JERE 9 (1) (2020) 31-39

Journal of Educational Research and Evaluation

http://journal.unnes.ac.id/sju/index.php/jere

Development of Assessment Instruments The Learning Process Chemistry Based Scientific Approach to High School Students

Muh Ikhsanul Yakin✉, Kartono Kartono, Endang Susilaningsih

Universitas Negeri Semarang, Indonesia

Info article

Published 31 March 2020

Keywords: development of assessment instruments, learning chemistry, scientific approach

Abstract

Instrument based assessment of the scientific approach is very important learning process developed to measure scientific students. The instrument is said to be good and true if valid criteria, the reliability, practical, and effective. This research aimed to know the instrument character, to get a valid and reliability instrument, to examme instrument practicality, and to test the instrument effectiveness level in measuring students scientific learning process. This research used a qualitative and quantitative and adapted the method of research and development Borg and Gall models. The content validity of the instrument was validated by experts. Assessment instruments were tested at 30 students (limited) and 100 students (broad). The validity of the instrument items were analyzed with the program SPSS 16.0 version and obtamed which representated 20 items every scientific learning indicators. Reliability were analyzed with the SPSS program 16.0 version and obtamed the reliability by value 0.908 (very high). The validity of the assessment instruments constructs were analyzed by using LISREL program 8.8. version and model compatibility fulfilled a good of fit criteria with the grade of value (0.6159), RMSEA (0.000), RMR (0.061), CFI (1.00), NFI (0.98), GFI (0.97), AGFI (0.93), NNFI (0.98), PNFI (0.74), and PGFI (0.81). The results of the value calculation of construct reliability was 0.95 and variance extracted value was to 0.54. Assessment instruments had practicality and effectiveness level with "very good" category based on the user's response.

✉Correspondence address :
Kelud Utara 3 kampus Pascasarjana UNNES Semarang, Indonesia
E-mail : ihsanulyakin@gmail.com

p-ISSN 2252-6420
e-ISSN 2503-1732

INTRODUCTION

All this time the assessment of chemistry learning in high school have not used based assessment instruments based on scientific approach. The assessment used by the teacher in the form of a standard observation sheets and traditional test (paper and pencil test) that do not reveal the students learning process are complex. Assessment results do not often represent the level of mastery of the concept and the actual student learning conditions. Shaklee (1997:90), in exposing the students mastery concepts, assessment is not only unveil the concept that has been achieved, but also know the development of learning process how a concept is obtained.

The reality in the field, assessment of students learning process which is done by the teacher have not been carried out with the maximum. The chemical learning by using scientific approach which study is required in the curriculum 2013 requires assessment instruments can measure the students scientific learning process which is tested validity, reliability, practical, and effective. Assessment of current chemical subjects in high school are less relevant to the function, purpose, and assessment principles which is set by Kemendiknas and is recommended by expert assessment of education.

The problem is accordance with the research conducted by Wiggins in the Journal Phi Delta Kappan (1984:703) stated the design of the test that is used by teacher has not fulfilled the real test standardization, so it has not actual described directly the performance of students in detail and essential which is appropriate with educational objectives. The design of assessment prefer the learning result capacity but it is not the thought process of students, which indirectly will power off the creativity of the students.

The traditional test experts expressed that the test based traditional is difficult to measure and to assess the understanding of the nature and the process of how scientists work (Marzano, 1994:31). This has been assumed by Lee (1992:92), stated that " Standarded testing has a cumulative negative impact on student". According to the results of the study Deborah, L et al, (2007:176), that traditional tests like this was criticized for ignored the inquiries of students who play a major role in the learning activities.

Commenced with the establishment of the 2013 curriculum which demands a learning approach based scientific and strengthen learning in the through authentic assessment, researcher wants to develop a tool to measure the assessment valid, reliabil, practical, and effective. This research becomes the solution of non existence measurement tool appraisal in assessing the learning process of the scientific high school students.

Borg & Gall (2007:272), stated the development research is research oriented to develop and to validate the products that are used in education. Seel & Richey (1994:11), the development is a translation process which is designed specifications into physical form. Characteristics of the assessment instrument in this study refers to the measurement of learning scientific approach indicator which tailored to the content of the lesson material oxidation reduction reaction chemistry that is studied by students of class X high school and developed with the development of model measures Borg & Gall.

Meanwhile, there are five aspects of the competency based learning, scientific approach that is to observing, asking, trying, reasoning, and communicating which needs a measurement tool to assess scientific-based learning. Learning-based scientific approach is a learning that adopts scientists steps in constructing knowledge through scientific methods so that it will louild scientific thinking skills, develop "sense of inquiry" and the ability of the creative thinking of the students (Alfred De Vito, 1989:120).

Resnick (1985:38), in fact the assessment focuses on the process of assessment of student learning. Wiggins (1984:703), assessment means a medium chronologically can help teachers in monitoring student. Meanwhile, Popham (1995:31), the assessment was supposed to be a part of learning that can not be integral. Authentic assessment according to experts (Marzano, 1994; Popham, 1995; Bookhart, 2001), authentic assessment is used to describe

the assessment of performance because the assessment tasks are closer to real life.

The assessment activity is closely related to the learning activities, and it is an integral activity in the learning process. The assessment activity is an activity that can not be separated with learning and it is done continuously in every learning process both in the classroom and outside the classroom. Cangelosi (1995:21), argued that the assessment was a decision about values. The assessment has elements of decision making based on results measurement of data and assessment results of information during and after teaching and learning activities.

Grondlund (1993:86), the validity leads to accuracy of the interpretation results to an certain evaluation procedure user which complies with the purpose of the measurement. Wiersma (1990:50), validity content is actually basing on the logic analysis, so it is not constitute a validity coefficient which is measured in statistics. Allen and Yen (1979:108) constructs validity is the validity that shows the extent to which the results of tests capable of revealing a trait or a theorizing constructs which will be measured.

This research resulted in the assessment instrument of chemical learning process which is based on scientific approach that is content validity and constructs validity. The validity of the content is estimated through testing in the relevance or expediency of test content through rational analysis by panels which is competent or through expert judgement method (Stephen Hora, 2009:1-2). The validity of constructs describes the extent to which the results of the test are able to uncover a validity constructs of theoritical which will be measured (Allen and Yen, 1979:108).

The validity of constructs assessment instruments are analyzed by confirmatory factor analysis of second order using LISREL 8.8 version software. Confirmatory factor analysis (CFA) is the factor analysis procedure to identify the existence of relationships between variables by doing the test the reliability, instrument validity and reliability with calculating the factor loading value or coefficients factor or the value of the lamda (λ_i). The CFA procedures is used to measure the statistical model which is used in accordance with the empirical data (Waltz, Strickland, & Lenz, 2005:3-7).

The instrument has a good content validity if it consists of items that represent all the learning indicator based on scientific approach that became the object of measurement. The validity of the content is done to examine grills instruments to ensure the relevance of the content of the instruments content with the measurement purpose. The validity of constructs means that the measuring instrument can be valid if the assessment is in accordance with the theory construction in where the instrument was made. The validity of constructs in this study tested the fit models GOF good fit criteria is RMSEA, RMR, CFI, NFI, GFI, AGFI, NNFI, PNFI, and PGFI.

Gronlund (1993:169), "reliability refers to the consistency of measurement, that is to how consistent test scores or other evaluation result are from one measurement to other". Reliability refers to the measurement results. Regularity a good reliability model if the construct of reliability $\geq$0.70 and variance extracted $\geq$0.5 (Allen and Yen, 1979:110). An outcome measurement can only be trusted when the measurement execution in recent times in the same subject obtained a relatively same measurement results for aspects that are measured in the subject indeed has not changed.

Akker (1999:10), stated that the practicality refers to the user level in assessing the instrument can be used and preferred in normal conditions. The level of practicality in the development of the learning material is seen from whether teachers and experts consider that the material is easy and can be used by teachers and students (Nieveen, 1999:12). This assessment instrument can be said as practical if the user in this case teachers stated that the assessment instrument model can be applied in the field and the implementation level categorizing at least "good".

Nitko and Brookhart (2007:60), the effectiveness of the instruments must pay attention to four issues, namely financing (cost), efficiency, practicality, and instructional features. The instrument is said to be effective if the assessment instruments meet the valid criterion, reliability, economical, objective, systematic and practical

(Sudiyatno, 2010:239). Akker (1999:14), effectively a refers to the levels that the experience and the results of interventions are consistent with the intended purpose.

METHODS

This research is a research and development (research and development). Borg and Gall (2007:626), research development is a strategy and efforts to develop an effective education is used to overcome learning problems.

This research development refers to the development model of the Borg and Gall (2007:766) which is just 10 steps, namely : (1) preliminary studies, (2) planning, (3) model hipotetik development, (4) review of model hipotetik, (5) revision, (6) trial limited, (7) the revision of the trial results, (8) broad test, (9) revision last model, (10) desiminasi.

Data collecting instrument that is used the interview guidelines sheet validation expert, interview, scoring rubrics, teacher response questionnaire. Qualitative analysis of instruments, is started from the stage of examination by the hypothetical expert (expert judgment) in the grills and granule instrument with walktrough technique. Quantitative analysis is to do an analysis the result of trial limited and extensive trials. Quantitative analysis was conducted to find out the grains characteristics of the item in the instrument validity and reliability form.

The validity of the grain items empirically are sought with pearson product moment correlation significance level at 5%, if the chances of error ≤0.05. Reliability of the instrument are considered good if it has the value cronbach's alpha ≥0.800. Validity and reliability analysis using SPPS program version 16.0. The validity of constructs is analysed by applying the confirmatory factor analysis of second order. Confirmatory factor analysis conducted with confirmatory factor analysis (CFA) software using LISREL 8.8.

The notation of LISREL, latent variable exogenous is assessment instruments based on scientific. Endogenous latent variables declared by "eta" symbolized (η) and indicators expressed by lamda (λ) is called the factor loading value. A conceptual structural model of the assessment instruments based on scientific which become endogenous latent variables (η) are five indicators of learning scientific approach: observing (η1), asking (η2), gathering information (η3), association (η4), and communicating (η5). Endogenous variable indicator measurement errors revealed by the EPSILON (ε).

A conceptual model of the structural instruments assessment basesd on the scientific approach has 20 indicators which is symbolized by X1-X20 presented in Figure 1 below.

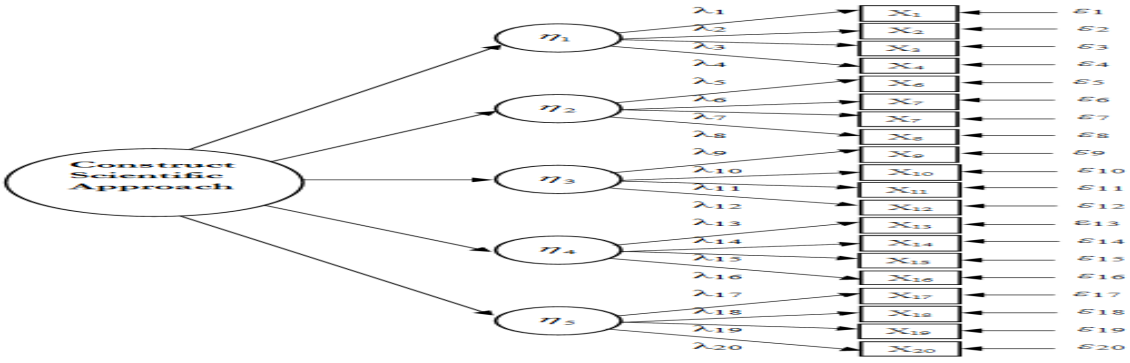


Figure 1. A Conceptual Model Of Instrument Based Assessment Of The Scientific Approach

A construct model of assessment instruments which based scientific approach is two confirmatory factor analysis of model that are not correlated. It means that there is no direct connection between each 5 (five) indicator based scientific learning approach.

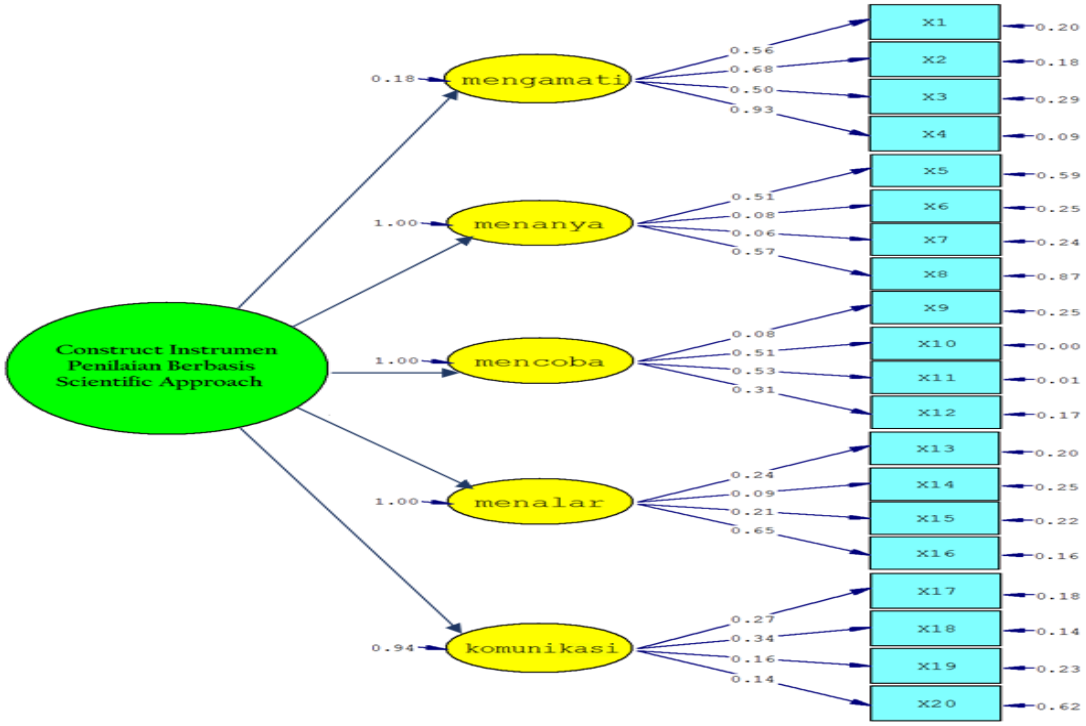
RESULTS AND DISCUSSION

The first product is tested at SMAN 5 Semarang (2016-2017). Analysis design of model design advisability test is done by LISREL program 8.8 version. The calculation result of the test instrument will be outlined as follows.

Thes test validity of the instrument is carried out with pearson correlation analysis using program SPSS version 16.0. The prerequirement that is used pearson correlation coefficient is $r_{calculate} \geq r_{table}$ of 0.195 (for 100 samples). The results of the grains validity analysis shows that all items are valid and become items which is compiled into assessment instruments device in chemistry learning process based scientific approach.

The test of reliability of the instrument is carried out by using cronbach alpha analysis software SPSS version 16.0. The results of test reliability calculation is gained reliability coefficient value of acquiring 0.908, it constitutes that 90.8% variance is observed it is caused by the real score variance individual group caused by a true individual Group score and the correlation between the observed score and true score same as the 0.908 (Nur in Sappaile, 1987:61, 2005:27). Analysis results can be concluded that the assessment instruments have a high reliability.

The results of confirmatory factor analysis second order model of a path diagram value of lamda (factor loading) can be seen in Figure 2. the following.



Chi-Square=45.31, df=98, P-value=0.6159, RMSEA=0.000

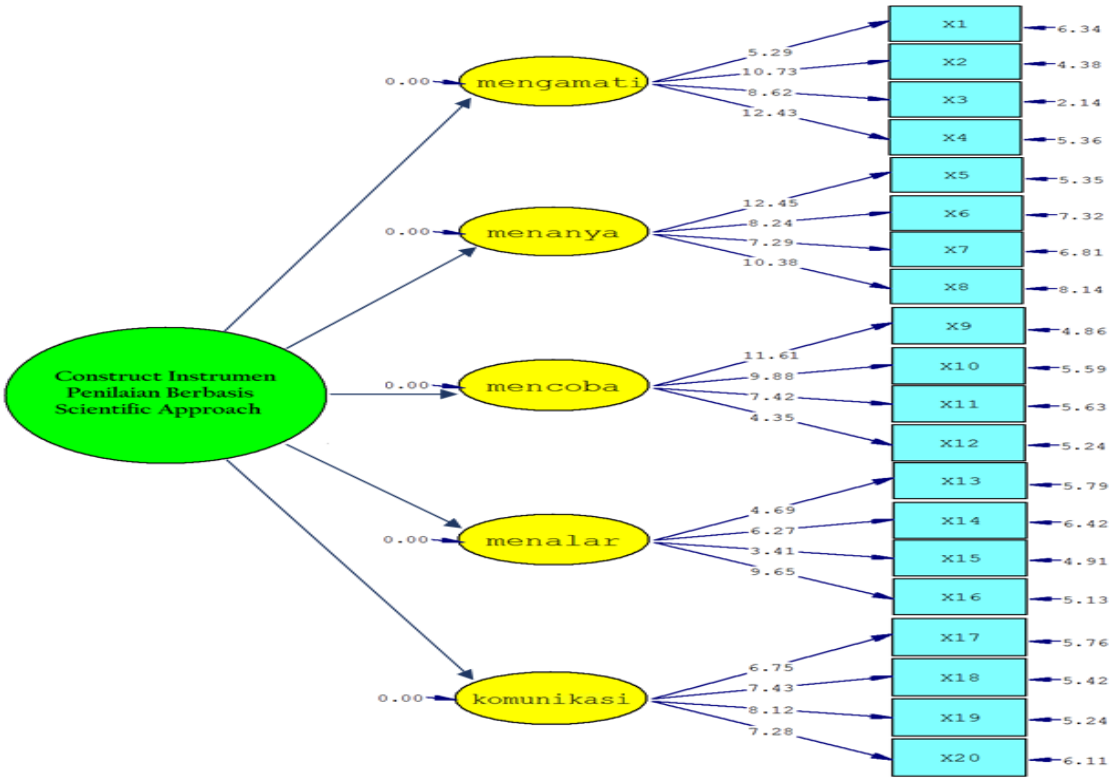
Figure 2. The Path Value of The Factor Loading Diagram

Confirmatory factor analysis results already fulfilled the criteria of goodness of fit. The value of probability testing goodness of fit indicates value 0.6159>0.05, and RMSEA 0.000<0.05. Test results of the other models such as the fit like CFI, NFI, GFI, AGFI, and NNFI it also produces the value >0.90 so that the models is revealed goodness of fit. Output path diagram that contains the value of the factor loading (λ) of five variables latent with 5 its indicators will be statistically significant because of all the 20

indicators have a value factor loading (λ)>0.05 (fit models).

The results of the calculation of construct reliability and value variance extract shows that the value of the construct of reliability is 0.95>0.70 (very high) and the value of the variance extracted 0.54>0.50 (good). The calculation result shows that all aspects have a value >0.50 it is according to analysis of variance extracted. it can be inferred that the reliability and variance of assessment instruments constructs in the learning process of chemical-based scientific approach has been fulfilled.

The results of confirmatory factor analysis second order model of a path diagram of the t-$_{value}$ can be seen in Figure 3. the following.



Chi-Square=45.31, df=98, P-value=0.06159, RMSEA=0.000

Figure 3. Diagram of The Path Value of T-$_{Value}$ Line

The analysis result of the value $t_{calculate}$ in Figure 3, it was concluded that there was no value in red. All grains are declared as valid and it can measure scientific learning variables. The analysis result of the value t-$_{value}$ indicates the value of the regression weight (standart estimate) is significant with the value $t_{calculate}$>1.95 (significance level at 5%). This analysis result proves that obtaining the empirical data support and prove that the instrument structural conceptual model developed is correct.

Analysis results of assessment instrument constructs based scientific approach are presented in table 1 and table 2 below.

Table 1. The Calculation Of The Value Of Assessment Instruments Are Invalid Constructs The Learning Process Chemistry

Variabel Laten	Perhitungan Estimates Nilai Factor Loading (λ)			Perhitungan Nilai Maximum Likelihood		Perhitungan Nilai Construct Reliability & Nilai Variance Extracted		Perhitungan Nilai Signifikansi Jalur (Uji T)		
	Indikator	Estimasi (λ)	Standar Error	Nilai R^2	Syarat	Construct Reliability	Variance Extracted	T_{hitung}	Syarat	Standar Error
Obsserving	Butir1	0.56	0.20	0.71	>0.5	0.90>0.70	0.71>0.50	5.29	>1.95	6.34
	Butir2	0.68	0.18	0.84	>0.5			10.73	>1.95	4.38
	Butir3	0.50	0.29	0.79	>0.5			8.62	>1.95	2.14
	Butir4	0.93	0.09	0.93	>0.5			12.43	>1.95	5.36
Asking	Butir5	0.51	0.59	0.92	>0.5	0.76>0.70	0.56>0.50	12.45	>1.95	5.35
	Butir6	0.08	0.25	0.78	>0.5			8.24	>1.95	7.32
	Butir7	0.06	0.24	0.79	>0.5			7.29	>1.95	6.81
	Butir8	0.57	0.87	0.64	>0.5			10.38	>1.95	8.14
Trying	Butir9	0.08	0.25	0.85	>0.5	0.78>0.70	0.58>0.50	11.61	>1.95	4.86
	Butir10	0.51	0.00	1.00	>0.5			9.88	>1.95	5.59
	Butir11	0.53	0.01	1.00	>0.5			7.42	>1.95	5.63
	Butir12	0.31	0.17	0.94	>0.5			4.35	>1.95	5.24
Reasoning	Butir13	0.24	0.20	0.82	>0.5	0.83>0.70	0.59>0.50	4.69	>1.95	5.79
	Butir14	0.09	0.25	0.57	>0.5			6.27	>1.95	6.42
	Butir15	0.21	0.22	0.69	>0.5			3.41	>1.95	4.91
	Butir16	0.65	0.16	0.94	>0.5			9.65	>1.95	5.13
	Butir17	0.27	0.18	0.76	>0.5	0.89>0.70	0.54>0.50	6.75	>1.95	5.76
Communicating										
	Butir18	0.34	0.14	0.64	>0.5			7.43	>1.95	5.42
	Butir19	0.16	0.23	0.96	>0.5			8.12	>1.95	5.24
	Butir20	0.14	0.62	0.78	>0.5			7.28	>1.95	6.11

Table 2. The Calculation Of The Value Of Test Match Criteria Model Goodness

Criteria Goodness of Fit	Testing	Requirements	Value Calculate	Fulfillment
Absolut	RMSEA	<0.05	0.00	Good Fit
	RMR	<0.05	0.016	Good Fit
Incremental	CFI	>0.90	1.00	Good Fit
	NFI	>0.90	0.98	Good Fit
	GFI	>0.90	0.97	Good Fit
	AGFI	min 0.90	0.93	Good Fit
	NNFI	min 0.90	0.98	Good Fit
Parsimony	PNFI	0.60-0.90	0.74	Good Fit
	PGFI	0.60-0.90	0.81	Good Fit
Decision				Good Fit

The calculation of the estimation coefficient value factor loading (λ) shows all the latent variables observing (η_1), asking (η_2), gathering information (η_3), association (η_4), and communicating (η_5) significant. The results of the test of goodness of fit indices have fulfilled the criteria which is set so that the theoretical models submitted are stated in accordance with the empirical data.

The calculation of the value of maximum likelihood (R^2) showed all indicators is significant and has a value of R^2>0.5. The greatest contribution of 20 indicators is the grain of 10 and 11 with a value of R^2 granules of 1.00.

The value of R^2 represents the reliability of indicators, so that 20 indicators can be expressed from the 10 and 11 grain is the most reliability.

The calculation of the value of the construct reliability showed high reliability value. The value of the variance extracted all indicators show the recommended appropriate >0.50. Calculation of all aspects of learning scientific approach obtains the value of the construct of reliability of 0.95>0.70 and value variance extracted 0.54>0.50. These calculations showed that the assessment instrument has a reliability model.

The evaluation result of the goodness of fit looks the constructs has already fulfilled the criteria for goodness of fit. The value of probability testing goodness of fit showed 0.6159<0.05, RMSEA 0.00<0.05, and RMR 0.061<0.05. The results of a test match is such a like CFI 1.00>0.90, NFI 0.98>0.90, GFI 0.97>0.90, AGFI 0.93, NNFI 0.98, PNFI 0.74, and PGFI 0.81 have met the criteria for testing goodness of fit of the model so that the instrument is declared fit.

Teacher response is obtained by practicality questionnaire scoring technique. The calculation result of the response towards all aspects have practicality value by category "very good". The calculations result are also obtained a score based on every indicators aspects of the clarity of the instructions and "very good" aspect of the implementation "very good". The calculation result shows the assessment instrument based scientific approach has practicality value with the category "very good".

Teachers respons to the effectiveness of the instruments was obtained by effectiveness questionnaire scoring technique. The results of the analysis of the teacher's response towards the effectiveness of the instruments obtained information that all aspects of the Organization has a value of "very good". The results of the calculations are also obtained a score based on every aspect of the indicators aspects of objectivity "very good", "very good" systematic and economical aspects of "very good". The calculation result shows the instrument-based assessment of the scientific approach has value effectiveness with the category "very good".

CONCLUSION

This research has resulted the development of products in the form of assessment instruments to measure learning chemistry based scientific approach. The research results obtained the conclusion that all indicators are significant and can measure the latent variables learning scientific indicators. This means that 5 indicators in this study significantly can measure student learning process based scientific approach. An assessment of the practicability and effectiveness is instrument shows the practicality and effectiveness of the good based on the users response. The analysis of data generated 20 grains valid, test then it is compiled into a assessment instrument device based on scientific approach. This 20 indicators, become the final draft of the assessment instrument in the chemistry learning process based scientific approach at high senior school students.

REFERENCES

Akker, J. V. 1999. Principles and Methods of Development Research. Design Approaches and Tools in Education and Training. Dodrecht: Kluwer Academic Publisher.

Alfred, De Vito. 1989. Creative Wellsprings for Science Teaching. West Lafayette. Indiana: Creative Venture.

Allen, M.J. & Yen, W. M. 1979. Introduction to Measurement Theory. Monterey: Brooks Cole.

Brookhart, S.M. 2001. The Art of Science of Classroom Assessment. Eric Digest.

Cangelosi, James S. 1995. Merancang Tes untuk Menilai Prestasi Siswa. Bandung: ITB.

C., Stephen., Hora. 2009. Expert Judgment in Risk Analysis. University of Hawaii. Non-Published Research Reports. Paper hlm 120.

Deborah, L et al. 2007. Collaborative Action Research to Improve Classroom Assessment in An Introductory Physics Course for Teacher. Journal phys. tchr. educ online, 4 (2). winter 2007.

Gall, M.D., Gall, J.P., and Borg, W. R. 2007. Educational Research: An Introduction (7th Ed.). Boston. MA: A & B Publications.

Gronlund, Norman., E. 1993. How to Make Achievement Test and Assessments. Boston: Allyn and Bacon.

Lee, F. Y. 1992. "Alternative Assessments". Chilhood Education. 69 (2), 72-73.

Marzano, R.J., Pickering, D.J., Mctighe, J. 1994. Assesing Student Outcomes: Performance Assessment Using the Dimensions of Learning Model. Alexandria: Association for Supervision and Curriculum Development.

Nieveen, Nienke. 1999. "Prototyping to Reach Product Quality". In J. Vam Den Akker,R Branch,K Gustafson, N Nieveen and Tj. Plomp (Eds). Design Approaches and Tools in Education and Training (hlm. 125-136). Dodrecht: Kluwer Academic Publisher.

Nitko, A.J., Brookhart, S.M. 2007. Educational Assessment of Student (3rd Edition). New Jersey: Pearson Education.

Popham, W. J. 1995. Classroom Assessment, What Teachers Need it Know. Oxford: Pergamon Press.

Resnick, D. P. & Resnick, L. B. 1985. Standards, Curriculum, and Performance: a Historical and Comparative Perspektive. Educational Researcher 9, 5 -19.

Sappaile, B.I. 2005. "Pengaruh Model Mengajar dan Ragam Tes terhadap Hasil Belajar Matematika dengan Mengontrol Sikap Siswa". Jurnal Pendidikan dan Kebudayaan. 056, 669-671.

Seels, Barbara B. and Richey, Rita. C. 1994. Instructional Technology: The Definition of The Field. Washington.D.C., Association for Educational Technology.

Shaklee, B.D. et. al. 1997. Designing and Using Portfolios. United States of America: Allyn & Bacon.

Sudiyatno. 2010. "Pengembangan Model Penilaian Komprehensif Unjuk Kerja Peserta Didik pada Pembelajaran Berbasis Standar Kompetensi di SMK Teknologi Industri". Disertasi. Yogyakarta: Program Pascasarjana Universitas Negeri Yogyakarta.

Waltz, C. F., Strickland, O. L., & Lenz, E. R. 2005. Measurement in Nursing and Health Research (3rd Ed.). New York: Springer.

William Wiersma and Stephen G,. Jurs. 1990. Educational Measurement and Testing. Boston London Sydney Toronto: Allyn and Bacon.

Wiggins, G. 1984. A True Test: Toward More Authentic and Equitable Assessment. Phi Delta Kappan. 70. (9) 703-713