6.32 | 0.60 |
| | Create | Formulate | - | 0.66 |
| | | Plan | 4.77 | 0.61 |
| | | Product | 4.14 | 0.47 |

Table 6 shows that t-value of all sub aspect was bigger than 1.96. this indicates that all sub aspect are significant in supporting the construct of HOTS. The Standarized Loading Factor (SLF) value also shows that all sub aspect meet the criteria of good construct validity.

## 4. Conclusion

PhysTeBloHOTS is valid according to the content and construct validity. The PhysTeBloHOTS items are valid according to the expert judgments with Aiken's V index ranged from 0.83 to 0.97, respectively. The result of EFA shows that PhysTeBloHOTS consist of three aspects: analysis, evaluate, and create. The result of CFA shows that three aspect of PhysTeBloHOTS are significant in supporting the construct.

## References

[1] Jensen, J. L.; McDaniel, M. A.; Woodard, S. M., &Kummer, T. A. 2014. Teaching to the Test or Testing to Teach: Exams Requiring Higher Order Thinking Skills Encourage Greater Conceptual Understanding. *Journal EducPsychol Rev*, 1-24.

[2] Kusairi, S. 2012. Analisis Asesmen Formatif Fisika SMA Berbantuan komputer. Jurnal Penelitian dan Evaluasi Pendidikan, 16/Edisi Dies Natalis ke-48 UNY, 68-82.

[3] Anderson, L.W., &Krathwohl, D.R. 2001. *A Taxonomy of Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

[4] Madhuri, G.V., Kantamreddi, V.S.S.N., &Goteti, L.N.S.P. 2013. Promoting higher order thinking skills using inquiry-based learning. *European Journal of Engineering Education*, 37/2, 117-123.

[5] Istiyono, E. 2013. *Pengembangan Instrument Untuk Mengukur Kemampuan Berpikir Tingkat Tinggi Dalam Mata Pelajaran Fisika Di SMA*.Disertasi Doktor, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta.

[6] Cullinane, Alison, & Liston, M. 2011. *Two-tier Multiple Choice Question: An Alternative Methodof Formatif Assessment for FirstYear Undergraduate BiologyStudents*. Limerick: National Center for Excellence In Mathematics and Education Science Teaching and Learning (NCE-MSTL).Faster, Smarter, And More Secure Approach To Pre-Employment Testing [Versielektronik]. *Journal of Business Psychology*, 26, 227–232.

[7] Winarti.; Cari.; Suparmi.; Sunarno, W. &Istiyono, E. 2017. Development of two tier test to assess conceptual understanding in heat and temperature. *Journal of Physics: Conf. Series,* 1-5.

[8] Widhiarso, W. 2010. *AplikasiTeoriRespon Item untuk Pemodelan Respons Menipu pada Skala Kepribadian*. Laporan Hasil Penelitian UGM.

[9] Istiyono, E., Mardapi, D., & Suparno. 2014. Pengembangan Tes Kemampuan Berpikir Tingkat Tinggi Fisika (PhysTHOTS) Peserta Didik SMA. Jurnal Penelitian dan Evaluasi Pendidikan, 18/1, 1-12.

[10] Hassan, S.R., Rosli, R., &Zakaria, E. 2016. The Use of i-Think Map and Questioning to Promote Higher-Order Thinking Skills in Mathematics. *Creative Education*, 7, 1069-1078.

[11] Hambleton, R.K., Swaminathan, H., & Roger H.J. 1991. *Fundamentals of Item Response Theory*. California: SAGE Publications, Inc.

[12] Kantrowitz, T.M., Dawson, C.R., &Fetzer M.S. 2011. Computer adaptive testing (CAT): A Faster, Smarter, And More Secure Approach To Pre-Employment Testing [Versielektronik]. *Journal of Business Psychology*, 26, 227–232.

[13] Veeravag, J., Muthusamy, C &Marimuthu, R. 2010. Using Bloom's Taxonomy To Gauge Students' Reading Comprehension Performance. *Canadian Social Science*, 6/3, 2010, 205-212.

[14] Mindyarto, B.N.; Rusilowati, A.; Kartono&Sugiyanto. 2011. PengembanganSistem Diagnosis Kognitif Fisika Online Untuk SMP. *JurnalPenelitianPendidikan*, 29/2, 86-95.

[15] Kowsalya, D.N., Lakshmi, V., & Suresh, K.P. 2012. Development and Validation of a Scale to Asses SelfConcept in Mild IntellecctualyDisabled Children. *International Journal Soc. Sci and Education*, 2/4, 669-709.
[16] Hair, J.F, Black, W.C., Babin, B.J., & Anderson, R.E. 2014. Multivariate data analysis (7thed). Edinburg Gate: Pearson Education Limited.
[17] Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. 2006. Multivariate data analysis (7th ed.). Prentice Hall
[18] Wiersma, W. (2000). Research methods in education: an introduction (7th ed.). Boston: Allyn & Bacon