The Analysis of Instrument of The Ability to Acting and Thinking Creatively Based Rasch Model

Suraji [1✉]**, Totok Sumaryanto, F.L** [2]**, Muhammad Khumaedi** [2]

[1.] SD Negeri Banyubiru 05 Kab. Semarang**,**Jawa Tengah, Indonesia
[2.] Universitas Negeri Semarang, Indonesia

Article Info

Abstract

The development of student's creativity was to answer the challenges of the 21st century curriculum. To know the aims of learning were reached, it was needed valid and reliable assesment. Research was conducted instrument development. The total sample testing instrument need 43 students from two schools who have used 2013 curriculum. Rater reliability data analysis using Hoyt's formula with SPSS software application . Validity and reliability items analyzed using Rasch model with Winstep program. Results: The 22th items tried out form 2 dimensions factor, they were namely the figural and verbal factors. Conclusion: results from test run on reliability was 0,94. The instrument based on the catagories was of good quality item. Based on the analysis factor it was will necessary to do improvement toward the arrangement of measuring instruments. The instrument consisted of two dimensions of the main dimensions which were figural and verbal dimension. Figural dimension had 10 measuring items, while verbal dimension had 12 measuring items. Instrument was arranged based on the lowest to highest logit. Sugestion: it required test run using larger sample. Items was arranged rom the easiest to hardest items based on logit number on each dimension. The benefits of this study: to arrange the model formula to assess the ability of acting and thinking creatively based on comic in the learning.

© 2019 Universitas Negeri Semarang

✉Correspondence Address
Jl. Kyai Djojo Projo No. 10 Krajan, Banyubiru, Semarang, Jawa Tengah 50664, Indonesia
E-mail: surajibelima@gmail.com

INTRODUCTION

The development of student's creativity is kind of curriculum demands to answer the challenge in 21st century. Skills in the 21st century consist of three components, which are learning skills, literacy skills and life skills. There are four skills that are often found in learning skills known as 4C (communication, collaboration, critical thinking, creativity and inovation). The development of creativity is listed in the core competence of curriculum 2013 and is known as think creatively and work creatively. These two creativity is a part of creativity concept definition. (Kementerian Pendidikan dan Kebudayaan, 2016).

One of the creativities in 21st century is creativity in curriculum syllabus 2013 about the ability to act and think creatively. This competence is stated in curriculum syllabus 2013 in core competence that says "to show the creativity in thinking and work creatively, productively, critical, independent, collaborative, and communicative. In clear words, systematic, logical and critical, in an aesthetic work,…" (Kemdikbud, 2017, p. vii). One of basic competencies that derives from "Kompetensi Inti" (KI) 4 is "Kompetensi Dasar" (KD) 4.1, to make comic. To evaluate the comic is used assessment rubric found in teacher's book (Kemdikbud, 2017, p.26). However, creativity assesment rubric in teacher's book only include product assessment. Psychomotor aspect contained in KI 4 that says "to show the creativive in thinking and to work creatively…." does not represent the valid and reliable assessment instrument that can measure the ability of thinking and creativity of making comic product.

Valid and reliable assessment is neccesary to understand the acquisition of learning objective appropriately. In order to know the level of student creativity, valid and reliable authenthic assessment system that can help to measure creativity of student is needed. This is because of psychomotor aspect that has less attention. It happens because teacher does not get into the authentic of a comprehensive assesment of learning.

Teacher it self does not have any preparation neither in psychomotor assessment aspect nor the instrument (Fuadi, Sumaryanto, & Lestari, 2015, p.2). The result of research came from Majid, Raharjo, & Supriyadi (2017, p.56) explained that primary schools in Bawang sub-district, Batang district did not run the assessment steps in learning outcomes as requested by assessment method against curriculum 2013.

According to Appulembang's research (2017, p.41-57) said that student creativity aged 6-12 years old could be measured through verbal and figural torrance assessement test. Student creativity can be measured by using figural test which is potential test to make comic and verbal test to test the ability of student to arrange a script and to deliver it properly.

Based on analysis by Guilford that there were five characteristics of creativity, for examples : fluency, flexibility, originality, elaboration (Putra & Iqbal, 2014, p.71). those characterisics are applied to measure comic creativity either in figural dimension or verbal dimension.

METHOD

This research was conducted at the primary school that inserted curriculum 2013 for the fifth grade students at regional education coordinator in Banyubiru sub-district and was stated in 2018-2019. The validity of instrument was assessed by three proffesionals. Reability or internal consistency of the assessment was analyzed using Hoyt's formula with anova using SPSS software. Research of this instrument development used Rasch model to analyze test run. Factor analysis used dimensionality analysis to recognize factors that create measurement contruct. The results were used to facilitate product revision.

There were some steps to develop the instrument refered to Mardapi, Allen & Yen, and ministry of education. Those steps were known as 1). Arranging specification of measurement instrument, 2). Writing points of measurement, 3). Examining

the points, 4). Conducting measurement trial, 5). Analyzing the points, 6). Improving the points, 7). Assembling measurement, 8). Conducting measurement, 9). Interpreting the results

Construct validity has been done through instrument trial. The trial took place at two schools, SDN Banyubiru 04 and SDN Banyubiru 05 and took sample from the fifth grade students. Total of students that became subject in doing this reasearch was 43 students. The result was analyzed as an advice to do more improvement on instrument before it was tested directy on the test run. Analysis of construct validity data on test run used Rasch model with Winsteps software.

RESULT AND DISCUSSION

Instrument development in arranging the spesification of measurement instrument has been arranged in the form of blue print. Blue print consisted of the purpose of meausrement, instument clues, instrument form, total of instrument points. "Kompetensi Inti" (KI) 4 and "Kompetensi Dasar" (KD) 4.1 were competence spesialized on skills. Therefore, the developed instrument was instrument to evaluate the ability to think and work creatively. Creating the comic such as product observing instrument and working performance. product observing instrument was to measure figural ability, while working performance was interpeted as media to measure verbal dimension ability.

The capability to act and to think creatively in representating the story was assessed by two aspects, figural and verbal aspect. Characteristic of creative thinking ability covered fluency, flexibility and originality. Test that could be applied to measure it was story writing test (Rhosalia, Laksono,& Sukartiningsih, 2016, p.167).

Figural aspect was measured using 9 items. Verbal aspect consisted of two aspects, script writing and story telling. Script writing aspect was measured using 9 items, while story telling aspect used 5 items. Number of items used to measure the whole three aspects was 22 items. The clues were transformed to instrument points in the form of check list with Likert scale. Results of the research discussing about total of likert scale response point showed that number of 5,7,11 was compatible (Budiaji, 2013, p.131). scale range used in this research showed that from 1 until 5 was moving from the most negative to the most positive.

Validity used in developing this instrument was content validity and construct validity. To know content validity, instrument was examined upon the truth of a concept, writing technic, and language used. This examination has been done by three proffesionals based on their background. The result of content validity by the expert was measured using SPSS software. Result from experts took reliability analysis using Hoyt formula by two ways of anova.

Computacy result (with SPSS application) presented the data in table 1. Computacy of reliability assessment among three raters got coefficient number of 0.78. Assessment from the experts could be reliable if reliability coefficient was more that 0.70. this statement matched with Wells & Wollack's opinion that said reliability coefficient for test used by teachers should be at least 0.70 or more (Azwar, 2017, p.98). According to Khumaedi (2012, p.29), number of reliability coefficient of research instrument on education that could be accepted was based on the policy from the researcher it self. The highest number of reliability coeficient was, the more consistent the result got if it was applied repeatedly.

Table 1. Result of computacy using two ways of anova.

Tests of Between-Subjects Effects

Dependent Variable:Skor Penilaian

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	55.773[a]	65	.858	.	.
Intercept	1129.227	1	1129.227	.	.
R	4.364	2	2.182	.	.
B	18.439	21	.878	.	.
R * B	32.970	42	.785	.	.
Error	.000	0	.		
Total	1185.000	66			
Corrected Total	55.773	65			

a. R Squared = 1,000 (Adjusted R Squared = .)

Analysis result with Hoyt formula could be calculated by

$$r_{xx'} = 1 - \frac{S^2r}{S^2s} = \frac{0,755}{3,470} = 0,782 = 0,78$$

$r_{xx'}$ = inter rater reability

S^2r = residual variant which is squared mean which has interaction among items and subjects.

S^2s = Subject variants that is squared mean among subjects (*rater*)

The assessment by three experts could be said reliable if the reliability coeffcient was more than 0.70. In other words, it could declare that assessment by three experts was reliable. It matched to Wells & Wollack's opinion that realibility coefficient for test used by teachers should be at least 0.70 or more (Azwar, 2017, p.98).

The result of instrument improvement upon the suggestion and decision of experts was tested using 43 samples of students. Total of samples was gained from two schools, 25 students were samples from SDN Banyubiru 05 and 18 students became samples from SDN Banyubiru 04. The test run was conducted on 13th and 14th of May 2019. In first session, students were given the chance to make a comic in 2 hours of learning session (70 minutes). After that, they proceeded to present their stories in front of class.

Construct validity was done by using Rasch Model of dimensionality analysis with Winsteps software to gain information about factors that could create a dimension. Dimensionality test were included in item function : dimensionality (Sumintono, 2016, p.12). Refers to dimensionality analysis, it was known that 22 tested items created 2 factors of dimension; figural and verbal factors. Output result of dimensionality analysis informed that there were two verbal items belonged to figural group: V9 and V10 item.

Item Fit Order analysis aimed to find out unfitted item criteria. Item Fit Order informed whether an item worked normally in a measurement or not. The unfitted ones would appear to the top of the parts. Criteria that used to decide the level of matching point was outfit means square, outfit Z standard, and point measure correlation. An item was consider as fit if the point from Outfit Means Square was stated at 0.5<MNSQ<1.5. number of ZSTD was stated around -2. 0<ZSTD<+2. 0. The number of Point Measure Correlation stated at around 0.4 < Pt Measure Corr < 0.85.

Table 2. Item Fit Order

```
TABLE 10.1 C:\Users\user\Desktop\DATA SKALA KECI ZOU3/4WS.TXT  May 30 15:38 2019
INPUT: 43 Person  22 Item  REPORTED: 43 Person  22 Item  5 CATS    WINSTEPS 3.73
--------------------------------------------------------------------------------
Person: REAL SEP.: 3.23  REL.: .91 ... Item: REAL SEP.: 3.81  REL.: .94

            Item STATISTICS:  MISFIT ORDER
```

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PT-MEASURE CORR.	EXP.	EXACT OBS%	MATCH EXP%	Item
19	120	43	.72	.19	3.20	6.8	3.17	6.7	A .75	.64	4.7	46.3	V19
2	170	43	-1.07	.19	1.04	.3	1.46	1.7	B .51	.60	55.8	47.9	F2
22	137	43	.10	.19	1.22	1.0	1.26	1.2	C .65	.65	46.5	45.7	V22
16	103	43	1.38	.20	1.21	1.0	1.18	.9	D .38	.61	55.8	49.0	V16
5	173	43	-1.19	.20	.93	-.3	1.16	.7	E .45	.59	53.5	48.0	F5
20	95	43	1.72	.21	1.13	.6	1.16	.8	F .32	.60	62.8	50.2	V20
1	173	43	-1.19	.20	.90	-.5	1.13	.6	G .58	.59	53.5	48.0	F1
10	153	43	-.46	.19	1.02	.2	1.01	.1	H .66	.64	48.8	44.4	V10
9	159	43	-.67	.19	1.00	.1	1.02	.2	I .69	.63	53.5	48.1	V9
3	136	43	.14	.19	.99	.0	1.01	.1	J .41	.65	41.9	46.1	F3
21	134	43	.21	.19	.90	-.4	.94	-.2	K .70	.65	51.2	46.3	V21
14	145	43	-.18	.19	.88	-.5	.85	-.6	k .75	.65	55.8	45.3	V14
12	143	43	-.11	.19	.86	-.6	.82	-.8	j .74	.65	48.8	45.8	V12
13	129	43	.39	.19	.83	-.8	.82	-.8	i .77	.65	39.5	46.1	V13
4	154	43	-.49	.19	.83	-.8	.78	-.8	h .56	.64	60.5	44.6	F4
17	116	43	.87	.19	.78	-1.1	.73	-1.3	g .65	.63	53.5	47.5	V17
18	140	43	.00	.19	.71	-1.5	.74	-1.2	f .72	.65	58.1	46.4	V18
8	127	43	.46	.19	.73	-1.3	.72	-1.4	e .52	.64	55.8	46.5	F8
11	150	43	-.35	.19	.70	-1.6	.66	-1.7	d .79	.64	55.8	44.6	V11
6	173	43	-1.19	.20	.64	-1.9	.57	-1.9	c .76	.59	55.8	48.0	F6
15	110	43	1.10	.20	.58	-2.3	.54	-2.6	b .76	.62	62.8	47.3	V15
7	145	43	-.18	.19	.53	-2.7	.52	-2.7	a .75	.65	60.5	45.3	F7
MEAN	140.2	43.0	.00	.19	.98	-.3	1.01	-.2			51.6	46.7	
S.D.	22.0	.0	.81	.01	.52	1.8	.53	1.9			11.8	1.5	

Analysis from Item Fit Order gave information that whole item of Oufit means square was stated at 0.5 < MNSQ < 1.5 around 0.52 at F7 item until 1.46 at F2 item except V19 item that had the number of outfit MNSQ more than 1.5 or it was around 3.17. Outfit Z standard was stated around -2. 0 < ZSTD < +2. 0. There were three items not included in ideal number; -2.7, 2.6, and 6.7. Point measure correlation was stated at point around 0.4< Pt Measure Corr < 0.85 except two items which were V16 and V20 that had 0.38 and 0.32 of Point measure correlation.

Summary Statistic analysis result was about the summary of statistic person and item. Information gained from table 2 was the quality of all students response, instrument quality, and interaction between person and point. Detailed information from table 2 was about a) Person Measure mean that showed average all response number from students toward the item point. The average that was lower that logit 0. 0 meant that ability of students was lower than item difficulty level, b) Alpha Cronbach number measured reability level between person and the whole item. The category for Alpha Cronbach : <0.5 : Very bad, 0.5-0.6 : bad, 0.6-0.7 : good, 0.7-0.8 : very good, >0.8 : excellent, c) Person number of reliability and item reliability was marked with coefficient : <0.67 : low, 0.67-0.80 : enough, 0.80-0.90 : high, 0.91-0.94 : very high, d) infit MNSQ and outfit MNSQ in person table had ideal number at 1 (the closer the point was to one, the better it was). While for infit ZSTD and outfit ZSTD for person and point tables had ideal point at 0. 00 which meant that the closer it got to 0. 00, the person quality and item would be better, e) separations number showed the classification of person and point. The quality of instrument would get better if the number of separations was higher then it could identify respondent and point groups (Sumintono & Widhiarso, 2015, p.85). the result of analysis in Summary Statistic could be seen at table 3.

Table 3. Ouput analysis of Summary statistic.

```
TABLE 3.1 C:\Users\user\Desktop\DATA SKALA KECIL ZOU374WS.TXT  May 30 15:38 2019
INPUT: 43 Person  22 Item  REPORTED: 43 Person  22 Item  5 CATS    WINSTEPS 3.73
-------------------------------------------------------------------------------

      SUMMARY OF 43 MEASURED Person
--------------------------------------------------------------------------------
|          TOTAL                      MODEL        INFIT       OUTFIT        |
|          SCORE     COUNT   MEASURE  ERROR    MNSQ   ZSTD   MNSQ   ZSTD     |
|--------------------------------------------------------------------------|
| MEAN      71.7     22.0      .55     .27      .97   -.3    1.01   -.2      |
| S.D.      14.3      .0      1.00     .02      .60   1.7     .68   1.8      |
| MAX.     101.0     22.0     2.79     .36     3.18   5.3    3.48   5.7      |
| MIN.      46.0     22.0    -1.36     .25      .40  -2.5     .44  -2.3      |
|--------------------------------------------------------------------------|
| REAL RMSE   .30 TRUE SD   .96  SEPARATION  3.23  Person RELIABILITY  .91  |
|MODEL RMSE   .27 TRUE SD   .97  SEPARATION  3.56  Person RELIABILITY  .93  |
| S.E. OF Person MEAN = .15                                                 |
--------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .93

      SUMMARY OF 22 MEASURED Item
--------------------------------------------------------------------------------
|          TOTAL                      MODEL        INFIT       OUTFIT        |
|          SCORE     COUNT   MEASURE  ERROR    MNSQ   ZSTD   MNSQ   ZSTD     |
|--------------------------------------------------------------------------|
| MEAN     140.2     43.0      .00     .19      .98   -.3    1.01   -.2      |
| S.D.      22.0      .0       .81     .01      .52   1.8     .53   1.9      |
| MAX.     173.0     43.0     1.72     .21     3.20   6.8    3.17   6.7      |
| MIN.      95.0     43.0    -1.19     .19      .53  -2.7     .52  -2.7      |
|--------------------------------------------------------------------------|
| REAL RMSE   .20 TRUE SD   .78  SEPARATION  3.81  Item  RELIABILITY  .94   |
|MODEL RMSE   .19 TRUE SD   .78  SEPARATION  4.06  Item  RELIABILITY  .94   |
| S.E. OF Item MEAN = .18                                                   |
--------------------------------------------------------------------------------
UMEAN=.0000 USCALE=1.0000
Item RAW SCORE-TO-MEASURE CORRELATION = -1.00
946 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 2123.04 with 879 d.f. p=.0000
Global Root-Mean-Square Residual (excluding extreme scores): .7853
```

Output analysis of variable maps showed the result of spreading map on difficulty level item and students response. The easiest item that could be reached by students was F1, F2, F5, F6 items. The hardest one was V 20. Students with the lowest ability was number 6A while students with highest ability were students from 35B. reliability item index and separation item showed whether the used points could classify the students or not (Sumintono & Widhiarso, 2015, p.113-117).

Table 4. Misfitted Item based on analysis indication item fit order

Item	0.5<MNSQ<1,5	Fit/Misfit	-2,0<ZSTD<2,0	Fit/Misfit	0,4<PtMC<0,85	Fit/Misfit	Used /Unused
F1	1.13	fit	0.6	fit	0.58	fit	Used
F2	1.46	fit	1.7	fit	0.51	fit	Used
F3	1.01	fit	0.1	fit	0.41	fit	Used
F4	0.78	fit	1.0	fit	0.56	fit	Used
F5	1.16	fit	0.7	fit	0.45	fit	Used
F6	0.57	fit	-1.9	fit	0.76	fit	Used
F7	0.52	fit	-2.7	fit	0.75	fit	Used
F8	0.72	fit	-1.4	fit	0.52	fit	Used
V9	1.02	fit	0.2	fit	0.69	fit	Used
V10	1.01	fit	0.1	fit	0.66	fit	Used
V11	0.66	fit	-1.7	fit	0.79	fit	Used
V12	0.82	fit	-0.8	fit	0.74	fit	Used
V13	0.82	fit	-0.8	fit	0.77	fit	Used
V14	0.85	fit	-0.6	fit	0.75	fit	Used
V15	0.54	fit	-2.6	fit	0.65	fit	Used
V16	1.18	fit	0.9	fit	0.38	misfit	Used
V17	0.73	fit	-1.3	fit	0.65	fit	Used
V18	0.74	fit	-1.2	fit	0.72	fit	Used
V19	3.17	misfit	6.7	misfit	0.75	fit	Unused
V20	1.16	fit	0.8	fit	0.32	misfit	Used
V21	0.94	fit	-0.2	fit	0.70	fit	Used
V22	1.26	fit	1.2	fit	0.65	fit	Used

Notes: F= figural aspect, V= verbal aspect

According to table 2, V19 item could not be used because it could not complete two of three fit item criteria. Outfit means square point was more than 1,5. The number of outfit Z standard was more than 2. 0. V16 item could be used because it still fulfilled two of three fit item criteria.

Validated measurement instrument and fulfilled reability coefficient would proceed to be tested in small scale. Total of sample was 43 people from two schools. According to Widoyoko (2017, h.146), minimum sample needed for the instrument test was 30 people.

After obtaining the measurement result, the next step was to analyze the measurement result using Rasch Model analysis. This model analyzed each point that had different quality (Sumintono & Widhiarso, 2017, p.106). the output of Rasch

politomy modeling could visualize tested item quality.

Reability test and emphirical instrument validity on field used analysis data from Rasch model for statistic test. This aimed to identify whether there was invalid factors or points among the data although it was valid in terms of the construct (Widoyoko, 2017, p.146). emphirical validity (factor validity) was analyzed using dimensionality analysys method. While, emphirical validity (point validity) was analyzed using item fit order analysis on Winsteps application.

Winsteps software could check unidimensionality item toward item menu function: dimensionality. Accurated item or point could be identified using model (infit-outfit) and measurement location could be seen on item menu: fit order (Sumintono, 2016, p.12). Rasch Model data analysis

used 1 logistic parameter (1PL) which was item difficulty level. Distinguished ability and guessing parameter were considered the same. To get information about item difficulty level could be seen on output table on item measure menu. The point was considered fit if the level of item is fitted on outfit means square $0.5<MNSQ<1.5$ and outfit Z standard was stated around $-2,0 <ZSTD<+2,0$ (Sumintono & Widhiarso, 2015, p.98).

The suitability of response pattern could be seen on output table and pick person fit order menu. Response pattern was considered fit it the point of outfit means square was $0.5< MNSQ < 1.5$, the acceptable number of outfit Z standard (ZSTD) was $-2.0< ZSTD<2.0$, Point Measure Correlation (Pt Mean Corr) number was $0.4 < Pt Mean Corr < 0.85$. (Sumintono & Widhiarso, 2015, p.81).

To know item refraction on winsteps program could be seen on item menu: **DIF, between/within.** DIF (Differential Item Functioning) was refraction point in measuring. in psychometric, refraction item is an item affected by differential item function. Item that has P number (PROB) under 0.05 showed that the item is affected by DIF (Sumintono, 2016, p.14).

Analysis on Summary Statistic declared that Person Reliability coeffisien number was 0.91 and item reliability was 0.94. it referred to coeffisien category :< 0.67: low, 0.67-0.80: enough, 0.80-0.90: good, 0.91-0.94: very good, it could be concluded that if the number of person coefficient and item between 0.91-0.94 had very well category. Either consistency answer from students or the quality of points from item has very well reliability aspect instrument.

Alpha Cronbach number that measured the whole reability interaction between person and item was 0.93. based on category for Alpha Cronbach: <0.5: very bad, 0.5-0.6: bad, 0.6-0.7: good, 0.7-0.8: very good, >0.8: excellent. The conclusion was number of Al/pha Cronbach was classified excellent. It meant that interaction between person and measurement item was categorized excellent. Separation number was 3.81 calculated with formula. The result was the higher separation number was, the better instrument quality was

because it could classify respondent and point group more detail (Sumintono & Widhiarso, 2015, p85). There were five strata items from the easiest to the hardest item to fulfill.

Rasch model analysis explained that measurement instrument of act and think creatively could proceed to another research with bigger sample. Item that was well-catagorized for visual aspect: proportion or size among visual component, fluency in sketching the picture, to match the title to theme, attracting title, composition or stability of picture and field, detail on each part of picture, the originality of picture, flexibility of picture.

Well-conditioned item for verbal aspect : to arrange the script based on the theme, the originality of writing idea/uniqueness, to arrange the script as ilustrated image, to put coherence between the script, the choice of words, total of row in each script, complete sentence structure, fluency in delivering the story, flexibility of body language when telling the story, to tell the story with their own language fluently, and to tell the sentence in detail.

CONCLUSION

The result of instrument informed that quality of an item reliability had well-category. Yet, based on factor analysis, it still needed to do improvement and perfection toward measurement instrument. Instrument itself was categorized as two main dimensions, those were figural and verbal dimension. Figural dimension consisted of 10 measurement items, while verbal dimension had 12 measurement items. Instrument were arranged from the lowest logit to the highest logit.

SUGGESTION

To gain valid and reliable instrument, it was a must to do testing on field with bigger sample. Item arrangement can be ordered based on the difficulty level from the easiest ones to the hardest ones based on logit number in each dimension.

REFERENCES

Allen, M.J & Yen, W.M. 1979. Introduction to Measurement Theory. California: Brooks/ Cole Publishing Company

Appulembang, Yeni Anna. 2017. "Norma Kreativitas Menggunakan Torrance Test Of Creativity Thinking Untuk Anak Usia 6-12 Tahun".Provitae Jurnal Psikologi Pendidikan, 9 (1): 41-57

Azwar, Saifuddin. 2017. Reliabilitas dan Validitas. Yogyakarta: Pustaka Pelajar.

Budiaji, W. 2013. "Skala Pengukuran Dan Jumlah Respon Skala Likert". Jurnal Ilmu Pertanian dan Perikanan. 2 (2): 127-133

Fuadi, Sumaryanto, T. & Lestari, W. 2015. "Pengembangan Instrumen Penilaian PsikomotorPembelajaran IPA Materi Tumbuhan Hijau Berbasis StarterExperiment Approach Berwawasan Konservasi" Journal of Educational Research and Evaluation, 4 (1): 1-11.

Kementerian Pendidikan dan Kebudayaan. 2016. Dinamika Kurikulum/ Perubahan KI & KD. Jakarta: Direktorat Pembinaan Sekolah Dasar.

Khumaedi, Muhammad. 2012."Reliabilitas Instrumen Penelitian Pendidikan". Jurnal Pendidikan Teknik Mesin, 12(1): 25-30.

Majid, N.K, Raharjo, T.J,& Supriyadi 2017. "Pengembangan Instrumen Asesmen Otentik Unjuk Kerja pada Mata Pelajaran IPA di SDN Jlamprang dan SDN Wonosari 03 Kabupaten Batang" Journal of Educational Research and Evaluation,6 (1): 55-62.

Mardapi, Djemari. 2015. Pengukuran Pengukuran & Evaluasi Pendidikan. Yogyakarta: Nuha Medika.

Putra, P.D.A & Iqbal, M. 2014."Implementation of Digital Comic to Improve Creative Thinking Ability in Integrated Science Study". International Conference on Mathematic Science and Education 2014. Faculty of Mathematic and Natural Sciences Semarang State University, 71-74.

Rhosalia, L.A, Laksono, K., Sukartiningsih, W. 2016. "Kemampuan Berpikir Kreatif Dalam Menulis Naratif Siswa Kelas V Sekolah Dasar Negeri Di Kecamatan Gayungan Surabaya". Jurnal Review Pendidikan Dasar: Jurnal Kajian Pendidikan dan Hasil Penelitian. 2 (2): 166-174.

Rusilowati, Ani. 2017. Pengembangan Instrumen Pengukuran. Semarang: Unnes Press

Sumintono, B. & Widhiarso, W. 2015. Aplikasi Pemodelan Rasch Pada Assesment Pendidikan. Cimahi: Trim Komunikata.

Sumintono, B. 2016. "Aplikasi Pemodelan Rasch pada Asesmen Pendidikan: Implementasi Pengukuran Formatif (assessment for learning)". Research Gate: Conference Paper March 2016. publication at: https://www.researchgate.net/publication/298787148.

Widoyoko, E. P. 2017. Teknik Penyusunan Instrumen Penelitian. Yogyakarta: Pustaka Pelajar