Research and Evaluation in Education Journal
Volume 1, Number 1, June 2015 (45-54)
Available online at: http://journal.uny.ac.id/index.php/reid

FACTOR ANALYSIS TO IDENTIFY THE DIMENSION OF TEST OF ENGLISH PROFICIENCY (TOEP) IN THE LISTENING SECTION

[1)]Heri Retnawati; [2)]Sudji Munadi; [3)]Yosa Abduh Al-Zuhdy
[1)2)3)]Yogyakarta State University, Indonesia
[1)]retnawati.heriuny1@gmail.com; [2)]sudji.munadi@gmail.com; [3)]alzuhdy@gmail.com

Abstract

This research is aimed at identifying the amount of the ability dimensions contained in the Test of English Proficiency (TOEP), particularly in the listening section. This study is an explorative descriptive quantitative research. The data of the study are the responses of the TOEP participants in the whole Indonesia in 2010, in some TOEP components. The participants are grade IX students of senior high schools. The collected data were gained from the data documentation of the Diretorate of Senior High Schools Founding of the National Education Ministry, Jakarta. The data analysis for identifying the dimension was done by implementing exploratory and confirmatory factor analysis. The exploratory factor analysis was done by using SPSS computer program. The result of the research shows that of the seven sets of listening which were analyzed, all of them contain listening dominant dimension if they are analyzed using graphics method, explainable variance, and Eigen value ratio.

Keywords: *dimension, TOEP, listening*

Introduction

The development of science, technology, and communication is getting rapid. The develpoment can become a global world challenge to answer, one of which is by preparing human resources who are able to communicate in the international world. One of the competences which is needed in this case is English competence. In Indonesia, an English test which is consiered to be standardized is Test of English Proficiency (TOEP), which has been calibrated and proved to be able to predict the participants' competence in International English Language Testing System (IELTS) or Test of English as a Foreign Language (TOEFL). TOEP measures listening and reading ability.

To be able to measure English competence, all this time, test method is employed. The result of the test is then analyzed, using classical test theory assumptions or item response theory. One of the assumptions to analyze with the approach of the two theories is a test to measure one competence only, known as unidimension.

In designing, assembling, and analyzing test items, the theory approach which is employed is unidimensional test approach, which only measures one single dimension. In this item response theory, there are assumptions to be fulfilled, namely local independence and unidimension (Hambleton, Swaminathan and Rogers, 1991; Hulin et al., 1983). Local independence occurs when the factors influencing achievement becomes constant, so that the subject responses towards the pair of any items will be statistically independent to each other. Unidimension means that each item of the test measures only one ability. The assumption of unidimension can only be shown if the test contains only one dominant component which measures the achievement of a subject.

In the real fact, unidimensional assumption is hard to be fulfilled. This is in line with the opinion that most educational and psychological tests in some levels are multidimensional (Bolt and Lall, 2003; Ackerman et al., 2003). Unidimensional analysis on the data which reality is multidimension will cause the occurrence of sistematic mistakes in the test administration. As a result, the information which is gained will be misleading and inflict detriment for the participants of the test.

The existence of multidimensional contents in the test components which are analyzed by using unidimensional model causes inaccurate ability estimation and gives misleading information. Related to this, a research about the amount of the dimension contents of TOEP sets, especially listening section, which is able to be utilized as a prerequirement using item response theory approach to do test set further analysis.

In the test, the presentation of the content is a vital thing in the test validity. Through an evaluation on the test which can be executed by a subject-mater expert (SME), the items that form a test and its relevance towards the planned domain can be revealed (Sireci and Geisinger, 1995). The result of the evaluation will show the dimensions which are consistent with the content structure, or dimensions which are not consistent with the content structure. This result also underlies a multidimensional scaling (MDS).

Multidimensional scaling positions the items in a space in a certain coordinate location. This space is determined by certain dimensions as the axis. The relative distance between the item pairs reflects the differences of the items (Bolt, 2001). The nearer an item towards others/its pair is, the bigger the characteristics similarity between the two items is. Based on the closeness of the distance or the similarity of the characteristics, the items can be categorized according to their substance. The analysis is known as a hierarchical cluster validity analysis (Sireci and Geisinger, 1995).

The test in education and psychology which measures latent variables is multidimensional. If the item analysis uses unidimensional approach, thus, inaccurate result of ability measurement is produced

(Wang, Chen, and Cheng, 2004). This happens because unidimensional approach ignores the correlation between latent competences. Multidimensional measurement approach gives attention to the relation between the latent competences which causes the increasing of measurement accuracy. Another advantage of the item response theory is proposed by De la Tore and Patz (2005) that the analysis with this approach gives additional information which increases the accuracy of item parameter estimation. In this situation, unidimension is a case of multidimension, that is, when the inter-latent-variable correlation is equal to zero. There are two types of dimensional structure, namely double-dominant type with inter-dimension correlation and one dominant dimension with some minor dimensions (Kirisci, Hsu and Yu, 2001). In line with this, Wang, Chen and Cheng (2004) state that there are two kinds of multidimension, namely inter-item multidimension and within-item multidimension.

In a research, instrument which involves a lot of items is usually employed. In order to understand such data, factor analysis is usually used. Factor analysis is used to reduce the data by finding the inter-variable relationship which are independent to each other (Stapleton, 1997), which then are assembled in less variables to find out the latent dimensional structure (Anonymous, 2001; Garson, 2006), which is known as factor. This factor is a new variable, and also called latent variable, construct variable, and has being directly unobservable as its characteristics. Factor analysis can be done in two ways, namely exploratory factor analysis and confirmatory factor analysis.

The basis of both exploratory and confirmatory factor analysis is reducing the large amount of variable. For instance, the initial variable is x_1, …, x_q, a latent factor compilation that will be found is ξ_1, …, ξ_n (with $q > n$). An observable variable depends on the linear combination of latent factor ξ_1 which is asserted by

$$X_i = \lambda_{i1}\xi_1 + \lambda_{i2}\xi_2 + ... + \lambda_{in}\xi_n + \delta_i \quad .....(1)$$

With δ_i (measurement error) is a unique part from x_i which is assumed not to be correlated with ξ_1, ξ_2, ..., ξ_n. For $i \neq j$, so $\delta_I \neq \delta_j$. The unique part comprises special factor s_i and a random measurement error e_i. In the confirmatory factor analysis, the amount of latent variable ξ is less compared to the exploratory factor analysis.

In the factor analysis, there is a squared factor loading. This squared factor loading indicates the amount of variants in the observed variables which can be explained by the factors (Van de Geer, 1971). The observed variable which can be explained by factor is usually reflected in the form of relative percentage towards the variants total from the whole observed variables.

Exploratory factor analysis is a technique for detecting and accessing latent resources from variance or covariance in a measurement (Joreskog & Sorbom, 1993). Exploratory factor analysis is defined as exploring empirical data to find out and detect the characteristics and inter-variable relationship without determining the model in the data. In this analysis, the researchers did not have a priori theory to arrange a hypothesis (Stapleton, 1997). Considering its exploring characteristics, the result of the analysis of this exploratory factor is weak. The result of the analysis, which explains only inter-variable relationship, is not also based on the applied theory. The result of the analysis only depends on the empirical data, and if the observed variables are in the large amount, the result of the analysis will be difficult to be interpreted (Stapleton, 1997). Factor analysis is usually strongly related to the questions about validity (Nunally, 1978). When identified factors are linked, exploratory factor analysis answers the question about construct validity, if a score measures what is supposed to be measured.

In exploratory factor analysis, factor analysis is aimed at explaining variance in the observed variable which can be explained by latent factor. Then, in order to interpret the

result of exploratory factor analysis, rotation is conducted. There are two kinds of rotation which can be employed to interpret factor, namely Varimax, Quartimax, Equamax rotation (which is Orthogonal) and Direct Oblimin, Promax rotation (which is non-orthogonal or Oblique). The result will show the rotated loading factor matrix, which then is named based on the dominant items in a particular factor (Wells & Purwono, 2009).

- $\frac{\lambda_1}{\sum \lambda_i} > 20\%$
- $\lambda_1 / \lambda_2 > 5/1 \ \ or \ \ 4/1$ .....(1)

In order to know the English competence, language communicative competence needs to be defined. Various communicative competence models which are proposed by language experts basically share the same concept which includes four main competence, namely grammatical/linguistic competence, strategic competence, sociocultural/sociolinguistic competence, and discourse competence (Bachman, 1990; Bachman & Palmer, 1996; Savignon, 1997).

The second model is a modification result of the first model which is developed by language experts (Bachman, 1990; Bachman & Palmer, 1996) based on the results of the research in the field of language evaluation. Therefore, this second model may be more appropriate to be the conceptual basis of the evaluation system development in language field. According to Bachman & Palmer, the term communicative competence is equalized with the term language ability as a construct that is supposed to be measured by a language test. Language ability covers two components: knowledge on language and strategic competence (or also known as metacognitive strategy). A language user needs the combination of these two competences to be able to produce or interprete a discourse, both in doing the language test and in using the language in real life. The development of TOEP question items refers to the taxonomy of language ability. In this case, Munby (1981) has identified micro-language ability which is resulted in the point that micro-language ability can be divided into observing, speaking, reading, and writing ability. Thus, this research aims to identify the amount of ability dimensions contained in Test of English Proficiency (TOEP), particularly in the listening section.

Method

This research employed quantitative method with explorative descriptive approach, because multidimensional loading of TOEP elements would be identified in this research. This study was conducted in Yogyakarta Special Region, from April 2013 until November 2013.

The data in this study were the responses of TOEP participants in the whole Indonesia in 2007-2010, at some TOEP sets. The participants were grade IX senior high schools students. The data collection in this study referred to the data of the documentation result from the Directorate of Senior High Schools Founding of the National Education Ministry, Jakarta. The data sources were in the form of students' answer sheets which had been documented in the form of computer data.

The data which had been gained were in the form of TOEP participants' responses and the sets of TOEP employing exploratory factor analysis, then the amount of dimensional loading in the test sets were estimated. The quantity of the dimension would be known by counting the amount of the factors contained in the test sets in the factor analysis, both exploratory and confirmatory. This exploratory factor analysis was done using a computer program, Statistical Package for Social Sciences (SPSS).

Findings and Discussion

Findings

In this research, the dimensionality of TOEP sets was proven through three ways, namely graphics, percentage of the explainable variance, and the ratio of the first and the second Eigen value. Analysis

was done by employing SPSS to find out the Eigen value, then, with the help of Microsoft Excel to draw the graphics, the percentage of the explainable variance and the ratio of the first and second Eigen value were counted. Each result is presented below.

Dimensional Validation with Graphics

In the TOEP 1A set, for the listening section, there is one part of the graphics which is steep. This indicates that the 1A listening section measures one main dimension, that is, listening ability. More complete result is presented in Figure 1.

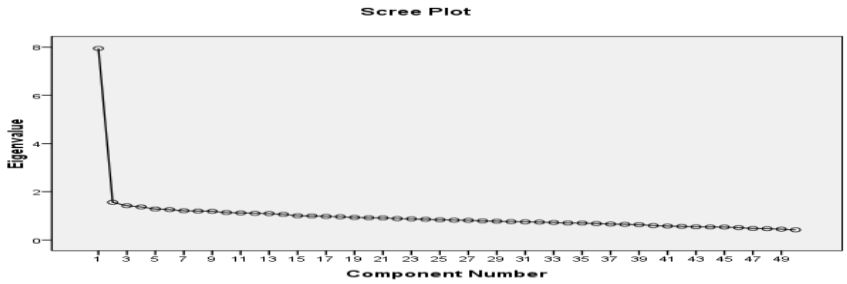
TOEP 1A Listening	

Figure 1. Scree plot for the TOEP 1A set

The same thing happens to the TOEP 2A set. For the listening section, there is one part of the graphics which is steep. It indicates that the listening section 2A measures one main dimension, that is, listening 2A set is proven to measure listening empirically. The result is presented in Figure 2.

TOEP 2A Listening	

Figure 2. Scree plot for the TOEP 2A set

In the TOEP 3B set, there is also a steep part in the graphics. It indicates that the listening 2B set measures one main dimension. This result shows that listening section 2B is proven empirically to measure listening ability. This different dimension is reading dimension. The more complete result is presented in Figure 3.

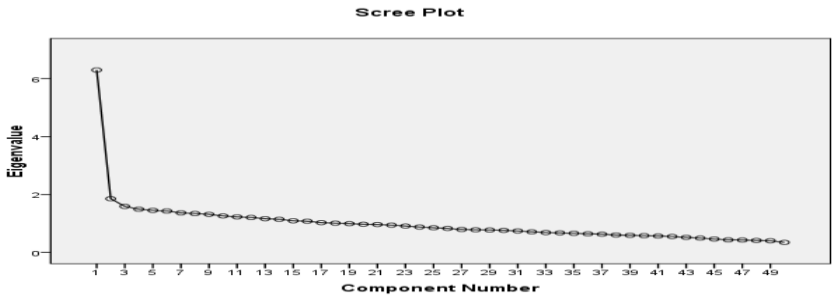
TOEP 2B Listening	

Figure 3. Scree plot for TOEP 2B set

In the TOEP 3A set, for the listening section, there is also a steep part in the graphics. It is an indication that the listening 3A set measures one main dimension, that is, this set is proven empirically to measure the listening ability. The result is presented in Figure 4.

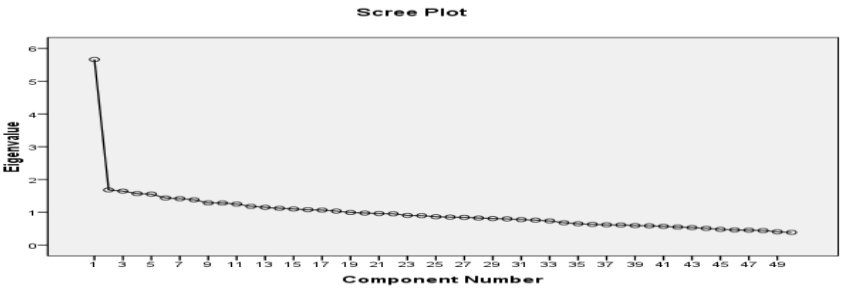
TOEP 3A Listening	

Figure 4. Scree Plot for TOEP 3B set

In the TOEP 3B set for the listening section, there is a steep part in the graphics. It indicates that the listening section measures one main dimension, that is, the listening 3B set is proven empirically to measure listening ability. The result is presented in Figure 5.

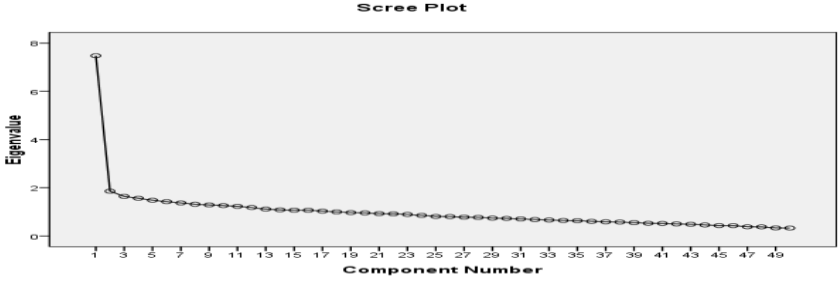
TOEP 3B Listening	

Figure 5. Scree Plot for TOEP 3B set

The graphics in the TOEP 4A for listening set shows the same thing. There is also one part of the graphics which is steep. It also indicates that the listening 4A set

measures one main dimension, that is, listening 4A set is proven empirically to measure listening ability. The complete result is shown in Figure 6.

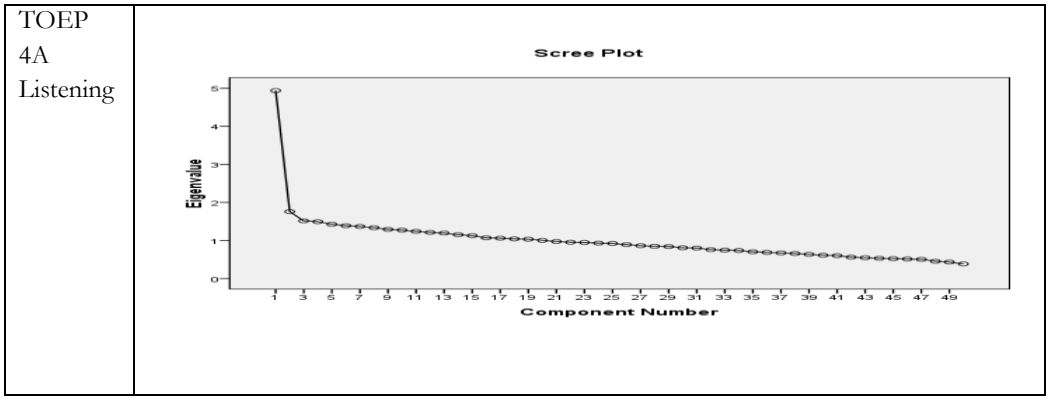


Figure 6. Scree Plot for TOEP 4A set

A little different result occurs in TOEP 4B set. In the TOEP 4B for listening section, there is one part in the graphics that shows a steep line, one other part which is rather steep, and another part which is slope. It indicates that listening 4B measures at least two dimensions, which are, listening dimension and other dimension. The more complete result is shown in Figure 7.

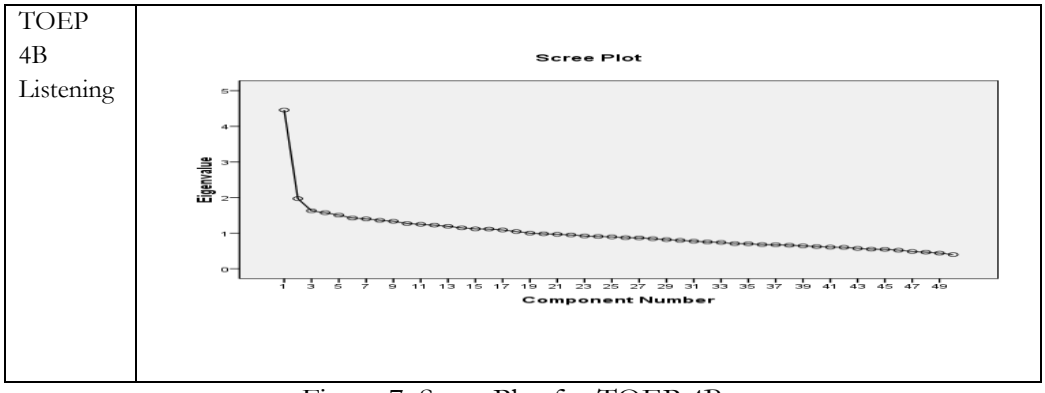


Figure 7. Scree Plot for TOEP 4B set

Based on those results, graphically, there are two main domains measured using TOEP set. In listening section, the measured domain is listening. In reading section, the measured main domain is reading.

Dimensional Validation with Explainable Variance Percentage

By employing analysis, the Eigen value, the result of explainable variance analysis, can be figured out. Explainable variances percentage is used to find out and explain how the score of the measurement result and its variation is. An instrument is said to be unidimension if the value is above 20%. Based on the analysis result presented in Table 1, the explainable variable percentage is still far below 20%. Some sets such as listening 1A, 2A, and 3B are close to 15% so it can be said as containing dominant dimension.

Table 1. Explainable variance percentage from TOEP set

TOEP	L
1A	15.890
2A	15.500
2B	12.600
3A	11.329
3B	14.953
4A	9.872
4B	8.910

Dimensional Validation with the First and Second Eigen Value Ratio

The first and second Eigen value ratio analysis result is presented in Table 2. Based on Table 2, a relatively equal result with explainable variance percentage result is gained. TOEP in listening section 1A, 2A, 3B, and reading 1A, 4B shows that those sets are unidimensional, while others contain dominant dimensions.

Table 2. Result analysis of the first and second Eigen value ratio

TOEP	Listening		
	λ1	λ2	λ1/λ2
1A	7.945	1.561	5.090
2A	7.750	1.686	4.597
2B	6.300	1.856	3.394
3A	5.664	1.681	3.369
3B	7.479	1.855	4.032
4A	4.936	1.759	2.806
4B	4.455	1.969	2.263

Discussion

Based on the analysis result with graphics method, the explainable variance, and the ratio of the first and second Eigen value, it can be said that listening set is proven to contain one dominant dimension, which is, listening only, although there is another dimension which is measured. Another dimension is vocabulary. In doing the listening test, the participants need not only listening skill, but also understanding on the words which are listened to. This is related to the vocabularies and expressions which have been understood by students. In the listening section, there are three points which are measured: responses, conversation, and mini talk. These three components require vocabulary mastery to understand, so it can be understood that listening does not merely measure dominant dimension only, but also other dimensions. The example of responses, conversation, and also mini talk are presented below.

An example of responses question

(Man): Hi, Amir. Long time no see. Where have you been?

A. I see. I have a long story to tell you.
B. Hi, Budi. I was overseas for a short course.
C. Yeah, Steve has been out for a long time.
D. I have been waiting here for you since dawn.

An example of conversation question

Man : Sorry, I'm late.
Woman: What happened? Did you lose your way?
Man : No. I had to work overtime finishing the report for tomorrow's meeting. It's a very busy time for us this week.

Question: Why was the man late?

In your test book, you read:
A. He lost his way.
B. He missed his bus.
C. He was in a meeting.
D. He had to work extra hours.

An example of mini talk:

The powerful healing properties of plants, spices, minerals, and fruit have been used for centuries. Ten everyday ingredients, gathered from all over the world, can be used to treat common ailments and injuries. There is no need for expensive prescriptions, you will find most of these remedies in your cupboard, or under the sink.

The first ingredient is aloe vera. Scientists are not sure how it works, but the gel you get when you cut a leaf of an aloe vera plant is rich in anti-inflammatory compounds as well as a chemical called bradykininase that acts as a topical painkiller. You can buy products containing aloe vera, but there's no substitute for the real thing. The plant is easy to grow on a kitchen windowsill and thrives on neglect.

To soothe sunburn, cuts, piles, and minor burns, wash the affected area thoroughly with soap and water. Then cut a chunk off a leaf, slice it lengthways and squeeze out the gel. Apply a

generous coating to the injured area and repeat two or three times a day.

What does the talk mainly discuss?

A. The curing power that aloe vera offers.
B. Ten ingredients that aloe vera consists of.
C. How to treat common ailments and injuries.
D. Why there's no substitute for aloe vera.

What can be said about the chemical called *bradykininase?*

A. It is anti-inflammatory.
B. It can relieve pains.
C. It can kill tropical animals.
D. It substitutes soap and water.

Assumptions which emerge during the development of the research seem to influence this analysis result. Since formulating the objectives of the test, it has been intended that the test which is being developed measures only one dimension, that is, listening ability. Related to this, it has been attested that TOEP set measures only one dimension, namely English listening ability.

Conclusion

From the seven listening sets which were analyzed, all of them contain listening dominant dimension when they were analyzed with graphical method, explainable variance, and Eigen value ratio. Based on this result, it can be said that listening section in TOEP contains dominant dimension, which is, listening only. This result brings implication towards further analysis related to the utilization of unidimensional item response theory. According to this case, analysis with the use of logistic model can be applied, in case to reveal the items quality quantitatively. Further mapping which is related to the factor loading and items substance needs to be done in order to figure out which item which measures more than one dimension. This can be a consideration for TOEP developer.

References

Ackerman, T. A., et al. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement*, 22, pp. 37-53.

Anonymous. (2001). Factor analysis. *Journal of Consumer Psychology*, 10 (1&2), pp. 75-82. Lawrence Erlbaum.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bachman, L. F. & Palmer, A. S. (1996). *Language Testing in Practice.* Oxford: Oxford University Press.

Bolt, D. M. (2001). Conditional covariance-based representation of multidimensional test structure. *Applied Psychological Measurement*, Vol. 27, No. 3, pp. 244-257.

Bolt, D. M. & Lall, V. M. (2003). Estimation of compensatory and noncompensatory multidimensional item response models using Marcov chain Monte-Carlo. *Applied Psychological Measurement*, 27, pp. 395-414.

De la Torre, J. & Patz. (2005). Making the most of what we have: A practical application of multidimensional item response theory in scoring. *Educational and Behavioral Statistics*, 30, pp. 295-311.

Garson, D. (2006). *Factor analysis.* Retrieved on 24 September 2006 from http://www2.chass.ncsu.edu/garson/pa765/index.htm .

Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). *Fundamental of item response theory.* Newbury Park, CA: Sage Publication Inc.

Hullin, C. L., et al. (1983). *Item response theory: Application to psychological measurement.* Homewood, IL: Dow Jones-Irwin.

Joreskog, K. & Sorbom, D. (1993). *Lisrel88: Structural equation modeling with the SIMPLIS command language.* Hillsdale, NJ: Scientific Software International.

Kirisci, L., Hsu, T. & Yu, L. (2001). Robustness of item parameter estimation programs to assumtions

of unidimensionality and normality. *Applied Psychological Measurement*, 25, pp. 146-162.

Munby, J. (1981). *Communicative syllabus design: A sociolinguistic model for defining the content of purpose-specific language programs.* Cambridge: Cambridge University Press.

Nunally, J. (1978). *Psychometric theory (2nd ed.).* New York, NY: McGraw Hill.

Sireci, S. G., & Geisinger, K. F. (1995). Using subject-matter experts to assess content representation: An MDS analysis. *Applied Psychological Measurement.* Vol. 19. No. 3, pp. 241-255.

Stapleton. (1997). *Basic concepts and procedures of confirmatory factor analysis.* Retrieved on 25 September 2006 from http://ericae.net/ft/Cfa.HTM

Van de Geer, J. P. (1971). *Introduction to multivariate analysis for the social sciences.* San Francisco: W. H. Freeman and Company.

Wang, W. C., Chen, P. H., & Cheng, Y. Y. (2004). Improving measurement precission of test batteries using multidimensional item response models. *Psychological Methods,* Vol 9, No. 1, pp. 116-136.

Wells, C. S. & Purwono, U. 2009. Assesing the fit of IRT models to item response data. *Makalah Pelatihan Psikometri Kerjasama Pascasarjana UNY dengan USAID.*