

The Characteristic of Analytical Thinking and Science Process Skills (ATSPS) Test for Senior High School Students

Irwanto

Chemistry Teacher, Budya Wacana Yogyakarta Senior High School
Jl. Cik Di Tiro Blok GK V/248 Yogyakarta, Indonesia, *email: irwan_uny@yahoo.com

Abstract

This research aims to describe the characteristic of ATSPS test for measuring students' analytical thinking and science process skills of chemistry. This research involved 188 students of XI grade of four senior high schools in Yogyakarta, Indonesia. ATSPS test consists of 15 essay items. The result of tryout was polytomous data and was analyzed with Partial Credit Model 1-Parameter Logistic (PCM 1-PL) using Winsteps 3.73. The result on the field tryout stage based on the lowest and the highest limit Outfit MNSQ 0.5 and 1.5, all tryout test items were stated fit with the PCM 1-PL. Alpha Cronbach coefficient based on CTT is 0.70. The difficulty level of question items was good, because of the range -0.74 to $+1.24$ logit. Maximum score of test information function is on the ability -0.1 with information score 9.164 and standard error of measurement 0.33. Therefore, ATSPS test is suitable for measuring the analytical thinking and science process skills of senior high school students.

Keywords: ATSPS test, analytical thinking, science process skills

Introduction

Assessment can mean as a process for getting information in any form of becoming the basic of decision taking about the students. The assessment can be done using assessment instrument in doing any activities of learning process evaluation or towards the learning result of students. The learning process will make the teacher as the reference in doing any assessment. Then, the assessment can be the guideline for evaluating the further learning process. The result according to Redecker & Johannessen (2013) states that the assessment can improve and increase the learning quality. Otherwise, the learning activities can mean to evaluate the students' success, determining the follow-up, acknowledging the attained abilities, improving learning activities, and helping the students study harder (Fautley & Savage, 2008).

The Government Regulation Number 19 year 2005 and Permendiknas (Regulation of National Education Minister) Number 41 year 2007 states that the assessment in the learning system functions as the way of assessing the learning

process and result by students (President RI, 2005; Kemendiknas, 2007). Then, the assessment of learning result aims at measuring the competency level by the students. The actualization of learning result assessment aim at evaluating the students achievements, either knowledge, attitudes or skills. Therefore, the assessment in the level of chemistry is demanded to be able to include the above competencies.

The learning process can be actualized effectively and efficiently when the systematic process of planning, actualizing, and assessing are available. The assessments are in accordance with the Government Regulation Number 19 year 2005 about national standard of education, paragraph 64, article 1 stating that the learning result assessment done by the teachers continuously to monitor the process, development and improvement. On the other hands, the Paragraph 19 article 3 states that the assessment can use many techniques according to the basic competency acquired by the students. The assessment technique can be in the form of written tests, assignments, observation, and practices (President RI, 2005).

One of assessment techniques often used in the chemistry, i.e. assessment in the practicum. The practicum is one of the learning activities correlating with those as parts of a series of learning process. As the practicum in chemistry correlating directly with the science process skills, so the assessment for the practicum should not only include the aspects of cognitive, but focussing on the overall aspects of skills (Sundari, 2008).

The practicum activities cannot be separated from the science process skills of the students in the laboratory activities. The condition happens as the practicum activities of chemistry are very interesting, so they can provide the students to cooperate, looking for the new ideas, and solving many problems. Otherwise, the practicum activities can functions as skill in the scientific work, such as observing, asking questions, predicting, formulating hypothesis, using tools and materials, processing data, and communicating the results.

The assessment of either cognitive aspect or skill process can be measured using written test (Subali, 2011). The cognitive aspect includes the high level

thinking ability of the students. One of the high level thinking ability is analytical thinking ability. The analytical thinking is students' abilities in describing concept into part of the more detailed concept and explaining the correlation between parts. Analytical thinking is not just memorizing the facts, but also their effort to increase the understanding to the complexities and creative to reach the learning objectives better. The analytical thinking ability is the high level of cognitive thinking they can reach after acquiring the lower level of cognitive thinking such as remembering, understanding, and applying.

Permendikbud (The regulation of Minister of Education and Culture) Number 66 year 2013 about the education standard assessment mentions the some principles such as objective, integrated and economical ones (Kemendikbud, 2013). Objective assessment means that the assessment is based on the standard and never influenced by subjective factors of the teachers. Integrated assessment means scheduled, integrated with the learning activities and the continuity. Economic assessment means effective and efficient in planning, actualizing, and reporting.

Assessment in the chemistry learning process aims at knowing the students proficiency after the learning process. One of the competencies to reach in the chemistry learning is mastering to the science process skills and analytical thinking ability. Temiz, Taşar & Tan (2006) state that the science process skills are part for which it is never separated all at once having the central role in developing the conceptual understanding of students in the learning activities. The condition makes the reason of why the science process skills are very important to develop and assess.

Durmaz & Mutlu (2014) explain that when the learning done with more accentuated to the science process skills, it will impact more positive to the students' learning results. Thus, the assessment of science process skills are very important component in the chemistry learning. The condition happens as the assessment can support the students to be more hard to study continuously and support the teachers to increase their qualities in the learning process (Badu, 2012).

The assessment is one of the teachers' responsibilities to know the students capabilities in obtaining their learning objectives. The success of learning result assessment is determined by the teachers' abilities in constructing and using the measurement tools correctly, and analyzing data resulted (Guntur, Sukadiyanto & Mardapi, 2014). However, in fact, there are still many teachers have no instruments of assessments to the science process skills in specific.

The science process skills are those of thinking used to build knowledge in solving the problems. This condition happens as the measurement of science process skills measurement of the students in the practicum activities is difficult to achieve. The difficulties encountered when doing the science process skills measurement is when there are so many students. The other challenges are when there are no assessment standards, so the teachers have many difficulties in assessing the students' learning results (Setiani, 2011). Meanwhile, one of competencies whom the teachers should have is developing instruments of assessments (Prasetya, 2012).

The lower capabilities of teachers in constructing tests can be seen based on their activities whose never use skill assessment instrument in the process of practicum. The assessment of science process skill aspects are only based on the teacher subjective assumption, i.e. when the students seems doing good practicum, they will get good marks without considering whole aspects concerning with the skills to be measured. Otherwise, the existing assessment instruments in some schools have been not referred to the special skills to measure their capabilities in the practicum.

Based on the problem, it is necessary to have instruments to provide their optimum smartness. On the other hand, it needs an instrument able to evaluate, increase knowledge and understanding of analytical thinking, and facilitating them to increase their science process skills in learning chemistry.

Method

Research Type

This type of research is that of educational research and development (R&D). The educational research and development is a method used to produce

some certain products, examining the validity, practicality, and effectiveness of the products (Patwary & Surjono, 2015). The products of educational research and development can be a learning media, books, moduls, learning instruments, school policies, or assessment instruments. The educational research and development aims at increasing products effectively and efficiently to use in schools.

Research Subjects

The study was conducted in SMA Negeri 2 Yogyakarta, SMA Negeri 4 Yogyakarta, SMA Negeri 6 Yogyakarta, and SMA Negeri 10 Yogyakarta. The determination of research subjects used a purposive sampling, so it can be obtained about 188 students for academic year of 2015/2016.

Procedure

This research used the development model of Wilson and Oriundo & Antonio. There are some steps in developing tests for chemistry including planning, experimenting and constructing tests.

Data, Instruments, and Data Collecting Technique

The data in this educational research and development is the characteristic of assessment instrument for chemistry. The characteristics of the data were gained from the empirical experiment including models, estimate reliability, item difficulty level, information function test, and measurement error. The instruments of data collecting such as essay to measure students' analytical thinking and science process skills in chemistry.

Data Analysis Technique

The data were analyzed and described according with the test to produce understandable information. The products revised then reconstructed becoming a ready-to-use test for the students. The experiment results were then analyzed using Partial Credit Model 1-Parameter Logistic (PCM 1-PL). The data analysis from the experiment was then tested using Winsteps 3.73. The item analysis was held using Item Response Test (IRT). The item analysis was done to determine the fit item using the model, estimating test reliabilities, and knowing the level of difficulties to the items.

The test item is then said fit with the model when having values of Outfit MNSQ around 0.5 to 1,5 (Linacre, 2002), while the good items have the difficulty index more than -2.0 or less than $+2,0$ (Hambleton & Swaminathan, 1985). The PCM is the explanation of Rasch model (1-PL), by the assumption that each item has the same differential and just contains the parameter of item difficulty level.

Result and Discussion

The Goodness of Fit

The result of empirical test was analyzed using Partial Credit Model 1-Parameter Logistic (PCM 1-PL) by calibrating using Winsteps program. The calibration of test item aims at determining overall fit item test using PCM. The test item is said fit with the model when having value of Outfit MNSQ around 0.5 to 1.5, while the value out of the reach is said misfit with model (Thomas, Anderson & Nashon, 2008). The value of Outfit MNSQ to the analysis result of experiment data is evenly distributed between 0.63 to 1.50. The goodness of fit of the item with model of partial credit with 1 logistic parameter is presented in Table 1.

Table 1. Goodness of Fit

Items	MNSQ	Notes	Items	MNSQ	Notes
1	1.50	Accepted	9	0,63	Accepted
2	0.80	Accepted	10	1,47	Accepted
3	0.92	Accepted	11	0,65	Accepted
4	0.98	Accepted	12	1,34	Accepted
5	0.86	Accepted	13	0,85	Accepted
6	1,46	Accepted	14	1,00	Accepted
7	0,99	Accepted	15	1,40	Accepted
8	0,67	Accepted			

Based on the item analysis result using value of Outfit MNSQ, there are 15 items stated fit with model. Therefore, whole grains tests be regarded as a final product which has been empirically valid.

Reliability Estimation

Reliability is the consistency of the scores obtained by the person taking the test when tested repeatedly using the same test on different occasions. Based on the analysis items using the program of Winsteps Cronbach alpha, the coefficient is obtained by Classical Test Theory (CTT) of 0.70. The estimation results of

reliability are based on an index of separation test use Winsteps program as presented in Figure 1. Therefore, all items in the final product expressed empirically reliable.

SUMMARY OF 188 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	43.4	15.0	.02	.16	1.05	.1	1.03	.1
S. D.	13.3	.0	.32	.03	.49	1.1	.51	1.0
MAX.	73.0	15.0	.60	.44	2.61	2.6	3.09	3.2
MIN.	11.0	15.0	-1.46	.13	.29	-2.2	.29	-1.8
REAL RMSE	.18	TRUE SD	.27	SEPARATION	1.50	Person RELIABILITY	.69	
MODEL RMSE	.16	TRUE SD	.28	SEPARATION	1.77	Person RELIABILITY	.76	
S. E. OF Person MEAN = .02								

Person RAW SCORE-TO-MEASURE CORRELATION = .97
 CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .70

Figure1. Result of Test Reliability Estimation

Rated reliability of the test is likely to be lower if the students are not consistent, meaning that at a certain point, there are many tests about as a higher precisely one capable, otherwise many tests are less capable in really doing about it (Subali & Suyata, 2012). The condition can occur if the students careless, not working, or working on the test item at random.

Item Difficulty Level

The parameter of item difficulty (*b*) describes the approximate level of difficulty of the item expressed in logic. An item is said to be good when the index has more difficulty than -2.0 or less than +2.0 logit. Items that have a level of difficulty below -2.0 is categorized as too easy one, whereas items that have a level of difficulty above +2.0 is categorized as too difficult.

The level of item difficulty is one of the factors that affect the probability of the test students' answers in response to a specific item. Items with a relatively high degree of difficulty is high abilities need to be answered correctly, while the items are easily requiring a low level of abilities to be answered correctly.

In this case, the items are categorized easy when the value of *b* (measure) less than -1, categorized fairly when $-1 < b < +1$, and categorized difficult when the value of *b* more than +1 (Adedoyin & Mokobi, 2013). Analysis items trial results showed that the level of difficulty of items in the range of -0,74 to +1,24

logit. Therefore, all items (15 items) in the test classified into good category. Therefore, all items (15 items) have either category according to Table 2.

Table 2. Item Difficulty Level

Items	Measure	Category	Items	Measure	Category
7	1.24	Difficult	12	-0.20	Fair
3	0.93	Fair	5	-0.25	Fair
4	0.50	Fair	1	-0.30	Fair
14	0.35	Fair	6	-0.32	Fair
9	0.26	Fair	11	-0.62	Fair
8	0.04	Fair	13	-0.66	Fair
2	-0.08	Fair	10	-0.74	Fair
15	-0.14	Fair			

Table 2 provides information on the level of difficulty of each item about the test results from the instrument items as the most difficult to easiest. Item 7 is the most difficult item (+1.24 logit), while item 10 is the easiest item (-0.74 logit). Information difficulty level of each items in the program output Winsteps measure of value. Items classified as difficult to have a measure value is positive, while the items are relatively easy to have a value measure is negative.

The analysis shows that the item 7 has the parameter *b* of +1.24, meaning that the required minimum test students' ability of +1.24, while the item 10 requires only minimal ability logit -0.74 on a scale to be able to answer correctly. The greater value of the parameter *b*, the greater the capabilities required to be able to answer correctly with probability 50%. In other words, the greater the value of the parameter *b*, then the item is more difficult. Histogram distribution level of difficulty of items on the analysis using SPSS Statistics 17 is presented in Figure 2.

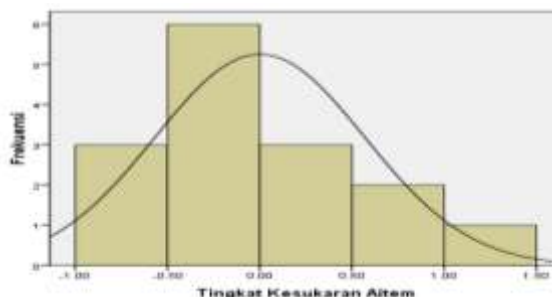


Figure 2. Item Difficulty Level

Based on the analysis items using Winsteps program, obtained by the 15 items are fit for use in the measurement phase. The condition is due to the entire item with a good level of difficulty, ranging from -2.0 to +2.0 logit.

Test Information Function

The test information function is a strength test in revealing trait (ability) as measured by the instrument. Function test information will vary according to the value of θ (ability) of students. At a certain value of θ , the value of the information function will reach a maximum point. The maximum point indicates that if the items was done by students with certain θ , it will obtain the highest information (Naga, 1992).

Information function tests provide a more accurate measure in estimating the ability of test takers than the coefficient of reliability (Samejima, 1994). The TIF curve for educational tests is generally bell-shaped with a maximum amount of information to be near a specific point (DePascale & Dunn, 2008). The ability of the test students are on the horizontal axis, while the function test information is on the vertical axis. The higher peak information function tests, the higher the reliability of a measurement value is.

Based on calculations using Microsoft Excel 2010 is obtained maximum value information function test in the students' ability (θ) -0.1 with the information value of 9.164 as well as measurement error (SEM) of 0.33. In general, the visualization of information function test and measurement error are presented in Figure 3.

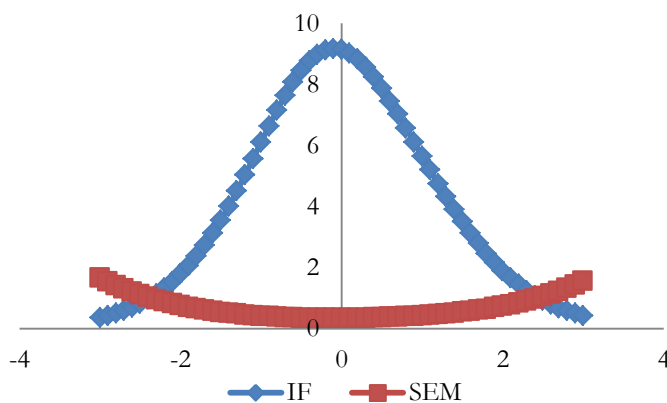


Figure 3. Test Information Function

Figure 3 shows that the range of very low ability of students obtains the information function were low. Likewise, in the range of very high ability students obtains information function relatively low. The condition indicates that the test produces high information function when given to students who have the medium ability.

The results obtained by the parameters itemsfit, reliability estimation, and item difficulty level can be concluded that test have been a reliable and standardized. Therefore, a chemistry test is feasible used to measure the analytical thinking ability and science process skills of senior high school students.

Conclusion

Based on the analysis, it can be concluded that: 1) all items (15 items) in the ATSPS test declare fit with model, 2) the estimation of reliability corresponding to Cronbach alpha coefficient based on the classical test theory (CTT) is 0.70, 3) the item difficulty level in the range of -0.74 to +1.24 logit, 4) the maximum value of information function test in the students' ability (θ) -0.1 with the information value of 9.164, and 5) standard error of measurement is 0.33.

References

- Adedoyin, O.O., & Mokobi, T. 2013. Using IRT psychometric analysis in examining the quality of junior certificate mathematics multiple choice examination test items, *International Journal of Asian Social Science* 3: 992-1011.
- Badu, S.Q. 2012. Implementasi evaluasi model Kirkpatrick pada perkuliahan masalah nilai awal dan syarat batas, *Jurnal Penelitian dan Evaluasi Pendidikan*, Edisi Dies Natalis ke-48 UNY Tahun 2012.
- DePascale, C.A., & Dunn, J.L. 2008. Simplifying psychometrics with SAS. Diakses dari <http://www.lexjansen.com/nesug/nesug08/ap/ap07.pdf> pada tanggal 26 Januari 2016.
- Durmaz, H., & Mutlu, S. 2014. The effects of an instructional intervention on 7th grade students' science process skills and science achievement. *Cukurova University Faculty of Education Journal* 43: 155-168.
- Fautley, M., & Savage, J. 2008. *Assessment for Learning and Teaching in Secondary Schools*. London: Learning Matters Ltd.
- Guntur, Sukadiyanto, & Mardapi, D. 2014. Pengembangan asesmen hasil belajar penjasorkes siswa sma pada permainan bolavoli, *Jurnal Penelitian dan Evaluasi Pendidikan*, 18: 13-29.

- Hambleton, R.K., & Swaminathan, H. 1985. *Item Response Theory Principles and Applications*. Boston: Kluwer Nijhoff Publishing.
- Kemendikbud. 2013. *Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 66 Tahun 2013 Tentang Standar Penilaian Pendidikan*.
- Kemendiknas. 2007. *Peraturan Menteri Pendidikan Nasional Republik Indonesia Nomor 41 Tahun 2007 Tentang Standar Proses untuk Satuan Pendidikan Dasar dan Menengah*.
- Linacre, J.M. 2002. Optimizing rating scale category effectiveness, *Journal of Applied Measurement* 3: 85-106.
- Naga, D.S. 1992. *Pengantar Teori Sekor pada Pengukuran Pendidikan*. Jakarta: Universitas Gunadarma.
- Patwary, M.A., & Surjono, H.D. 2015. The implementation of e-learning with team builder at vocational high schools, *Research and Evaluation in Education Journal* 1: 25-44.
- Prasetya, T.I. 2012. Meningkatkan keterampilan menyusun instrumen hasil belajar berbasis modul interaktif bagi guru-guru IPA SMP N Kota Magelang, *Journal of Educational Research and Evaluation*, 1: 106-112.
- Presiden RI. 2005. *Peraturan Pemerintah RI Nomor 19 Tahun 2005 Tentang Standar Nasional Pendidikan*.
- Redecker, C., & Johannessen, O. 2013. Changing assessment – towards a new assessment paradigm using ICT, *European Journal of Education* 48: 79-96.
- Samejima, F. 1994. Estimation of reliability coefficients using the test information function and its modifications, *Applied Psychological Measurement* 18: 229-244.
- Setiani, F. 2011. Pengembangan asesmen alternatif dalam pembelajaran matematika dengan pendekatan realistik di sekolah dasar, *Jurnal Penelitian dan Evaluasi Pendidikan* 15: 250-268.
- Subali, B. 2011. Pengukuran kreativitas keterampilan proses sains dalam konteks assessment for learning, *Jurnal Cakrawala Pendidikan* 15: 130-144.
- Subali, B., & Suyata, P. 2012. *Pengembangan Item Tes Konvergen dan Divergen dan Penyelidikan Validitasnya Secara Empiris*. Yogyakarta: Diandra Pustaka.
- Sundari, R. 2008. An evaluation on the use of laboratory in teaching biology in public madrasah aliyahs in Sleman regency, *Jurnal Penelitian dan Evaluasi Pendidikan* 12: 196-112.
- Temiz, B.K., Taşar, M.F., & Tan, M. 2006. Development and validation of a multiple format test of science process skills, *International Education Journal* 7: 1007-1027.
- Thiagarajan, S., Semmel, D., & Semmel, M. 1974. *Instructional Development for Training Teachers of Exceptional Children: A Sourcebook*. Minneapolis, Leadership Training Institute/Special Education: University of Minnesota.
- Thomas, G., Anderson, D., & Nashon, S. 2008. Development of an instrument designed to investigate elements of science students' metacognition, self-efficacy and learning processes: The SEMLI-S, *International Journal of Science Education* 30: 1701-1724.