

## Using the Rasch Model for the Affective Assessment of EFL Learners

**Siti Mariam**

Post Graduate School of Universitas Negeri Semarang, Indonesia

**Mursid Saleh**

Semarang State University of Semarang, Indonesia

**Warsono**

Semarang State University of Semarang, Indonesia

**Januarius Mujiyanto**

Semarang State University of Semarang, Indonesia

### Abstract

Affective assessment is one of the components of authentic assessment which requires English teachers to assess the students comprehensively. This study aims to (1) explain the ways in which affective aspects improve students' English achievement, (2) explain the formulation of an affective assessment for EFL learners using the Rasch model. This research and development approach was employed as the basis to develop rating scale model as an instruments measuring the students' affective aspects in EFL classroom. The data were analyzed using the Rasch Model. The validity and reliability, in the small-scale field tryout were item reliability 1.00 and person reliability 0.93. Meanwhile, item validity are 0.90 and person validity 0.87 were used in the large-scale field one. The results of this study indicate that: (1) there are five affective dimensions or variables developed in this study. They are attitude, motivation, interest, self-concept, and personal value. The five dimensions consist of 24 aspects or sub variables and 35 indicators that become the basis of constructing the 120 item instrument, and as inventory rating scale model. It can be inferred that the affective measuring instrument with the 120 items of statement has a model fit with the data. It means that the model is able to estimate population covariance matrix which is not different from the sample covariance matrix so that the estimation result becomes a basis for generalization.

**Keywords:** affective, assessment , EFL learner, English competence, Rasch model

**Cite as:** Siti Mariam, Saleh, M., Warsono, & Mujiyanto, J. (2018). Using the Rasch Model for the Affective Assessment of EFL Learners. *Arab World English Journal*, 9 (2). DOI: <https://dx.doi.org/10.24093/awej/vol9no2.29>

## Introduction

Affective assessment in higher education is a difficult area to write about since everyone has a view and many assumptions which are commonly expressed 'assessment drives the learning. Students are more strategic and mark-oriented than they were. Interestingly, though, assessment is sometimes the last thing that lecturers think about when designing their courses, we tend to think about the curriculum and what should be covered and only when that has been determined do we turn our attention to how we might assess what our students have learned. The implementation of affective education in college or university must be done through a special approach, meaning it can not be equated with an affective education approach for groups of primary and secondary education, because adolescents who are considered 'immature' can still be fully controlled by parents and teachers which is the largest affective factor in their lives (Slavin, 2006). Therefore the application of affective domains in the education process feels easier, whereas for groups of students considered to be adults, the application of affective domains must go through a specific strategy in order to have a real impact on the next journey of life as a citizen. In practice, the impact of affective domains in education is often overlooked by researchers. This is because they assume that affective domains are very difficult to measure and have dependence on other factors such as economic, political, social and cultural factors as well as psychological or personal life factors (Lynch, Baker & Lyons, 2009). On the other hand, the affective domain actually has a very significant influence on the cognitive domain (Krathwohl, Bloom & Masia, 1964, pp.49-50), so the outcome of the affective domain application is very much needed in the learning process. Affective learning domains related to feelings, emotions, or students' responses to their learning experiences. Affective behavior, among others, are shown by the attitude, interest, attention and awareness. However, to implement these three domains or skills in assessment is not easy because of the various factors, including (1) low commitment, (2) insufficient knowledge (3) limited facilities and supporting funds, (4) political will either from central and regional government; and (5) dissemination of less effective information. Affective problems are important to everyone, but their implementation is poor. (Mariam, 2018, p.1). This is because designing the achievement of affective learning goals is not as easy as cognitive and psychomotor learning. The educational unit should design the right learning activities to achieve affective learning objectives. The success of educators in implementing affective domain of learning and the success of learners achieving affective competencies needs to be assessed. Therefore it is necessary to develop a reference for the development of affective domain appraisal tools and their measurement results.

The study aims to develop an affective assessment for English as a Foreign Language (EFL) learners by using the Rasch Model. The findings of the study will surely contribute to English Language Teaching (ELT), mainly in English language evaluation in higher education.

## Literature Review

Affective aspect plays an important role in man's life, mainly in making decision, perception, interaction, communication and intelligence. According to Airasia & Russell (2008) "a second behavior domain is the affective domain. The affective domain involves feelings, attitudes, interests, preferences, values, and emotions. Emotional stability, motivation, trustworthiness, self-control, and personality are all examples of affective characteristics" (Pp.69) Various research

results showed that effectiveness of cognitive achievement occurred in accordance with effectiveness of affective achievement. In general, students who have good academic achievement (cognitive), they also have high learning motivation and positive attitude towards the subject (affective). On the other hand, when they have low achievement, usually their motivations are low besides that their attitude towards the subject is also negative. According to the research results, around 25 percent of variant of learning cognitive achievement is contributed by affective characteristics owned by a student individually in the beginning of learning. Basically, students' learning achievement cannot be seen from cognitive and psychomotoric domain only as practiced today in our education, but also must be seen from affective achievement. The three domains have reciprocal relationship, although the power of the relationships varies from one case to another.

Assessment and giving feedback to learners is one of the eight specified areas of activity, core knowledge and professional values articulated in the Education of National Standards framework for teaching and supporting learning in higher education. Such recognition of the centrality of assessment to the learning process means all who teach and facilitate students learning need to reflect critically on assessment practices in higher education.

The use of the word 'affect' in the course of general conversation is rare, although the use of its derivatives (such as 'affectionate': a disposition to act from a kindly feeling or love towards one another) are more common. In psychology, the term is commonly used in conjunction with cognition (e.g. Clark & Fisk, 1982; Tomkins & Izard, 1966), but not so in educational discourse, where references to 'affective education' or 'affective learning' are infrequent.

Affective domain is difficult to define and measure. There are limited evaluation instruments to measure and assess. There is an unwillingness to give mark in affective domain because it is related to the validity and reliability aspects. It is difficult to determine behavior standard that reflects affective domain and there are less direct consequences that reflect in affective behavior. The reality shows that not many teachers use affective assessment to measure students' English proficiency. There are many reasons why affective assessment should be used by English teachers to assess students' competence comprehensively. From the identification of the problem above, it seems that teachers need a means that can facilitate them to assess their students comprehensively. Up till now, there is no affective assessment model of English subject for students of non-English department. Therefore, a model of affective assessment of English subject for students of non-English department is needed to facilitate English. There is an interaction between language learning and the environmental components in which the students were grown up. Both negative and positive attitudes have a strong impact on the success of language learning teachers in higher education

The researcher proposes that affective dimensions consist of five aspects namely: attitude, motivation, self concept, interest, and personal value.

First, according to Gajalakshmi (2013), attitude is determined by the individual's beliefs about outcomes or attributes of performing the behavior (behavioral beliefs), weighted by

evaluations of those outcomes or attributes. Thus, a person who holds strong beliefs that positively valued outcomes will result from performing the behavior will have a positive attitude toward the behavior. Conversely, a person who holds strong beliefs that negatively valued outcomes will result from the behavior will have a negative attitude. Attitude concept can be viewed from these three dimensions. Each one of these dimensions has different features to bring out language attitude results. Accordingly, the attitude concept has three components i.e. behavioral, cognitive and affective. These three attitudinal aspects are based on the three theoretical approaches of behaviorism, cognitivism, and humanism respectively.

Second, motivation is a key force in understanding adult learning. Motivation is inherent in humans and is the driving force for people to behave in certain ways to achieve objectives. The expectancy x value theory (Feather 1982) states that adult learners are motivated when they see value in doing something which drives them to exert the effort to achieve the expected outcome. Two types of motivation are extrinsic and intrinsic. With extrinsic motivation, learners are motivated to learn to achieve rewards or avoid punitive actions. With intrinsic motivation, learners are motivated to learn because of the personal satisfaction gained from acquiring new knowledge or skill. Knowles (1984) identifies five factors that distinguish andragogy (adult learning) from pedagogy children's learning. These factors were, as people mature, they a) become self-directed learners, b) bring a wealth of experience to learning, c) have a readiness to learn according to their roles in society, d) like to engage in problem-based learning, and e) become more internally motivated to learn. Two models of adult learning were presented: the expectancy-valence model (Vroom, 1964) and the force-field-analysis model (Lewin, 1951). The assumption of the expectancy-valence model was that learners' motivation would increase if they felt that their expectations and values were being met. The basis of the force-field-analysis model was that positive and negative forces exist in the mind-set of learners and the push and pull effect of the forces will affect a learner's motivation to learn about things that require changes in their thinking or practices. In some contexts, it was suggested (Brookfield, 1983) that adults are more focused on learning in non-credit, non-academic, vocational or recreational fields. However, it was also argued that both male and female adults actively participate in professional upgrading programs when they have the opportunity or are sufficiently motivated, and can afford the fees or be sponsored.

Third, the skill-development model highlights that academic achievement influences academic self-concept (Marsh, 2006; Marsh et al. 2002; Marsh et al. 1999). Lastly, according to Guay et al (2003) both academic self-concept and academic achievement directly influence each other, that is, they are reciprocal. The debate among researchers concerning whether prior academic self-concept influences academic achievement or prior academic achievement results into subsequent academic self-concept has been considered an egg-chicken question (Marsh et al. 2002).

Forth, interest is often thought of as a process that contributes to learning and achievement. That is, being interested in a topic is a mental resource that enhances learning, which then leads to better performance and achievement (Hidi, 1990). Indeed, research has demonstrated that both

situational and individual interest promote attention, recall, task persistence, and effort (Ainley, Hidi, & Berndorff, 2002; Hidi, 1990; Hidi & Renninger, 2006). From this perspective, then, interest appears to play a very important role in learning and academic achievement. As important as interest is for performance and achievement, however, we believe that interest is critically important in its own right. Indeed, one of the primary goals of college education is to help students discover their true interests and chart a life course based on interests developed and nurtured in college. Thus interest may be viewed as essential with respect to adjustment and happiness in life. Relegating interest to the role of a mediator (i.e. a motivational process that is important only because it influences performance) misses the central importance of interest in our lives. Researchers in positive and health psychology have demonstrated that happiness life satisfactions is an important components of well-being (Lucas, 2007; Sheldon & Lyubomirsky, 2007). Pursuing activities and topics that we find interesting play an important part in determining how fulfilled we are with our lives, and not doing so leaves us with a feeling of unease and discontent (Sheldon & Elliot, 1999). Therefore, we endorse the perspective that interest is an important outcome, and we believe that it is a crucial component of success in academics, sports, or other areas of our lives (Harackiewicz, Durik, & Barron, 2005; Hidi & Harackiewicz, 2000; Maehr, 1989; Nicholls, 1979). Researchers (Elliot & McGregor, 2001; Pintrich, 2000) have found that individuals can pursue these two types of goals in one of two ways: by trying to attain the desired outcomes, such as learning as much as possible mastery-approach and doing better than others performance-approach, or trying to avoid negative outcomes such as not learning the material mastery-avoidance or doing worse than others performance-avoidance. As mentioned earlier, the goals that individuals choose to pursue in achievement settings provide purposes or reasons, for task engagement, and serve to orient their attention and effort while engaged in the activity. Of particular relevance to interest development, a focus on task mastery and skill development may encourage the individual to explore all aspects of the task, master it and develop skills, and experience positive affect (Flum & Kaplan, 2006; Renninger, 2000).

Fifth, personal values have been proposed to have a significant influence on decision-making (e.g. Rokeach, 1973). Specifically, it may be argued that an individual's values may propel him or her to behave in an ethical or unethical manner (Baird & Zelin, 2007). Personal values, although individualistic in nature, are largely influenced by societal and cultural factors and tend to vary across nations (Lan et al., 2009). Research indicates that personal values influence an individual's behavior and attitude, and this can at times conflict with the values held by colleagues or organisations within which they work (e.g. Lan et al., 2009). Personal values may be regarded as deep-seated, pervasive, core-beliefs or guiding principles that transcend specific situations to direct or propel human behavior in decision-making. This belief coincides with Rokeach's (1973) definition of a value described as "an enduring belief that a specific mode of conduct or end-state of existence." (p.5). Rokeach (1973) purports that values are central to an individual's thought processes, and therefore instrumental in the formation of attitudes and the execution of purposive behaviors in many circumstances or issues. It was further suggested that an individual's values are arranged or classified according to a value system which subsequently influences acceptable behavior responses.

## Research Methods

### *Design*

This research and development approach aimed at developing product in the form of instruments of rating scale model to measure students' affective aspects in EFL classroom. The research process was carried out through the stages of (1) pre-development: theoretical review on affective aspects in English language teaching and learning, (2) development process consisting of (a) arrangement of test specification and instrument items, (b) the evaluation by English Language Teaching (ELT) experts and measurement by psychometric experts, (c) analysis of tryout data quantitatively, and (3) presentation, that is to rearrange the instruments after revision so that it was ready to use. The validity and reliability, in the small-scale field tryout were item reliability 1.00 and person reliability 0.93. Meanwhile, the item validity was 0.90 and person validity was 0.87.

### *Subjects*

The tryout subjects were respondents who had the same characteristic or almost same as the respondents in data collection. The tryout subjects were students of Education and Teacher Training Faculty, Walisongo State Islamic University of Semarang, Indonesia. The location of the tryout test was non-English department and faculties. They were the Islamic Education department, the Arabic Education department, the Islamic Education Management department, the Chemistry department, the Physic department, the Biology department, and the Mathematic department. The tryout subjects from each department were chosen using the purposive sampling technique.

### *Research Instruments*

The instruments that were developed or produced in this research were the instruments of affective measurement for non-English department students. They are inventory, self report, or affective scale. The kind of scale developed is summated rating scale or scale of modified Likert model into variety five alternative points of answers in accordance with context of statements. The affective scale becomes the instruments of data collection for try-out product activity in field testing.

### *Data Analysis Techniques with Rasch Model Approach*

The technique of data analysis of the empirical test result was done by descriptive qualitative and quantitative. Techniques of analyzing the results of empirical empirical measurements in the research are derived from such activities, (1) readability testing of English Affective product, (2) test of English Affective product implementation, (3) model suitability or dimensionality, (4) Calibration of test items, (5) Measurement of Students' English Affective (6) practicality and (7) the effectiveness of English Affective products. The empirical analysis employs the Rasch polytomous and dichotomous model. The activity of the data analysis. The measurement result analysis was performed using RSM (Rating Scale Model) measurement model to test the polytomous IRT model. In order to apply the IRT approach, both IRT assumptions i.e. unidimensionality and local independence must be met (Hambleton, Swaminathan & Rogers, 1991). The affective instrument satisfies unidimensionality assumptions based on CFA testing as it proves to measure one factor, namely affective. Regarding with the assumption of local

independence, the participant's response to an affective measuring instrument item is not related or influenced by other items, so that local independence assumptions are also met. Therefore, the application of the IRT model can be done.

The model in the Item Response Theory (IRT) is selected based on the mathematical form of the item characteristic function and the number of parameters involved in the model. Appropriate or fit model with a certain measuring scale, not necessarily fit with another scale. The researcher used polytomous IRT model because this affective measuring instrument is designed using Likert type scale with four answer response options. The applied polytomous model is the rating scale model, an IRT polytomous model that uses a single item parameter. This model is in line with the Rasch Model on the dichotomous IRT model. The model was chosen because this research seeks to develop affective measuring instrument that can be used over and over again, so that sampling invariant is required. For that reason, the researcher used a measurement model approach and used one parameter statement. The analysis is done using Winstep software. This software lets the researcher know the index of the threshold value ( $\delta$ ) of each answer option for each item, commonly called the degree of statement approval statement, or in this case the degree of propensity for statement approval. The calculation of fit statistics statement based on the infit value was also done to see whether the item is good or bad. Furthermore, the testing of differential item functioning (DIF) to see the possibility of a statement item that has a bias response to gender. DIF analysis also uses Winstep program.

## Findings

### *Rasch Model*

According to Linacre (2011), Rasch developed an analytical model of the Item Response Theory (IRT) in the 1960s which is commonly called "one logistic parameter" (IPL). This mathematical model was later popularized by Wright. With raw data in the form of dichotomous data (in right and wrong form) indicating the ability of the students, Rasch formulated this to be a model that connects students and items. A student who is able to do 80% problems correctly has better abilities than those who can only work out 65% of the problems. The data percentage indicates that the raw data obtained is called ordinal data type which is rank rather than linear in nature. Because ordinal data do not have the same interval, they need to be converted into ratio data for the purposes of statistical analysis. So that a person gets a score of 80%, then the odds ratio is 80:20, which is nothing but the data ratio that is more appropriate for the purpose of measurement. Through this ratio data Rasch develops a measurement model that determines the relationship between individual ability level and item difficulty level by using the functionality of the algorithm to produce measurements of the same interval. The result is a new unit called logit (log odds unit) that indicates students' abilities and the problem of the item, so that later from the logit score obtained, it is concluded that the success rate of the students in the work depends on the level of abilities and the difficulty level.

Within the scope of social science, it obtains data in the form of a common source number that can usually be in the forms of attitude and opinions on the statement items or questions in a given instrument. The instrument is designed from variables that have been satisfactorily defined,

then identified relevant constructs. From there the items are made and developed to be able to measure the variables intended. At the same time the choice of answers provided generally follows a scoring pattern followed by the classical test theory (CTT). In the context of the Rasch model, these 'settled' polymers are nothing but measurements whose results depend on the test dependent scoring; while what should be done in quantitative research in social science is objective measurement.

According to Wright & Mok (2004), the concept of objective and fair measurement in the social sciences must have five criterias : (1) Provide a linear measure with the same interval; (2) Conduct a proper estimation process; (3) Find items that are not exact (misfits) or not common (outliers); (4) Overcoming lost data; (5) Generate replicable measurements for independent of the parameters studied.

Of the five conditions earlier, so far it is only the Rasch model that meets the five requirements . In other words, the quality of measurement in the social sciences done with Rasch model will have the same quality as the measurements made in the field of Physics. When viewed further, the log odds unit generated in the Rasch model is a scale with equal intervals and is linear derived from the odds data ratio rather than the raw data scores obtained (1). Therefore the process of estimating one's abilities or the degree of difficulty of the problem will have a more precise estimation value and can be compared because it has the same unit (logit) (2). Since the algorithm used will perform a structured sorting between the respondents of high abilities into low, which simultaneously also sort the problems from the easy to the difficult, the presence of inaccuracies / consistency of answers from respondents (misfit) or patterns outside the habits (outlier) will be easily detected; as well as for the response patterns received by a particular item (3). The ordering of respondents' abilities and difficulty in structured ways also makes the Rasch model can predict when there is missing data (4). The resulting log scale generates a value that depends on the given response pattern, rather than on a predetermined initial score, so that the Rasch model will always result in independent measurement .(5) Analysis by Rasch model resulted in a fit statistics analysis that informed the researchers whether the data obtained were ideally described as having high abilities giving the answer pattern to the item according to the degree of difficulty. The parameters used are infit and outfit from mean square and standardized values. According to the Masters, inlier (inlier sensitive or information weighted fit) is the sensitivity of the response pattern to the target item on the respondent (person) or vice versa, while the outfit (outlier sensitive fit) measures the sensitivity of the response pattern to the item about a certain difficulty level on the respondent or vice versa.

Quantitative research in social science has always faced fundamental criticism in terms of testing its research instruments. The usual quantitative test instrument used in CTT is the reliability index (alpha Cronbach) which only measures the interaction between item and person, how individual quality of item can never be done because of the absence of a measurement index that can be done; the same time to detect an inconsistent respondent's answer is not available. Different from the classical test theory, in Rasch model analysis the items are done to the level of each item. In addition to the item, the model Rasch also simultaneously tests the person (respondent), which

will show a consistent answer pattern that tends to approve (in the attitude instrument) and identify the original answer. Test for instrument research can also be done in the form of dimensionality test, rank scale used or detection of the bias of the items tested. All that can be done because basically the Rasch model meets all the objective measurement requirements.

Data analysis techniques performed at the development and evaluation stage include theoretical and empirical analysis of the English Affective draft. Theoretical test results technique is analyzed descriptively qualitative and quantitative. Qualitative data were analyzed using qualitative descriptive technique. The quantitative data of Delphi results were analyzed using quantitative descriptive techniques. Expert input from the Delphi process is used as a foothold in quantitative and qualitative data analysis techniques. Quantitative descriptive technique using Rasch model. MFRM analysis using Facet 3.70 software help. Quantitative data analysis techniques Rasch model that involves experts as a composite assessor (rater) that is MFRM (Multi Facets Rasch Measurement). The MFRM results are the logit of the check list, the interagency rater and logit of the English Affective product observed. The results of MFRM analysis resulted in a rater agreement on content validity, content reliability, practicality and Affective of English . Content validity includes two aspects: (1) content relevance and (2) the content coverage of an English Affective product. The relevance of the content refers to the suitability of each of the English Affective products with theoretical content. Coverage refers to the scope / breadth of the content aspect of the English Affective product. Assessment of content relevance as well as the coverage of English Affective contents based on expert considerations. Focus Group Discussion( FGD ) results data are processed descriptively qualitative.

### ***Unidimensional Testing Using Rasch Model***

Testing of construct validity is done using factor analysis method that is with Rasch model which is done to test dimensionality as reference to test unidimensionality assumption of measuring instrument. Rasch also gave information on whether the fit model to measure affective uses of this measuring instrument. At this stage, the analysis is done using Winstep 3.71 program 2016 output. To assess whether the measurement model is really fit with the data, please note the value of fit index. The fit index generated from the analysis using the Rasch model method there is a wide range. An index indicating that the model is fit does not guarantee that the model is perfectly fit, or vice versa. So researchers do not just rely on one index fit to test the model. This study uses the criteria of the index fit forms residue between items or statement items.

### ***Unidimensional Analysis and Response Independence***

Rasch models are unidimensional measurement models that require statistically independent responses. The unidimensional measurement model means only one dimension of the individual being the focus, i.e. the ability on the dimension ( $\beta$ ). Testing is not unidimensional means there are other factors than an individual other than  $\beta$  who has responded to the test question. Marais & Andrich (2008) referred to it as trait dependence.

Independent response statistically means the probability of an outcome is independent of other outcomes. A person's response to more than one problem means one's response to a problem

is independent of its response to another problem (Andrich, 1988). Independent statically is not met when the response to an item is affected by its response to the previous item. Marais & Andrich (2008) referred to this as response dependence. In the literature of trait dependence and response dependence are generally not distinguished, both are categorized as local independence. Checking the unidimensionality through Winstep and Facet can be seen from the output (Tennant & Pallant, 2006). For the Winstep program the limit of unidimensionality is measured by the rules Raw variance explained by the following measurement.

- a. Sufficient,  $20\% \leq$  Raw variance explained by measurement  $\leq 40\%$ .
- b. Good,  $40\% <$  Raw variance explained by measurement  $\leq 60\%$ .
- c. Excellent,  $60\% <$  Raw variance explained by measurement  $\leq 100\%$ .

Local independence violation can be done by calculating residual correlation between items. Dependencies are suspected when the relative residual correlation between item is relatively high (Tennant & Gonaghan, 2007; Zenisky, Hambleton & Sireci, 2002). High residual correlations indicate the existence of dependencies between items that can not be explained by the ability parameters / approval levels of people and items. High residual correlation means that in addition to the measured variables there are dimensions or other factors that affect the individual response to a particular item. Relatively high residual correlations may indicate trait dependence or response dependence.

### ***Unidimensional Results of Affective Instrument Measuring***

Result of dimensionality of affective instrument based on scoring result according to Rasch modeling. Based on the Rasch instrument modeling it is of sufficient quality if Raw variance explained by measures yields more than 20% or 0.2. The full results of affective instruments are shown in the table 1.

**Table 1 Results of Unidimensionality identification based on tandardized Residual Variance (in Eigenvalue units)**

No	Standardized Residual Variance	Value		Category
		Empirical	modeled	
1.	Raw variance explained by measures	67.3%	64.4%	Excellent
2.	Raw variance explained by persons	31.0%	29.7%	Good
3.	Raw Variance explained by items	36.3%	34.7%	Good
4.	Raw unexplained variance (total)	32.7%	35.6%	Good

Results processed from Winstep 3.71

The result of identification of dimensional units on affective instruments according to Rasch criteria of the model stated good. Politomous scoring for raw variance explained provide greater empirical value than expected models, which means that the affective instrument gives good results. Raw unexplained variance (total) scoring result is smaller than expected model gives

good result. The result states that the dimension of the affective instrument of scoring only measures one unit of affective dimension.

Unidimension test results with other software are facets. Unidimensional facets have the meaning of a scoring model having unidimensional if the value of variance is explained by Rasch measures is less than 30%. The result of the identification of dimensionality using facets software is presented in table 2. The result stated that the distance distribution of variance value explained by Rasch measures 17.1% to 42.76%, this means the result of the affective instrument fulfills the unidimensionality element. This provides the basis that affective measuring instrument can be used to analyze the student's affective assertions.

**Table 2 Result of Identification Unidimensionality Affective Instrument with Facets**

No	Standardized Residual Variance	Value		Category
1.	Raw-score variance of observations	0.82	100%	Excellent
2.	Variance explained by Rasch measures	0.29	34.95%	Excellent
3.	Variance of residuals	0.54	65.05%	Excellent

Results processed from Facets 3.71

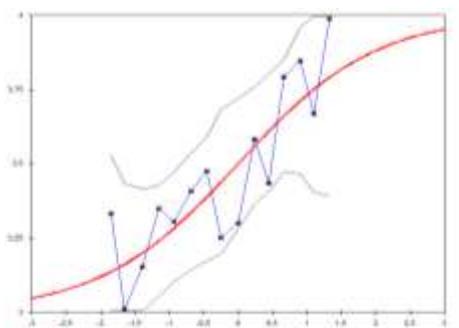
### Assessment Practicality of the Items Statement

The practicality of Affective products in addition to being judged theoretically by experts is also assessed empirically by the user. The practicality indicators of Affective products are easy to understand, easy to implement, easy to do and easy to administer. Assessment of the practicality of Affective product at the stage of initiation at the development and evaluation stage is carried out by the user using the same rating sheet in the practical part. The results of the practicality assessment were analyzed statistically based on the Rasch Model reference framework. Statistical results can be seen in Table 3.

**Table 3: Test Results Initiation Product Practicality of Affective Instruments**

Results of the analysis for Assessment	
Obs Agree.	53.0%
Expected Agree.	55.5%
Separation Ratio	0.33
Average infit	1.00 (MNSQ)
Average outfit	0.99 (MNSQ)
Average logit	-0,23

Description: Obs. Agr.= Observation Agreement, MNSQ=meansquare



**Figure 1 Graph of Trust**

(Results of data research with software Facet 3.70.3)

Based on the results of the assisted statistical analysis of the software of Facet 3.70.3. All assessors stated that Affective products begin conceptually to scoring guidelines based on Rasch Model Reference frame practically with 53.0% deal. The resulting separation ratio is also small 0.33 this means no different statements by the assessors. The average value of infit and outfit is still in the range of 0.5 to 1.5 ie 1.00 and 0.99 with a distribution of 0.77 to 1.25.

## Discussion

This study designs an affective measuring instruments based on the theory proposed by Rogers (1991) because it is considered more capable of describing the effectiveness as a psychological variable. However, it still needs to be tested to prove that the affective theory has a valid construct. The affective measurement test process gives the calculation results of the validity of the measuring instruments and the unidimensionality test based on the Confirmatory Factor Analysis (CFA) method which confirms that the construct of the affective measuring instrument proposed by Rogers (1991) valid. This research also performed testing using first order and second order CFA method to test unidimensionality assumption. The firstorder CFA test towards 120 items fitaffective statements gives a fit result on RMSEA, but chi-square is not fit. This can be considered fit considering that chi -square has sensitivity to the number of samples and data normality. But generally the chi-square index is more believed so that the researcher decided again to carry out the CFA first order test by eliminating the statement items that contribute small. Elimination resulted in a total of 120 affective questions items that carried out the CFA first order test. The 120 items of the question are also tested by the five CFA second order factors. The results of the first order test and the five factors of second order CFA indicate that the model used fit with the data so it can be ascertained that the statement items designed in this study fit to measure affective of English competence.

## Conclusion

There are five affective aspects that are developed in this study, namely attitude, motivation, interest, self-concept, and personal value. Attitude dimension has four aspects, namely attitude towards English subject, attitude towards learning English, attitude towards student's self, and attitude towards who are different from the student. Attitude aspects have five indicators, namely attitude of curiosity, attitude of critical thinking, attitude of honesty, attitude of carefulness, and

attitude of flexibility. Motivation dimension has six aspects, namely desire and eagerness to succeed, urging and need of learning, expectation and aspiration in the future, appreciation in learning English, interesting activities in learning English, and conducive learning environment so that it makes a student enable to learn English well. Motivation aspects have nine indicators, namely: strong willing to learn English, preserving against difficulties in English learning, the amount of time available to learn English, prefer to work autonomously, high willingness to follow English lessons, be able to defend his or her opinions, working hard on English assignment, participate to be the best in learning English, and like to look for and solve problems in learning. Interest dimension has four aspects, namely having glad feeling in learning English, students' involvement in English learning, attracting on learning English activities, and having a stable tendency to pay attention in learning English. Interest aspects have seven indicators, namely receive English lessons happily, learning English continuously, not to be forced to learn English, students' activeness in learning English, content of English lessons in accordance with students' need, following teacher's explanation, and doing homework and assignment. Self-concept dimension has two aspects, namely students' knowledge on English language and students' expectation on ideal English learning. While self-concept aspects have three indicators, namely students' point of view related to their English competence, students' point of view related to the advantages of English competence and students' active role in English learning. Personal value dimension has eight aspects, namely self-expectation, orientation towards life, working ability, self-actualization, self-integrity, self-confidence, self-direction, and intellectual thinking. Meanwhile personal value aspects have ten indicators, namely having English competence, the meaning of life, hard-working, having competence in working, wisdom, a wide perspective, honesty, helpful, brave to try, and being responsible. Thus, 35 indicators from five dimensions of affective become the basis of instruments item arrangement.

The formulation of affective instruments as the result of this research and development is in the shape of inventory or self-report rating scale model or Affective Scale. This affective scale consists of 120 statement items that covering the five of affective dimensions, namely attitude, motivation, interest, self-concept, and personal value. The affective scale has the validity and reliability, in the small-scale field tryout were item reliability 1.00 and person reliability 0.93. Meanwhile item validity 0.90 and person validity 0.87. The instrument validity is significant as it shows, an infit mean square value for the affective measuring instrument was 0.97. For that reason, it can be said that the affective measuring instruments with 120 items of statement has a model fit with the data.

The affective scale has the result that shows overall measurement model fit or the model is suitable with the data. It means that the model that hypothesized or proposed can estimate covariant matrix of population that is not different from covariant matrix of sample. It means that the estimation result that is obtained from the sample data can be the basis for making generalization.

### **Suggestions**

Based on the research conclusions above, it can be put forward several suggestions related to product benefits, namely:

The affective scale that becomes the research product is designed to be able to be used for measuring students affective aspects in learning English. This affective scale has two benefits, either for enhancing students English achievement or overcoming students' difficulties in learning English. So that is the reason why English teachers should design authentic assessment that involve the three domain of learning, namely cognitive, psychomotoric, and affective point of view.

Although this affective scale produced in this study had been tried out to non-English department students of Walisongo State Islamic University of Semarang, but indeed, the affective scale can be applied to all of non English department students to others university. Because this instruments are arranged based on English as a Foreign Language (EFL) that is implemented by all over higher education in Indonesia.

#### **About Authors :**

**Siti Mariam** is currently a lecturer in English Language Education Department of Walisongo State Islamic University of Semarang..Her research interests include ELT Method and Language Learning Evaluation.Currently, she is pursuing her Ph.D in Postgraduate School of Semarang State University of Indonesia. .Her ORCID ID is 0000-0002-3639-1301

**Mursid Saleh** a Professor in English Language Department of Languages and Art Faculty of Semarang State University of Indonesia, He earned his master's degree and Ph.D from Macquarie University of Sydney, Australia.

**Warsono** a Professor in English Language Department of Languages and Art Faculty of Semarang State University of Indonesia.

**Januarius Mujiyanto** a Professor in English Language Department of Languages and Art Faculty of Semarang State University of Indonesia.

#### **References**

- Ainley, M, Hidi, S., & Berndorff, D. (2002). Interest, Learning, and the Psychological Processes that Mediate Their Relationship. *Journal of Educational Psychology*,94, 545-561.
- Airasia, P. W & Michael K Russell. (2008). *Classroom Assessment Concept and Applications*. New York: McGraw-Hill Higher Education.
- Andrich, D (1982). *An Extension for the Rasch Model for Ratings Providing both Location and Dispersion Parameters*.*Psychometrika*,47 (1), 105-113
- Andrich, D. (2004). Controversy and the Rasch Model: A Characteristic of Incompatible Paradigma ? *Medical Care*, 442 (1), 1-7
- Bloom, B.S.(1956). *Taxonomy of Objective: Cognitive Domain*. New York: David McKay.
- Bloom, B.S., Madaus, G.F., & Hastings, J.T( 1981). *Evaluation to Improve Learning*. New York: McGraw-Hill Book Co.
- Elliot, A.J, & McGregor,H. (2001). A 2 X 2 Achiement Goal Framework. *Journal of Personality and Social Psychology*, 80, 501-519

- Guay, F., Marsh, H.W., & Boivin, M. (2003). Academic Self Concept and Academic Achievement: Developmental Perspective on Their Causal Ordering. *Journal of Educational Psychology*, 95, 124-136. <http://eric.ed.gov/?id=EJ662406>
- Hambleton, R. K., Swaminathan, H., Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. Newbury Park : Sage Pub.
- Harackiewicz, J.M., Durik, A. M., & Barron, K. E. (2005). Multiple Goals, Optimal Motivation, and the Development of Interest. In Laham, S.M., Forgas, J.P & Williams, K.D (Eds), *Social Motivation, Conscious and Unconscious Process* (pp.21-39). New York, NY, USA: Cambridge University Press.
- Hidi, S., & Renninger, K.A (2006). The Four-Phase Model of Interest Development. *Educational Psychologist*, 41, 111-127
- Hidi, S., & Harackiewicz, J.M. (2000). Motivating the Academically Unmotivated: A Critical Issue for the 21st Century. *Review of Educational Research*, 70, 151-179.
- Krathwohl, D.R; Bloom; & Masia, B.S (1973). *Taxonomy of Educational Objectives: Book II Affective Domain*. London: Longman Group.
- Linacre, J.M. (2011). Rasch Measures and Unidimensionality. *Rasch Measurement Transactions*. 24 (4), 1310
- Lynch, Baker & Lyons. (2009). *Affective Equality : Love, Care and Injustice*. New York: Palgrave Macmillan.
- Lucas, R.E. (2007). Personality and the Pursuit of Happiness. *Social and Personality Psychology Compass*, 1, 168-182.
- Maehr, M. (1989). Thought about Motivation. In Ames, C & Ames, R. (Eds), *Research on Motivation in Education: Goals and Cognitions* .(3), pp.229-315. New York: Academic Press.
- Mariam, S. (2018). Developing an Affective Assessment Model of English Competence for University Students of Non-English Department. *Unpublished Dissertation*. Semarang : State University of Semarang, Indonesia.
- Marsh, H.W., & Craven, R.G. (2006). Reciprocal Effects of Self-Concept and Performance from a Multidimensional Perspective: Beyond Seductive Pleasure and Unidimensional Perspective . *Perspectives on Psychological Science*, 1, 133-163. <http://pps.sagepub.com/content/1/2/133.short>.
- Nicholls, J.G. (1984). Achievement Motivation: Conceptions of Ability, Subjective Experience, Task Choice, and Performance. *Psychological Review*, 91, 328-346
- Pallant, J.F. & Tennant, A. (2007). An Introduction to the Rasch Measurement Model : An Example Using the Hospital Anxiety and Depression Scale ( HADS), *British Journal of Clinical Psychology*, 46 (1), 1-18.
- Pintrich, P.R. (2000). A Motivational Science Perspective on the Role of Student Motivation in Learning and Teaching Context. *Journal of Educational Psychology*, 95, 667-686.
- Renninger, K.A. (2000). Individual Interest and Its Implications for Understanding Intrinsic Motivation. In Sansone, C & Harackiewicz, J.M. (Eds), *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance* (pp.373-404). San Diego, CA: Academic Press, Inc.

- Sheldon, K., & Elliot, A.H. (1999). Goal Striving, Need Satisfaction, and Longitudinal Well Being: The Self-Concordance Model. *Journal of Personality and Social Psychology*, 76, 482-497.
- Sheldon, K.M., & Lyubomirsky, S. (2007). Is it Possible to Become Happier? *Social and Personality Compass*, 1, 129-145.
- Slavin, R.E. (2006). *Educational Research in an Age of Accountability*. USA : Pearson.
- Wright, B.D., & Mok, M.M. (2004). An Overview of the Family of Rasch Measurement Models. Introduction to Rasch Measurement. *Journal of Applied Measurement*. pp.1-24.