COMMUNICATIONS IN HUMANITIES AND SOCIAL SCIENCES
Homepage: chss.kipmi.or.id

Validity and reliability: Social attitude assessment instruments

Bambang Yuniarto[a], Rivo Panji Yudha[b,*], Aulia Sholichah Iman Nurchotimah[a]

[a]*Postgraduate Program, IAIN Syekh Nur Jati Cirebon, Cirebon 45132, Indonesia*
[b]*Postgraduate Program of Early Childhood Education, Universitas Pancasakti Bekasi, Bekasi 17414, Indonesia*

Article history:
Received: 8 December 2021 / Received in revised form: 19 July 2022 / Accepted: 21 July 2022

Abstract

The purpose of this study is to develop an instrument for assessing social attitude competence whose constructs are valid according to experts and empirically valid and reliable based on field trials in Civics learning for high school students. This research was conducted through a quantitative descriptive approach at SMP Negeri 1 Gebang. Social attitude assessment instrument was used and it contained 28 items and an expert validation observation sheet. Design validation was done by reviewing the product design by three experts (expert judgment). The study was conducted to determine the feasibility of the instrument from the aspect of substance, construction, and language of the social attitude assessment instrument. Advice was given by experts after doing research to get approval, which was later used as a benchmark to state that the product was logically valid. The measurement of the content validity of social attitude statements was conducted using Likert scale and the reliability test used the Rasch Model approach. The results of the field test showed that the social attitude assessment instrument was valid in terms of content and constructs. The reliability test using the Rasch model approach showed that the product reliability was categorized as satisfactory. The final product is considered suitable for direct use by teachers and can be further developed with grid references.

Keywords: assessment instrument; social attitude; validity; reliability

1. Introduction

The Ministry of Education and Culture issued a new policy in education, i.e. changing the education unit level curriculum to the 2013 curriculum. The 2013 curriculum balances three aspects: psychomotor, affective, and cognitive. Hence, the assessment in the 2013 curriculum must also cover all three aspects. In the affective aspect, the attitude assessed consists of two attitudes, i.e. spiritual attitudes related to spirituality, and social attitudes. However, in practice, there are still many schools that have not been able to make learning activities that maximally develop social attitudes since so far attitude assessment activities have been carried out only based on observations (Tiara & Sari, 2019).

However, in this decade learning only focused on cognitive achievement has received many criticisms since it does not provide more benefits to the achievement of the function of national education. Law Number 20 of 2003 defines the function of national education, namely to develop competence in the context of educating the nation's life, forming a dignified national character and civilization with the aim of fostering the potential of students to be human beings of faith, very religious, fear of God Almighty, noble hearted, healthy, insightful, competent, creative, independent, democratic and responsible. It is clearly seen in the explanation that attitude learning plays a major role in shaping students based upon the function of national education. Learning objectives or cognitive domains are often contrasted with affective domains. Cognitive is associated with the synthesis, evaluation, and comprehension of knowledge or information, while affective is associated with attitudes, emotions, and feelings (Rimland, 2013). Currently, learning in schools concentrates on developing students' affective in addition to develop students' cognitive and psychomotor skills. Learning in schools is required to shape the attitudes of students in accordance with the noble values of the Indonesian nation.

The indications of the achievement of student competence in learning certainly must be supported by an assessment process in accordance with the aspects that need to be assessed. Making the right assessment is one of the skills that must be possessed by teachers in the classroom. The assessment given should be objective, accurate and cover the aspects of knowledge, skills, and attitudes. Such assessment requires an appropriate special instrument as a reference to measure the achievement of student competence. There are still teachers who have misconceptions or have not mastered well how to assess student attitudes, especially social attitudes. Many indicators of social attitudes that are assessed have made teachers find it difficult to make assessments. Even though some teachers have understood how to evaluate techniques, most of them still do not understand make social attitudes difficult.

It is proven by juvenile delinquency that has occurred

* Corresponding author. Tel.: +62 817-9073-668
Email: rivoyudha@yahoo.co.id
https://doi.org/10.21924/chss.2.1.2022.22

recently, ranging from mild cases to serious ones such as murder. The lack of inculcation of attitudes in school has resulted in the emergence of several cases of juvenile delinquency. Cultivation of attitude especially in junior high schools necessarily needs to concern that at the junior high school level, the character and attitudes of students are built.

One of the benchmarks for learning success can be seen from the assessment results. Regulation of Minister of Education and Culture Number 23 of 2016 stated that assessment refers to the process of collecting and processing information to measure the achievement of student learning outcomes. Furthermore, the assessment is used to improve the learning process and acts as material for preparing reports on student learning progress. The 2013 curriculum uses authentic assessment - an assessment focused on the development of students' attitudes, knowledge, and skills. However, in particular, the assessment in the attitude domain is considered difficult. The main difficulty found in the assessment of attitudes is in terms of their scoring (Suryani, 2016).

The domain of attitude is very important to be owned and developed since an early age by students, including scientific attitudes to make it embedded in adulthood. Pala (2012) (in Musyarofah, et al, 2013) stated that character education can be started at every grade level. This is important for a strong foundation today and for strengthening and building foundations for the future (Musyarofah, Hindarto, 2013).

Such assessment requires an appropriate special instrument as a reference to measure the achievement of student competence. Observations and interviews conducted at SMPN 1 Gebang, Cirebon Regency showed that the school has implemented the 2013 curriculum in all classes from grade VII to grade IX. The SMP mentioned above, in the implementation of learning has not been able to develop an instrument for assessing social attitudes, especially disciplinary attitudes to the fullest.

The reference used to determine the level of success in achieving student competence is referred to as a scoring guide. Kunandar (2015) explained the scoring guideline or rubric as a guide or hint that explains the limits or key words for scoring questions in the form of descriptions and the answer criteria used to score both non-objective and subjective description questions. Teachers can use guidelines to correct students' work or answers more accurately and avoid subjectivity (Kunandar, 2015).

The assessment instrument is required to fulfill several requirements of the assessment instrument. One of the conditions that must be met is validity. Validity is the ability of the assessment instrument to assess what will be assessed (Anisah, 2018).

According to Suryabrata (2014), the most important characteristics that must exist in the assessment instrument are validity and reliability (Sumardi Suryabrata, 2014). Meanwhile, Rivo (2020) stated that the validity of the assessment instrument consisted of content and construct validity (Yudha, 2020). Referring to this, researchers have developed alternative instruments to assess the competence of social attitudes that are logically valid, content and construct valid, and reliable. This assessment instrument can also be further developed by the teacher.

A test instrument is seen to be good if it is able to provide appropriate information regarding the student's abilities for the competencies being tested. In 2016, Susongko conducted a research related to the validation model of test questions, namely the validity of Messick including several aspects such as content, structural, substantive, consequential and external aspects. This validity was analyzed using the Rasch model supported by the Winsteps software. Other research related to test theory analysis was conducted by Kustriyono (2004) who used qualitative analysis to determine the characteristics of the test being tested in terms of material, construction and language (Kustriyono, 2004).

In this study, measurements were carried out using the social attitude instrument. This instrument was specifically designed to measure students' social attitudes on civic education subjects. The social attitude instrument was designed by considering the construct component and the descriptive component. The construct component is that the items must provide an indication of the answers that will be obtained, from low to high stages. In this study, the items would produce answers showing a certain qualitative range, namely five alternative answers. Hereinafter referred to as the response pattern politomy. The second component is a descriptive component, which is a component that explains several points on certain aspects. In this study, the aspect considered was how students' social attitudes to understand civics education learning and then analyzed using the Rasch model approach.

This paper aims to use the Rasch model approach to determine the quality of the assessment tool to determine students' social attitudes in Civics learning. This quality was measured by several metrics, namely stated items and item reliability according to the Rasch model. Therefore, an instrument was then designed based on the Rasch model to determine which items were suitable and which ones were not. In addition, with the help of the Winstep software, the Cronbach alpha value was determined for the reliability of the project.

2. Methodology

This quantitative research involved 100 students of SMPN 1 Gebang class VII as the research sample. Quantitative analysis was carried out by focusing on the numerical analysis of the data collected (Zanuddin, et al., 2021). This research is part of the research on the development of the HOTS assessment in social attitude learning. The instrument used was a description of 28 items.

The instrument was designed and contained 28 items, which were consulted with experts to later prove its empirical validity and content validity. Expert test or validation was carried out with respondents from instrument or product design experts. This activity was carried out for initial product reviews to provide input for further instrument improvements. The instruments that have been produced were then evaluated on whether the format produced is feasible, and how appropriate the content of the learning assessment material is. If the instrument is not feasible, then the instrument is revised to be feasible to be tested. More complete instrument design for work and expert judgment validation sheet.

The validity of the measurement results of the instruments used was carried out through validation by experts (expert

judgment). Expert judgment is a technique for validating or assessing the feasibility of a product design. In this activity, an evaluation is carried out by experts in their fields. The suggestions given are used to improve the materials and learning designs that have been prepared.

The way to prove the validity of educational instruments according to RP Yudha (Yudha, 2020) can be divided into two, namely the expert judgment and field trials. Testing the validity of the instrument is carried out through a panel of experts. It is called as non-empirical validity testing including appearance test, completeness test and conformity test. While the evidence of validity carried out by field trials is called as evidence of empirical validity. Empirical trials use data sources similar to the data sources or respondents to be studied. The types of empirical validity include the difficulty test of discriminant test items and factor analysis.

This validity is estimated through testing the feasibility or relevance of the test content through rational analysis by a competent panel or through expert judgment (Azwar, 2012). To find out whether the test is valid, it must be done through a review of the test grid. To ensure that the test questions represent or reflect the overall content or material should be mastered proportionally. Therefore, the content validity of a test does not have a certain magnitude calculated statistically, but it is understood that the test is already valid based on the examination of the test grid. Therefore, Wiersma and Jurs in (Djaali & Muljojo, 2008) stated that content validity is actually based on logical analysis, so it is not a statistically calculated validity coefficient.

One Citizenship Lecturer experts and two evaluation experts, two experts from lecturers and one expert from senior Citizenship teachers were asked to validate the social attitude instrument. This validation included evidence related to content, and construct validation. Content validity analysis was carried out by analyzing the results of content validation by experts using the Rasch model approach. This was because using the Rasch model was viewed as a solution to the problem of validity where the Rasch model is able to provide statistics and offers an opportunity to investigate the validity of test instruments based on the responses of research subjects. The results of the validity can be seen and analyzed with the Winsteps program in the Out fit order table to see the suitability of items that function in the normal category to be used as a measurement of student misconceptions.

In this review, the assessment of the dependability of the instrument utilizing the Rasch model methodology was utilized to gauge the unwavering quality of the test instrument. The instrument utilized acted as questions that have been created and dissected utilizing the Rasch model. The investigation was done by using the Winsteps application to decide the appropriateness of the things with the estimation model utilized, specifically the Rasch model. The circumstances that should be met in the Rasch model were the reasonableness of the evaluation instrument model with the INFIT MNSQ acknowledgment breaking point of 0.77 to 1.30 (Adam & Khoo, 1996).

The results of the reliability analysis can be viewed using the Winsteps program in the Summary Statistics table. Tables can provide general information about the quality of student response patterns, the tools used, and the relationship between students and items.

3. Results and Discussion

3.1. Expert judgment

As seen in Table 1, the analyst chose three specialists according to the alternate points of view with various rules in light of the desires of the scientist yet homogeneous as per their inclinations and their relationship to factors to be approved from scholastics, experts, and content, to track down the chose factors. From the three specialists, remarks or ideas were obtained as sentences on research factors, expansion and deduction of the quantity of factors, and information handling.

Table 1. Expert general information

Expert	Knowledge Field	Educational Background
I	Civic Education	Doctor
II	Assessment	Doctor
III	Civic Education	Bachelor of Education

Table 2. Assessment instrument validation results

Rated aspect	Number Item	Expert 1	Expert 2	Expert 3	Average	Aspect Average	Presentase
Aspect conformity with indicators	1	5	5	5	15	0,97	97%
	2	4	5	5	14		
	3	4	4	5	13		
Writing Procedure	4	4	4	4	12	0,82	82%
	5	4	4	4	12		
	6	4	4	4	12		
Language Usage	7	5	4	4	14	0,86	87%
	8	5	4	4	13		
Physical appearance	9	5	5	4	14	0,87	87
	10	4	4	4	12		
44	43	131			0,873	87,3%	

It then came to the specialists who met the prerequisites of scientists. The consequences of the evaluations of the three assessors of the social mentality appraisal instrument in metro schooling learning are summed up in Table 2. Parts of the evaluation included congruity of appraisal angles with existing pointers, similarity with markers, composing, language, and parts of actual appearance.

Table 2 shows that the overall approval consequences of the three social mentality evaluation instrument assessors should be visible from the master judgment scores, which get the most elevated typical incentive for the part of congruity with the markers with a level of 97%.

There are four separate aspects to the validation of performance assessment instruments including the suitability of aspects with indicators, aspects of writing, aspects of language, and aspects of appearance. It is expected that validators will vary in their abilities. This examination purposely is to reliably distinguish according to experts whether the instrument is feasible for use in the field.

A portion of the primary concerns that became input from the four assessors included 1) the strategy in composing language is not slick and right, for instance in consolidating or isolating sentences; 2) some things were not in accordance with the signs of social mentality; 3) the language in the proclamation thing is not vague yet; and 4) the instrument must measure explicit skills according to a social perspective. Furthermore, the above inputs were used to revise the social attitude assessment instrument for further confirmation to the validator as the input provider. The results of the validation of the three experts stated that the social attitude assessment instrument was feasible to use.

Figure 1 shows an outline of the consequences of expert judgment examination of social demeanor instruments, which were broken down utilizing Winsteps programming. It was then found in the result rundown about the things and the subjects. This test showed a detachment of 1.89 and a thing unwavering quality of 0.88. This implied that the test has worked very well for having a number of different degrees of trouble. While the subject just had a partition of 2.12 and dependability of 0.82 individuals. This implied that these subjects were exceptionally fluctuated in light of the fact that they just had many capacities.

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	109.3	27.7	1.03	.28	1.01	-.09	1.02	-.07
SEM	1.4	.3	.07	.01	.05	.17	.05	.17
P.SD	14.0	2.7	.68	.08	.53	1.66	.53	1.68
S.SD	14.1	2.7	.69	.08	.53	1.67	.53	1.69
MAX.	129.0	28.0	2.79	.96	1.03	1.02	1.04	1.09
MIN.	3.0	1.0	-.32	.19	.00	-2.11	.00	-2.97

REAL RMSE	.32	TRUE SD	.60	SEPARATION 1.89	PERSON RELIABILITY .88
MODEL RMSE	.29	TRUE SD	.62	SEPARATION 2.12	PERSON RELIABILITY .82
S.E. OF PERSON MEAN = .07					

PERSON RAW SCORE-TO-MEASURE CORRELATION = .74
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .95 SEM = 3.26

Fig. 1. Level of appropriateness of items

The mean measure esteem was 1.03. The mean worth more prominent than 0 showed that the propensity of the subject's capacity was more noteworthy than the degree of trouble of the inquiries. Cronbach's Alpha worth (KR-20) is a dependability coefficient determined in view of the traditional test hypothesis approach. This worth is the connection between the individual and the thing in general. Alpha worth was 0.95 indicating that the unwavering quality of the test overall was still extremely agreeable. The ideal incentive for INFIT and OUTFIT MNSQ was close to 1, while the ideal incentive for INFIT and OUTFIT ZSTD was close to 0. For the individual and thing tables, the mean qualities for INFIT and OUTFIT MNSQ and INFIT and OUTFIT ZSTD were highly closer. This implied that all things were as per the Rasch model and can be utilized as an evaluation instrument. On the other hand this can be deciphered that the things utilized have been cautiously ready to evaluate the respondents' responses, according to the development of fundamentalism.

3.2. Content validity

After the expert validation phase was complete, approximately 100 students of small class VII surveyed were tested on equipment to assess their social character in small learning. The content was analyzed purposely to find out if the instrument was empirical or practical in terms of content to know what the student's social character was. The trial results of Class VII equipment would be helpful for further development and improvement of the equipment.

Figure 2 shows the Winsteps output of Wright's map about the value of the student's ability level in the assessment of social attitudes. The data showed that seven students had the highest ability, while six students had the lowest one.

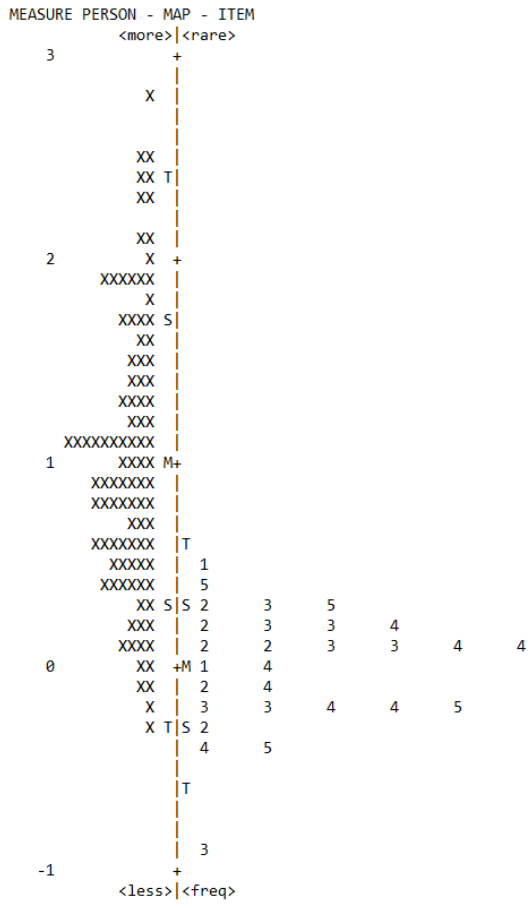

Fig. 2. Item wright map

Figure 3 shows the distribution of items that did not fit or were considered non-fitting to the model. Item limits were declared as model-aware if either or both of the following conditions were met: The first requirement is that the costume's MNSQ score is between 0.5 and 1.5; costume ZSTD values are

between -2.0 and 2.0 and the item correlation value (point measure correlation) with the total score is between 0.4 and 0.85 (Sumintono & Widhiarso, 2015).

As shown in Figure 3, based on the analysis results of the social attitude assessment tool using Winsteps Version 4.4.5 program, 28 items were obtained for non-matching items and 28 items for matching items, and the final means became 28 items.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
13	373	99	.34	.13	1.92	4.34	1.50	2.00	A .17	.43	40.4	53.8	3
4	388	99	.07	.14	1.45	2.27	1.48	1.92	B .35	.41	52.5	56.6	4
14	373	99	.34	.13	1.44	2.33	1.47	1.91	C .16	.43	54.5	53.8	5
1	400	99	-.18	.15	1.38	1.93	1.45	1.37	D .35	.39	53.5	58.8	4
12	405	99	-.30	.15	1.40	2.04	1.38	1.31	E .48	.39	59.6	59.2	2
7	408	99	-.37	.15	1.34	1.75	1.22	1.30	F .59	.38	53.5	59.2	4
2	401	99	-.20	.15	1.27	1.42	1.25	1.42	G .49	.39	56.6	58.8	3
3	388	99	.07	.14	1.19	1.05	1.19	1.10	H .54	.41	52.5	56.6	2
10	389	99	.05	.14	1.18	1.02	1.10	.61	I .57	.41	51.5	56.7	1
6	401	99	-.20	.15	1.16	.89	1.10	.61	J .62	.39	50.5	58.8	4
11	427	99	-.87	.17	1.12	.74	1.10	.70	K .38	.36	59.6	59.4	3
23	369	99	.40	.13	1.08	.55	1.11	.68	L .30	.44	46.5	51.5	5
9	403	99	-.25	.15	1.10	.57	1.01	.12	M .55	.39	60.6	59.0	5
19	383	99	.16	.14	1.03	.24	1.03	.22	N .41	.42	59.6	55.9	3
8	400	99	-.18	.15	.98	-.05	.92	-.43	n .53	.39	59.6	58.8	3
5	411	99	-.44	.16	.93	-.34	.87	-.75	m .60	.38	66.7	59.4	5
17	386	99	.11	.14	.85	-.84	.93	-.37	l .38	.41	59.6	56.4	4
21	393	99	-.03	.14	.89	-.58	.88	-.69	k .27	.40	58.6	57.9	4
16	394	99	-.05	.14	.83	-.91	.84	-.92	j .35	.40	59.6	57.9	2
22	388	99	.07	.14	.83	-.92	.83	-1.02	i .39	.41	64.6	56.6	3
28	368	100	.47	.12	.67	-2.24	.76	-1.52	h .41	.44	56.0	50.5	1
15	384	99	.14	.14	.75	-1.47	.74	-1.61	g .32	.42	58.6	55.9	2
18	383	99	.16	.14	.73	-1.62	.68	-1.06	f .48	.42	62.6	55.9	2
27	375	99	.30	.13	.62	-2.50	.70	-1.14	e .29	.43	65.7	53.9	2
20	394	99	-.05	.14	.66	-2.10	.64	-1.33	d .33	.40	70.7	57.9	4
25	382	99	.18	.14	.61	-2.53	.65	-1.26	c .38	.42	66.7	55.6	4
26	382	99	.18	.14	.60	-2.58	.58	-1.88	b .42	.42	69.7	55.6	3
24	386	99	.11	.14	.55	-3.01	.53	-1.30	a .37	.41	69.7	56.4	3
MEAN	390.5	99.0	.00	.14	1.02	.0	1.03	.0			58.6	56.7	
P.SD	13.4	.2	.28	.01	.32	1.8	.37	2.0			7.0	2.3	

Fig. 3. Item fit order

Figure 4 shows the measurement information obtained from the instrument for assessing students' social attitudes in civic education subjects. The x-axis shows the level of students' ability in the assessment of a given social attitude, while the y-axis shows the value of the information function. Based on the graph, at the medium ability level, the information obtained by the measurement was very high so that the development of an instrument for assessing student social attitudes in actuarial citizenship education subject is suitable or optimal if it is used for students with moderate abilities.

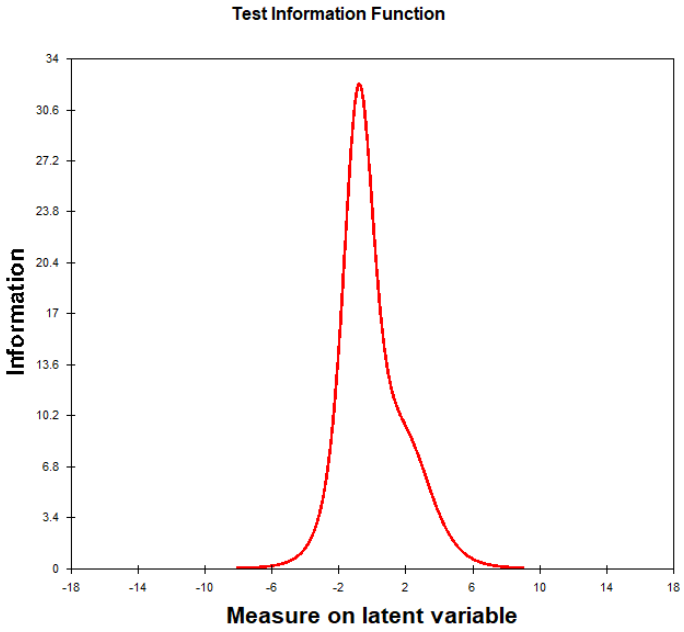

Fig. 4. Test information function

Nurhadi (2014) in his research produced a spiritual and social attitude assessment instrument that was logically valid and had a reliability coefficient value of 0.878, which was categorized as very good. In the study, it was stated that the product was appropriate to be used as a tool to measure student attitudes in learning based on divine values and love for the environment. Reviewing the statement, it is stated that the development product can be judged to meet the eligibility criteria as an instrument for assessing spiritual and social attitudes on the subject of vision and optical devices in junior high school science learning (Nurhadi et al., 2014).

Instruments of social attitudes certainly need to be made by concerning with the ease of use. The self-assessment sheet has been made in a format that makes it easier for teachers to use and develop it again. Teachers are given the convenience of managing student scores to obtain a specific picture of student attitudes. The score is given according to the response given by the student and then the number is calculated. Then, the number of student scores is calculated to be adjusted to the five-scale assessment criteria using the provided equation. Furthermore, the teacher can directly provide justification for adjusting the value obtained by students.

The use of the instrument also does not require a long time and a special place. Giving a response does not require special supervision, so students can do it anywhere and anytime according to an agreement with the teacher. In addition, the self-assessment sheet is also equipped with a grid containing indicators and the level of student attitudes being measured. This can make it easier for teachers who want to develop the product further. Through these factors, it can be said that the product is simple in use by both teachers and students.

Another advantage possessed by this social attitude instrument is that it is equipped with response instructions that use light language and are easily understood by students. This makes it easier for students to use the instrument properly. The use of instruments can provide opportunities for students to introspect themselves and find out their strengths and weaknesses. This can be an encouragement for students to motivate them to always improve themselves. The use of this instrument can also help teachers to provide more objective assessments and have more time for giving. The assessment process can be carried out at any time and place and does not require a long time. Even students can give responses at home; thus, it does not reduce student learning time in class.

However, this instrument also has its drawbacks. Some of the shortcomings of this instrument are that it is still limited to certain subjects, so it cannot be used for other topics. Instruments can be developed using universal topics in one subject for each semester. The use of universal attitude assessment instruments is considered more effective.

This instrument also does not have components that can be used as indicators of student honesty. Hence, it needs to be a companion instrument to obtain accurate data. The self-assessment sheet uses colored printed media, so it is considered less economical. This can be overcome by making a software-based self-assessment sheet that can be accessed by teachers and students.

4. Conclusion

The results of the validation of constructs by expert judgment from the three means of social attitude measurement should be read from the expert judgment score, which receives the highest average of the aspect of compatibility with the index

at a rate of 97%.. The result of the reliability of social attitude means using the Rasch model showed the Root-Means-squared Stats (Costume MNSQ)-value 1.02 displayed in the Individual; Item columns with an item score of 0.77 and personal reliability of 0.88 and Cronbach's alpha value of 0.95. There were 28 eligible items but other 28 items were not.

In conclusion it has been proven that the instrument in the form of a social attitude assessment sheet had valid, and reliable content and construct. The resulting instrument can be used as an alternative for teachers to assess the social attitudes of junior high school students in the subject of civic education. Self-assessment sheets can help teachers to provide assessments that are more objective, easy to use and require no much time in the learning process in class. This instrument can also be developed by the teacher based on the needs in the school.

Acknowledgements

The resulting instrument can be used as an alternative for teachers to assess the social attitudes of junior high school students on Civics subjects. Self-assessment sheets can help teachers to provide assessments that are more objective, easy to use and require no much time in the learning process in class. This instrument can also be developed by the teacher based on the needs of the school.

References

Adams, R. J. & Khoo, S-T. (1996). ACER Quest: the interactive test analysis system. [Computer software]. Version 2.1. Camberwell, Victoria: Australian Council for Educational Research. https://research.acer.edu.au/measurement/3/

Anisah, G. (2018). Validitas Instrumen Asesmen Menulis Cerpen Terintegrasi Pendidikan Karakter. *Al Ulya : Jurnal Pendidikan Islam*, *3*(1), 15–25. https://doi.org/10.36840/ulya.v3i1.148

Azwar, S. (2012). Reliabilitas dan Validitas edisi 4. Yogyakarta: Pustaka Pelajar.

Djaali, & Muljojo, P. (2008). Pengukuran dalam Bidang Pendidikan. Jakarta: Grasindo.

Kunandar. (2015). Penilaian Autentik (Penilaian Hasil Belajar Peserta Didik Berdasarkan Kurikulum 2013). Jakarta: Rajawali Press.

Kustriyono (2004). Penyusunan Perangkat Soal Ujian Akhir Mata Pelajaran Sains-Biologi SMP dalam Rangka Pengembangan Bank Soal. *Jurnal Penelitian dan Evaluasi Pendidikan*. 6 (2), 175-198, https://doi.org/10.21831/pep.v6i2.2048

Musyarofah, Hindarto N & Mosik (2013). Pendidikan Karakter Terintegrasi dalam Pembelajaran IPA Guna Menumbuhkan Kebiasaan Bersikap Ilmiah. *UPEJ Unnes Physics Education Journal*, *2*(2), 41-48. https://doi.org/10.15294/upej.v2i2.2665

Nurhadi, N., Rosidin, U., & Suana, W. (2014). Pengembangan Instrumen Penilaian Sikap Spiritual dan Sosial pada Pembelajaran IPA Terpadu. *Jurnal Pembelajaran Fisika Universitas Lampung*, 2 (4).

Sumardi Suryabrata. (2014). Metode penelitian. *Metode Penelitian*. Jakarta: Rajawali Press.

Sumintono, B., & Widhiarso, W. (2015). Aplikasi Permodelan Rasch pada Assessment Pendidikan. Cimahi: Trim Komunikata.

Suryani, I. (2016). Pengembangan Instrumen Penilaian Sikap Ilmiah Pada Pembelajaran Dengan Model Latihan Penelitian di Sekolah Dasar. *Pedadidaktika: Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, *3*(2), 217–227.

Tiara, S. K., & Sari, E. Y. (2019). Analisis Teknik Penilaian Sikap Sosial Siswa dalam Penerapan Kurikulum 2013 di SDN 1 Watulimo. *Edu Humaniora Jurnal Pendidikan Dasar, 11 (1), 21-30.* https://doi.org/10.17509/eh.v11i1.11905

Yudha, R. P. (2020). Validity And Reliability Rubric Of Performance Assessment Geometry Study In Junior High School Using The Many Facet Rasch Model Approach. *Eduma : Mathematics Education Learning and Teaching*. 9 (2), 26-36. https://doi.org/10.24235/eduma.v9i2.7100

Zanuddin, H., Shaid, N., & Halim, Z. (2021). Online learning platform during Covid-19 pandemic: An exploratory factor analysis on students' attitude and perceived usefulness in media and communication education. *Communications in Humanities and Social Sciences*, *1*(1), 7–15. https://doi.org/10.21924/chss.1.1.2021.7