ICMSE 2015
INTERNATIONAL CONFERENCE ON MATHEMATICS, SCIENCE, AND EDUCATION
Faculty of Mathematics and Natural ScienceSemarang State University
phone: +62 248508112,+62 85876267947, +62 81805844994
email:icmse2015@gmail.com
website: icmseunnes.com

APPLICATION PARTIAL CREDIT MODEL TO MEASURE THE QUALITY OF CHEMISTRY ADAPTIVE TESTS AT VOCATIONAL HIGH SCHOOL

Suwahono*, Budiyono[2], and A.K. Prodjosantoso[3]

[1]UIN Walisongo semarang, Jln walisongo 3-5, Semarang, 50151, Indonesia
[2]UNS,Jln Ir Sutami 36 A, Surakarta, 57126, Indonesia
[3]UNY,Jln Colombo, Yogyakarta 55281, Indonesia

ABSTRACT

This study aims to determine the quality of the adaptive test chemical in SMK is being developed. Chemical adaptive test is a test used to measure the ability of learning chemistry on vocational high school. The approach used in this study is a survey approach. Population of The study includes adaptive test package chemistry SMK Textiles and student responses to test items. Adaptive test chemical that was developed using the test system completely wrong on the grounds. The subjects of this study were 1021 vocational students Texmaco Semarang. item test is analysed using item response theory models. Model partial credit model is a polytomouss model of IRT that follow the pattern of Rasch. When we assume that an item follow the pattern of partial credit then the trait/ability higher individuals are expected to have higher scores than an individual who has the low trait/ability. The results using the winstep program shows that the requirement unidimensionality about adaptive test chemical SMK fulfilled this case because it has raw variance equal to 32.6 %, this shows that the 20 % minimum requirement has been met. Results of the person reliability of 0.89 with a good category. Reliability item about 0.99 with a special category. Andrich threshold for chemistry adaptive tests measure moves ranging from none , -1.43 , -1.29 to 2.72 sequentially move this means that the gradation sequence assessment guidelines for assessors are valid.

Key Word: partial credit model, Winstep, Andrich threshold

INTRODUCTION

Vocational High School is a high school that aims to prepare graduates who are intelligent, ready to work, and competitive. To prepare course conducted diverse learning activities. Starting from the productive or vocational learning as the basis, the normative - adaptive learning as a supporter , and extracurricular activities as a means to channel the interests and talents of students.

Productive or vocational learning has a large proportion of the teaching and learning activities in vocational, for at least three days a week or 18 hours of lessons with the assumption that every hour lesson is 45 minutes. This learning tailored to the majors where students are, could be in the laboratory or classroom. SMK majors textiles for the samples in the learning experiments conducted in the laboratory of textiles, factory direct, and in the classroom.

Adaptive chemistry for students vocational high school

An adaptive Chemistry aim to support productive learning that has been obtained by students in each department is one adaptive lesson in SMK. Chemical presence as adaptive materials in SMK should be placed at the actual path, the chemical as adaptive underlies material productive thought patterns (Suwahono, budiyono and Prodjo, 2015), this means that existing chemical learning must be tailored to productive learning. The need for standardized gauges is needed by teachers of chemistry field. This is the starting point of research and development of chemical measurement tools adaptive.

Partial Credit Model (PCM)

In the Item response theory (IRT), the measurement model differentiated by category of

response and the number of parameter items that were included in the model . Based on the response categories IRT measurement model can be divided into dichotomous IRT models and models IRT polytomous. Dichotomous IRT models used in its test item has two response categories. While polytomous IRT models used in its test items have more than two response categories. This study uses a polytomous models for items in measuring devices developed using two categories of response categories and four categories of response. Dichotomous response obtained from the response to the correct one testier derived from structured reason answered by testier.

This model is commonly used in typical performance test. Currently working on a test sheet and the chemical generic skills observer observation results when filling the observation sheet generic skills chemistry. Politomous response obtained from the testee phase of reasons generic skills test sheets chemistry (see figure.1).

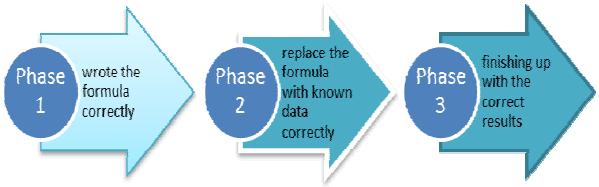

Figure. 1 The stage of completion in a matter of generic skills test chemical

According to the stage of completion of item, this research politomi models used is Partial Credit Model (PCM). Partial credit can be given to measures that could be addressed by individuals. Model PCM is suitable to be worn on the research conducted. This relates to a matter that will be used requiring phase identification of problems until a final solution (Masters & Wright, 1984, Embretson & Reise, 2000) .

PCM was originally developed to analyze the test items that require multiple steps to completion. PCM is the development of Model 1-PL and into the family of Rasch model. Model simple dichotomy in the Rasch model is a special case of PCM. The analytical results indicate PCM certainly fit with the data fit well in the model dichotomy (Wu and Adams, 2007). Model dichotomy and PCM can be said mixture in a single analysis.

1. **Equation**

   PCM is a development of the Rasch model item dichotomy applied to items politomi. Rasch Model dichotomy item contains only one item location parameter is then developed by describing the location of items into categories. This development is called the characteristic function (operating characteristic functions/OCF), which is defined by the following equation (Engelhard, 2005).

$$P_{i1}(\theta) = \frac{p_{i1}(\theta)}{p_{i0}(\theta) + p_{i0}(\theta)} = \frac{\exp(\theta_n - \delta_{i1})}{1 + \exp(\theta_n - \delta_{i1})}$$

Description;

θ_{ni} = individual trait levels (location of individual trait on the latent trait continuum)

δ_{in}= item location parameter (indicates the probability of obtaining a score of 0 and 1 each)

OCF became the prototype development for item politomi Rasch model . If i was item politomi by category score, 0 , 1 , 2 ... , m_i , then the probability of individual n score x on item i that will be described in Category Response Function (CRF) was realized in the following equation;

$$p_{ix}(\theta) = \frac{exp\left[\sum_{j=0}^{x}(\theta_n - \delta_{ij})\right]}{\sum_{r=0}^{m_i}\left\{exp\left[\sum_{j=0}^{x}(\theta_n - \delta_{ij})\right]\right\}}$$

Description;

θ_{ni} = individual trait levels (location of individual trait on the latent trait continuum)

δ_{ij} = the intersection of the lines between categories (j) in point (i)

Equation above can be elaborated based on the number of categories in items. For example a scales having a category 3 with a score of 0.1, and 2. Then we get categories (j) as many as 3 equation that the probability of individuals in each category

Probability in category 0

$$P_{i0}(\theta) = \frac{1}{1 + \exp(\theta_n - \delta_{i1}) + exp[(\theta_n - \delta_{i1}) + (\theta_n - \delta_{i2})]}$$

Probability in category 1

$$P_{i0}(\theta) = \frac{\exp(\theta_n - \delta_{i1})}{1 + \exp(\theta_n - \delta_{i1}) + exp[(\theta_n - \delta_{i1}) + (\theta_n - \delta_{i2})]}$$

Probability in category 2

$$P_{i0}(\theta) = \frac{exp(\theta_n - \delta_{i1}) + exp(\theta_n - \delta_{i2})}{1 + exp(\theta_n - \delta_{i1}) + exp[(\theta_n - \delta_{i1}) + (\theta_n - \delta_{i2})]}$$

Probability category 0 look number 1 in the denominator. This is because the PCM requires the following equation,

$$\sum_{j=0}^{0} (\theta - \delta_{ij}) = 1$$

a) Graph of PCM

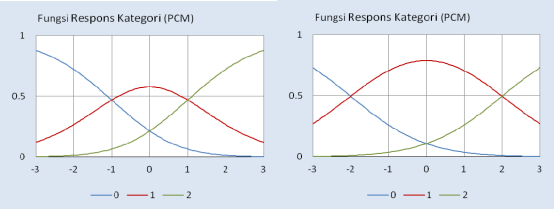

Figure 2. Graph of OCF and CRF

b) Interpret parameters δij

A lot of the terminology used to describe stage δij among other parameters (step parameters), difficulty stages (step Difficulties item) or intersection category (category intersections) (du Toit, 2003). δij shows the probability of the meeting point of two lines in the item category. The above equation shows the probability of an individual in response to the stage category x mi represents the difference between the level of the trait (Θ) and parameter intersection category (δij). In other words, the intersection of parameter categories can be considered as the level of difficulty associated with the transition phase from one category to the next, and there are difficulties steps mi (junction) for items with mi + 1 category of answers (Embretson & Reise, 2000). It should be noted that the value δij not always have to be sequential in point i as a relative magnitude of the two probabilities which is nearby. Further, when for example the individual probabilities Pi1 (Θ) and PI2 (Θ) is the same, the values δij also be the same (Tatsuoka, 1996).

δij can also be interpreted as the point at which the latent trait scale successive two categories of response curves intersect so called crossroads category (category response curves intersect). δij is the point where the two categories have the same probability to be selected by the associated trait level (Linacre 1998,2012).

Because it is the intersection of two lines categories , then it will be easier to interpret δij if seen directly in the item characteristic curve (curve characteristics item / ICC) . Shown in Figure 8 there are two points of intersection between the lines , namely line category 0 and category 1 characterized by δij and the intersection between the lines, namely line category 0 and category 1 characterized by δi2.

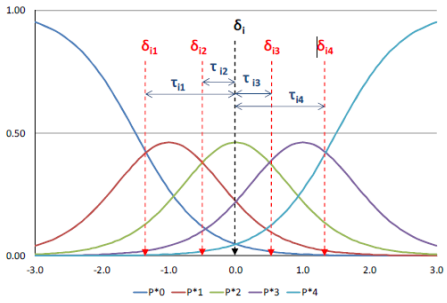

Figure 3. scale stage of completion of item

On the other hand δij not indicate the level of difficulty for success in the second phase or to achieve a score of 2, but rather shows the item difficulty for the second phase independent with earlier stages (Wu and Adams, 2007). In PCM all items are considered to have equal slopes on all categories that these terms appear in the discussion of the PCM. δij in this model does not show the individual has a probability of 0:50 in response to the threshold of the above categories as well as the parameters bij did in the GRM, but rather indicates the relative degree of difficulty at each stage. δij more indicates the position where the latent trait continuum intersecting response categories so that individual traits are more likely to later stages than the previous stage.

If we are forced to interpret δij as item difficulty level then how to interpret it is as follows. To δij based on Figure 8 we found that individuals who have high levels of abilitas under δij has a high probability to choose a category 0 compared with the category 1 (Wu and Adams, 2007).

c) Pure Score estimate

PCM also provides a procedure for estimating the score of expectation or pure scores

associated with item response. The equation used is as follows;

$$E(X) = \sum_{x=0}^{m_i} xP_x(\theta)$$

Description

χ = score item

$P_x(\Theta)$= Probility particular trait level in response to item

As with other Rasch model, armed with raw scores we've been able to estimate the pure scores. Individuals who have the same raw score will be located at the level of the same trait. The challenge is that all items should have the same relationship with the underlying latent trait. In practice, this challenge is difficult to achieve.

Subjects and populations

The subjects used in this study are vocational students texmaco located in the border region Semarang Kendal. The number of participants in this study a number of 1058 participants

RESULTS AND DISCUSSION

Result Measuring Chemical adaptive Testing Using PCM

In this test case the test results obtained fit , testing of the model fit, fit item and estimate the parameters of each item. Here are the results of testing the model fit, fit item and estimate the parameters of each item. Testing of the model fit in the IRT is composed of 2 forms of person fit testing and test fit items

a. unidimensionality.

Unidimensionality is an important measure that states the instruments developed can measure what should be measured, the results issued by Winstepin figure 5.

```
TABLE 23.0 ALL                                  ZOU463WS.TXT  May 14 18:57 2015
INPUT: 1058 Person  35 Item  REPORTED: 1058 Person  35 Item  4 CATS WINSTEPS 3.73
--------------------------------------------------------------------------------

    Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)
                                            -- Empirical --    Modeled
Total raw variance in observations     =         51.9 100.0%        100.0%
  Raw variance explained by measures   =         16.9  32.6%         31.3%
    Raw variance explained by persons  =         10.5  20.2%         19.4%
    Raw Variance explained by items    =          6.4  12.4%         11.9%
  Raw unexplained variance (total)     =         35.0  67.4% 100.0%  68.7%
    Unexplned variance in 1st contrast =          6.1  11.7%  17.4%
    Unexplned variance in 2nd contrast =          3.3   6.3%   9.4%
    Unexplned variance in 3rd contrast =          2.2   4.2%   6.2%
    Unexplned variance in 4th contrast =          2.0   3.8%   5.7%
    Unexplned variance in 5th contrast =          1.8   3.4%   5.1%
```

Figure 5. unidimensionality result

From the table issued by Winstep on raw variance is explained by measure is 32.6 % the price is above the limit of 20 % defined feasible measurement system developed for use because it only measures the ability of a chemical as adaptive materials in vocational high school.

b. Testing items fit

An item said to be fit if infit mean square value is in the range of 0.70 and 1.80 (Adams and Khoo, 1993). Testing of the items fit gives information about which items that do not fit, so it should not be used in further analysis. From the test results obtained using software Winstep results;

```
TABLE 13.1 ALL                                  ZOU463WS.TXT  May 14 18:57 2015
INPUT: 1058 Person  35 Item  REPORTED: 1058 Person  35 Item  4 CATS WINSTEPS 3.73
--------------------------------------------------------------------------------
Person: REAL SEP.: 2.91  REL.: .89 ... Item: REAL SEP.: 11.53  REL.: .99

        Item STATISTICS:  MEASURE ORDER

-------------------------------------------------------------------------------------------
|ENTRY  TOTAL  TOTAL          MODEL|   INFIT  |  OUTFIT  |PT-MEASURE |EXACT MATCH|      |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| Item |
|-----------------------------------+----------+----------+-----------+-----------+------|
|    19   1461   834     .80    .05| .73  -6.6| .75  -5.3|  .65   .56| 45.6  51.8| N19  |
|    22   1577   887     .79    .05| .75  -6.3| .74  -5.8|  .73   .55| 41.7  50.7| N22  |
|    24   1623   912     .79    .05| .81  -4.8| .80  -4.6|  .72   .55| 39.6  50.1| N24  |
|    21   1595   893     .79    .05|1.29   6.2|1.20   3.9|  .45   .54| 38.0  49.8| N21  |
|    20   1594   891     .77    .05|1.35   7.5|1.28   5.5|  .47   .55| 33.8  50.2| N20  |
|    13   1839  1058     .76    .04|1.32   7.4|1.46   9.1|  .37   .54| 39.5  51.6| N13  |
|    23   1611   892     .75    .05| .81  -4.9| .78  -5.1|  .71   .55| 36.7  49.3| N23  |
|    29   1902  1032     .56    .04|1.22   5.5|1.26   5.8|  .26   .53| 40.3  44.4| N29  |
|    14   1981  1058     .50    .04|1.30   7.5|1.38   8.4|  .38   .53| 27.6  42.2| N14  |
|    34   1912  1012     .43    .04|1.27   6.9|1.24   5.7|  .48   .51| 33.7  41.9| N34  |
|    17   1641   809     .36    .05| .83  -4.6| .83  -4.0|  .57   .52| 41.3  44.4| N17  |
|    12   2073  1056     .33    .04|1.59   9.9|1.57   9.9|  .42   .51| 24.5  42.2| N12  |
|    26   1697   844     .32    .05|1.22   5.4|1.32   6.7|  .35   .51| 45.6  42.4| N26  |
|    25   1825   883     .21    .05| .64  -9.9| .67  -8.9|  .61   .51| 45.0  43.1| N25  |
|    28   1953   909     .08    .04| .56  -9.9| .54  -9.9|  .69   .50| 59.2  43.3| N28  |
|    35   2205  1010    -.06    .04|1.08   2.2|1.08   2.1|  .33   .49| 42.9  42.9| N35  |
|    18   1873   826    -.07    .05| .82  -5.0| .94  -1.3|  .72   .49| 58.1  44.8| N18  |
|    10   2325  1058    -.09    .04|1.11   3.2|1.06   1.6|  .43   .48| 33.2  42.6| N10  |
|    11   2369  1058    -.16    .04|1.20   5.8|1.18   4.4|  .50   .47| 32.9  43.1| N11  |
|    31   2281  1019    -.17    .04| .81  -6.2| .90  -2.5|  .48   .46| 46.3  42.9| N31  |
|    27   2015   871    -.18    .05| .64  -9.9| .61  -9.9|  .63   .48| 54.4  45.3| N27  |
|    33   2354  1039    -.20    .04| .83  -5.5| .87  -3.4|  .56   .47| 47.7  43.8| N33  |
|    30   2349  1020    -.30    .04| .77  -7.5| .77  -6.1|  .41   .44| 56.4  44.2| N30  |
|    15   2467  1057    -.34    .04|1.01   .2|1.15   3.7|  .34   .45| 44.8  44.8| N15  |
|     1   2500  1058    -.39    .04| .92  -2.5| .98  -.4|  .45   .44| 56.0  45.2| N1   |
|    16   2502  1058    -.39    .04| .73  -9.0| .71  -8.1|  .57   .44| 54.5  45.2| N16  |
|    32   2455  1020    -.47    .04| .85  -4.5| .81  -4.8|  .62   .43| 47.6  46.0| N32  |
|     8   2597  1057    -.56    .04|1.09   2.5|1.23   5.1|  .23   .43| 42.4  47.7| N8   |
|     9   2610  1058    -.58    .04|1.05   1.3|1.16   3.6|  .41   .43| 42.5  48.1| N9   |
|     2   2620  1058    -.60    .04| .99   -.1|1.05   1.0|  .44   .42| 51.6  48.2| N2   |
|     4   2596  1047    -.62    .04| .94  -1.5|1.01   .2|  .42   .42| 43.4  48.1| N4   |
|     5   2640  1058    -.64    .04|1.19   4.7|1.41   8.4|  .26   .42| 47.3  48.8| N5   |
|     3   2702  1057    -.76    .04| .84  -4.3| .73  -6.7|  .52   .41| 60.2  50.6| N3   |
|     7   2632  1033    -.76    .04| .85  -4.0| .79  -5.1|  .47   .40| 59.4  50.0| N7   |
|     6   2761  1056    -.88    .05| .75  -6.4| .70  -7.1|  .39   .40| 68.0  52.7| N6   |
|-----------------------------------+----------+----------+-----------+-----------+------|
| MEAN  2146.8  985.4     .00    .04| .98  -1.1|1.00   -.4|           | 45.2  46.4|      |
| S.D.   397.9   86.0     .53    .00| .24   5.9| .27   5.9|           |  9.8   3.3|      |
-------------------------------------------------------------------------------------------
```

Figure 6. measurement result of item measure

Based fig 6 issued by Winstep software, of the 35 items tested are known;

a) Outlier items using Mean size and S.D ; 0.98 + 0.24 = 1.22 then the item numbers 20, 21, 13 and 12 is a misfit items or outliers,

b) Mean square (MNSQ) received 0.5 < MNSQ < 1.5, the number 12 with a value of 1.59

c) Point Correlation measure (Pt Mean corr) 0.4 to item number 12 could not be concluded item number 12 items outlier.
But if used measure of adams and Khoo then the item number 12 is acceptable.

c. Item reliability and Person reliability
Reliability item and the person is a hallmark of IRT in which the reliability of items and reliability of person there is no relationship, the results returned by Winstep in figure 7.

```
ABLE 3.1 ALL                                   ZOU463WS.TXT  May 14 18:57 2015
NPUT: 1058 Person  35 Item  REPORTED: 1058 Person  35 Item  4 CATS WINSTEPS 3.7
-------------------------------------------------------------------------------

    SUMMARY OF 1058 MEASURED Person
-------------------------------------------------------------------------------
          TOTAL                          MODEL         INFIT        OUTFIT    |
          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
-------------------------------------------------------------------------------|
 MEAN      71.0      32.6       -1.07     .25      1.01    -.1   1.01    -.1 |
 S.D.      15.8       2.3         .81     .03       .29    1.4    .31    1.4 |
 MAX.     106.0      35.0        1.31     .34      1.84    2.9   1.92    3.0 |
 MIN.      43.0      25.0       -2.38     .22       .37   -4.1    .34   -3.7 |
-------------------------------------------------------------------------------|
 REAL RMSE    .26 TRUE SD     .77  SEPARATION  2.91  Person RELIABILITY  .89 |
MODEL RMSE    .25 TRUE SD     .77  SEPARATION  3.13  Person RELIABILITY  .91 |
 S.E. OF Person MEAN = .02                                                   |
-------------------------------------------------------------------------------
       VALID RESPONSES:  93.1%  (APPROXIMATE)
'erson RAW SCORE-TO-MEASURE CORRELATION = .96 (approximate due to missing data)
RONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .93 (approximate du

    SUMMARY OF 35 MEASURED Item
-------------------------------------------------------------------------------
          TOTAL                          MODEL         INFIT        OUTFIT    |
          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
-------------------------------------------------------------------------------|
 MEAN    2146.8     985.4         .00     .04       .98   -1.1   1.00    -.4 |
 S.D.     397.9      86.0         .53     .00       .24    5.9    .27    5.9 |
 MAX.    2761.0    1058.0         .80     .05      1.59    9.9   1.57    9.9 |
 MIN.    1461.0     809.0        -.88     .04       .56   -9.9    .54   -9.9 |
-------------------------------------------------------------------------------|
 REAL RMSE    .05 TRUE SD     .53  SEPARATION 11.53  Item   RELIABILITY  .99 |
MODEL RMSE    .04 TRUE SD     .53  SEPARATION 12.06  Item   RELIABILITY  .99 |
 S.E. OF Item MEAN = .09                                                     |
-------------------------------------------------------------------------------
```

Figure 7 item reliability and person reliability

Results of the person reliability as much as 0.89 while for the item reliability 0.99. This shows that the person reliability in kategory good and item realiability for special category.

d. Andrich threshold
Analysis of the validity of the ranking scale is testing done to verify whether the rating of the given option to take measurements used confusing or not. Rasch model analysis provide verification process for asusmsi ratings given in the instrument. From the results of Winstep stated that the choice of polytomus (andrich threshold, 1988) moves from none, -1.43, -1.29 and 2.72 it states is a valid option given.

```
SUMMARY OF CATEGORY STRUCTURE.  Model="R"
-------------------------------------------------------------------------
|CATEGORY   OBSERVED|OBSVD SAMPLE|INFIT OUTFIT|| ANDRICH |CATEGORY|
|LABEL SCORE COUNT %|AVRGE EXPECT|  MNSQ  MNSQ||THRESHOLD| MEASURE|
|-------------------+------------+------------++---------+--------|
|  1  1    9812  28| -1.67 -1.68|   .98   .97||  NONE   |( -2.91)| 1
|  2  2    9571  28| -1.23 -1.21|  1.14  1.25||  -1.43  |  -1.38 | 2
|  3  3   14237  41|  -.52  -.52|   .96   .98||  -1.29  |    .81 | 3
|  4  4     868   3|   .45   .40|   .91   .98||   2.72  |( 3.83)| 4
|-------------------+------------+------------++---------+--------|
|MISSING    2542   7| -1.81      |            ||         |        |
-------------------------------------------------------------------------
OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.
```

Figure 8. Andrich threshold result

CONCLUSION

The results using the winstep program shows that the requirement unidimensionality about adaptive test chemical SMK fulfilled this case because it has raw variance equal to 32.6 %, this shows that the 20 % minimum requirement has been met. Results of the person reliability of 0.89 with a good category. Reliability item about 0.99 with a special category. Andrich threshold for chemistry adaptive tests measure moves ranging from none, -1.43, -1.29 to 2.72 sequentially move this means that the gradation sequence assessment guidelines for assessors are valid.

REFERENCES

Adams, R. J. & Khoo, S.T. (1993). Quest- The Interactive Test Analysis System. Hawthorn,Victoria: ACER.

Andrich, D. (1988). Rasch Models for Measurement. Newbury Park CA: Sage.Masters, G.N. & Wright, B.D. (1984). The essential process in a family of measurement models. *Psychometrika*, (49), 529-544.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum

Hambleton, R. K., & Swaminathan, H. (1985). Item response theory: principles and applications. Boston: Kluwer-Nijhoff Publishing.

Linacre, J. M. (1998). Structure in Rasch residuals: Why principal components analysis (PCA)? Rasch Measurement Transactions, 12(2), 636. Retrieved from **http://www.rasch.org/rmt/rmt122m.htm**

Lineacre, J. M. (2012). A User's Guide to Winsteps. Ministep: Rasch-Model Computer Programs (Program Manual 3.7.5.0).

Suwahono, Budiyono &Prodjo (2015) Needs analysis for the development of chemical adaptive assessment in vocational high schools, *Proceeding International Conference on Science & Science Education, Satya Wacana Christian University, held in August 1, 2015*.

Tatsuoka, K. K. (1996). Use of generalized person-fit indices, zetas for statistical pattern classification. *Applied Measurement in Education*, 9, 65-76.

Wu, M., & Adams, R. (2007). *Applying the Rasch model to psycho-social measurement: A practical approach*. Melbourne: Educational Measurement Solutions.