The development of affective assessment instruments in the new normal era (post-covid-19)

Muhammad Arfan Mu'ammar[1]**, Aris Sofandi**[2]
arfan.slan@gmail.com[1], arissofandi@gmail.com[2]
University of Muhammadiyah Surabaya[1], SMPN 61 Surabaya[2]

Abstract

This development research was conducted at MTs Ma'arif Ketegan Sidoarjo. The types of data obtained are qualitative and quantitative data. The survey results showed that 50% of teachers developed their own affective assessment instruments but 87.5% of teachers admitted that these instruments were not valid and reliable. Then as many as 81.3% of teachers chose to use the observation method in assessing student effectiveness during online learning, because it is considered the easiest. As many as 62.5% of teachers do not understand that affective assessment of students in online learning can be done using the self-assessment method. The instrument is presented in a Google Form which contains 50 contextual statement points with a Likert Scale. The content validation score with the Aiken's V formula obtained a score of less than 1 for each statement item. Construct validation is a follow-up to the input of the supervisor and the theory that has been analyzed. External validation shows an increase in instrument models that are more applicable and adaptive. Instrument stability obtained a score of 0.885 so it was concluded that it was less reliable for a significant level of 1% and reliable for a significant level of 5%. The instrument equivalence data shows that there is no significant change in the results of the original and alternative instruments. Internal consistency with the Cronbach Alpha coefficient obtained a score of 0.92 with a very high category. The instrument also has a practical and efficient title. Thus this instrument is considered valid, reliable, useful, and valid. The response given by this instrument is said to be valid, reliable, practical and valid.

Keywords: *Affective, Assesment intructional*

INTRODUCTION

Education in Indonesia is currently undergoing a tremendous transformation. The online learning process shows the new face of Indonesian education. In the learning process, assessment is a very important aspect. Assessment in learning is an effort to collect information to determine student progress (Imania & Bariah, 2019) and measuring student learning outcomes (Mendikbud RI, 2016). Assessment is different from measurement, measurement is only done by assigning numbers to a condition according to certain rules, while assessment leads to consideration of decisions on the results of measurements (Ratumanan & Rosmiati, 2019). In learning assessment, teachers need a tool to collect information on student learning outcomes which is called an assessment instrument " "(Riscaputantri & Wening, 2018). In education, there are three domains that must be assessed comprehensively, namely the cognitive domain (knowledge), psychomotor domain (skills), and affective domain (attitude) (Mendikbud RI, 2016).

Currently, the world has undergone a transformation in the field of education, namely with the increasing role of technology in supporting the learning process. Research shows that technology can change social interactions, for example, students become more open in conveying their ideas (Yee & Bailenson, 2007). Thus, students are more responsive, active, and open-minded. Technology is considered capable of improving learning in the cognitive and affective domains (Näykki, Laru, Vuopala, Siklander, & Järvelä, 2019). The phenomenon of the Covid-19 pandemic has also encouraged the involvement of a more massive role of technology in the learning process. The existence of large-scale social restrictions rules causes learning to be carried out remotely, both synchronously and asynchronously. Teachers and students are required to better master technology so that the learning process does not stop.

The role of technology in learning is not entirely positive. Online-based learning is considered less accommodating the affective needs of students. So far, not many researchers have focused on the affective domain. This is natural because the world of education tends to emphasize the cognitive aspects of students. The decision whether a student is smart or not is more in terms of his cognitive level so that student learning achievement is only measured by how high his cognitive scores are (Wong, 2020). Teachers who only focus on students' cognitive are considered to provide an incomplete educational experience (Hall, 2011).

Assessment in the affective domain is rarely carried out because the teacher does not understand the concept of competence in the affective domain (Setiawan & Tumardi, 2019).

The research results '(Imtihan & Ahyar, 2018) mentions that (1) many teachers assume that students' affective domain achievements are determined by the discipline and moral attitudes shown by students, (2) some teachers assess students' affections personally without using a valid assessment instrument. Similar research results (Riscaputantri & Wening, 2018) which states that around 83% of teachers do not have an affective assessment instrument, 96% of teachers do not understand the importance of affective assessment, and 93.3% of teachers do not conduct affective assessment. Many teachers need training, especially in the mastery of the affective domain (Velea & Farca, 2013). Thus, teachers feel difficult to develop the affective assessment instruments. As a result, teachers use instruments that have not been tested for validity and reliability. Empirical data states that so far the Government has only provided theoretical training related to the steps of preparing the instrument and has not set a standard format for the student affective assessment instrument (Setiawan & Tumardi, 2019). For this reason, this study intends to present a student assessment instrument in the affective domain which is valid and practical so it can be used by teachers, especially in online learning situations in this new normal era.

In the 2013 curriculum, the affective side can be divided into two, namely spiritual attitudes and social attitudes. Spiritual attitude relates between students as a human being with their creator. While social attitudes are related between students and their environment (Ratumanan & Rosmiati, 2019), both the family environment, school, and community (Mendikbud RI, 2017). The government has established character values that must be mastered by students, namely, religious, honest, tolerance, discipline, hard work, creative, independent, democratic, curiosity, national spirit, love for the homeland, respect for achievement, communicative, love peace, likes to read, cares about the environment, cares about social, and responsibility (Mendikbud RI, 2010). In further developments, the Government established 5 (five) main character values in the Character Education Strengthening Program, namely religious, nationalism, integrity, independence, and mutual cooperation (Mendikbud RI, 2017). These five character values are the magnitudes of the values derived from the previous value derivatives. The following describes the classification of the Government's determination related to character values that must be mastered by students.

Table 1 - Classification of the Determination of the Minister of Education and Culture on Character Values

Determination of the Minister of Education and Culture in 2017	Determination of the Minister of Education and Culture in 2017
Religious	Religious, tolerance, love peace
Nationalism	Discipline, national spirit, love for the homeland, appreciate achievements, care about the environment
Integrity	Honest, responsibility
Independence	Hard work, creative, independent, curiosity, like to read
Mutual Cooperation	Democratic, communicative, socially caring

The affective domain is an abstract realm which is difficult to assess. Affective assessment must be carried out continuously in the form of instructional to encourage students in a more positive way (Hall, 2011). Researchers agree that the best way to assess the affective domain is by paying attention to the behavior that underlies the student's character (McCoach et al., 2013). Thus, the content of the instruments that will be compiled later are statements of concrete behavior that can be observed that approach the affective realm of students. The affective assessment must be carried out comprehensively because the results of the current assessment may differ from the results of the next week's assessment. This is because humans are dynamic creatures whose thoughts and motivations can change.

Before compiling an affective assessment instrument, it is necessary to pay attention to the domain of affective characteristics to be assessed. There are eight affective characteristics that can be observed, namely, (1) Attitude; (2) Value; (3) Self-esteem; (4) Self-control; (5) Self-efficacy; (6) interest; (7) aspirations; (8) anxiety. Each characteristic should to contain of three criteria, namely a clear target, relevant intensity, and the right direction (Anderson & Bourke, 2013). In relation to the Government's Character Education Strengthening Program above, the relevant characteristics are aspects of attitudes and values. Thus, the criteria that meet the character targets are to achieve the goals of character education in Indonesia, implemented with high intensity with special attention at all levels of education in Indonesia, and lead to the mental development of the nation's generation.

Affective assessment can be done by using the method of observation, self-assessment, peer assessment, and teacher notes '(Anderson & Bourke, 2013; McCoach et al., 2013; Mendikbud RI, 2016; Phan, 2019; Ratumanan & Rosmiati, 2019). Assessment of affection can be in the form of textual, visual, physiological, and multimodal, but is often used is in the

form of textual, which can be in the form of self-assessment, questionnaires, and observations. The combination of textual form and student visualization is considered to be an effective combination in knowing the level of students' affection (Yadegaridehkordi, Noor, Ayub, Affal, & Hussin, 2019). Mc. Coach argues that observation has weaknesses, including the observed students tend to show untruthful attitudes (McCoach et al., 2013), so the implementation of the observation method must involve participants who come from the internal communities (Kozinets, 2010). Phan added that an effective method is the interview method '(Phan, 2019), however, this method has several weakness, including (1) the interviewer may only focus on completing the list of questions without delving into the respondent's answers; (2) the respondent may only give the answer the interviewer wants to hear; and (3) respondents may lie because they are worried that it will affect their portfolio. For this reason, the researchers suggest choosing a self-assessment method to get accurate and valid affective assessment results (Anderson & Bourke, 2013). In addition, the resulting assessment instrument must refer to valid, objective, fair, integrated, open, comprehensive, systematic, criteria-based and accountable principles (Mendikbud RI, 2016). Borg added that there are five criteria that are often used to improve the quality of the instrument, namely objectivity, test and score standards, content interpretation standards, and fairness (Gall, Gall, & Borg, 2002).

During the online learning process, teachers are required to be able to implement online assessments also. Online learning can be synchronous or asynchronous, namely through blogs, Wikis, You tube, social networks, social media, online communities, and interactive virtual games (Korucu & Atun, 2016). Thus, the assessment which is conducted cannot be equated with a face-to-face assessment. In the online affective assessment process, the teacher must pay attention to the netnographic aspect. Kozinets uses the term netnographic to refer to the anthropological approach to virtual human behavior (Kozinets, 2010). Indeed, ethical research need to pays attention to 3 (three) basic principles, namely 1) respecting the autonomy of research subjects; 2) avoid danger; and 3) pay attention to privacy and data protection (National Advisory Board on Research Ethics, 2009).

Kozinets (Kozinets, 2010) explains how to obtain data from netnographic research, namely research planning, main data sources, data collection, interpretation, ensuring ethical standards, and research representation. The netnographic research method to assess students' effectiveness is done by self-assessment '(Anderson & Bourke, 2013; McCoach et al., 2013; Mendikbud RI, 2016; Phan, 2019; Ratumanan & Rosmiati, 2019), observation

(Anderson & Bourke, 2013; Hetland & Mørch, 2016; Kozinets, 2010; McCoach et al., 2013; Mendikbud RI, 2016; Phan, 2019; Ratumanan & Rosmiati, 2019), and interview (Hetland & Mørch, 2016; Kozinets, 2010). Kozinets suggests to conducted mixed netnographic research (ethno-netnography) because the data generated from online observations are not sufficient to represent data conclusions. Data retrieval requires involvement, contact, interaction, and harmonious relationships (Kozinets, 2010). Boellstorff added that there are 5 (five) things that participant observers need to pay attention to, namely 1) researchers at least do not burden informants with questions about general information; 2) the researcher must initiate an emotional relationship with the informant; 3) the researcher's observations may contain some errors; 4) researchers must make extensive field notes; and 5) researchers must maintain the consistency of the taken the data (Boellstorff, Nardi, Pearce, & Taylor, 2012).

Most researchers focus on the design of emotion recognition and expression instruments for the world of educational computing (digital) (Yadegaridehkordi et al., 2019). Researchers in the world have formulated an affective self-assessment instrument called the Assessment Experience Questionnaire (AEQ). AEQ is a self-assessment instrument developed to measure students' effectiveness which focuses on 6 items, namely: 1) the demands of students' time and effort; 2) assignments and learning; 3) the quantity and time of the feedback; 4) the quality of the feedback; 5) the use of feedback; 6) exam and study (Gibbs & Simpson, 2003). Self-assessment is defined as a paper and pencil instrument whose items produce a numerical score from which conclusions can be drawn about how individuals distinguish various aspects of themselves, such as personality traits, self-concept, learning styles, attitudes, values, and interests (Gall et al., 2002).

In developing a self-assessment instrument, it is necessary to pay attention to several problems that arise, namely word choice, consistency of item dimensions, habits in assessing, habits of persuasive instrument statements, number of scales (the most effective scale interval is 4 to 7 points), bias in the response scale (McCoach et al., 2013). To produce a quality assessment instrument, it is necessary to pay attention to the following boundary formulations: 1) the item contains only one dimension; 2) item and response scale using clear language, not dialect or slang; 3) items must be unidirectional, unambiguous, and not accompanied by comparison items; 4) the response scale must be clear and not imply hidden answers; 5) The response scale contains the same interval and is able to accommodate all possible answers (McCoach et al., 2013).

The scale content contained in the instrument must be communicative, objective, reliable, and easy to interpret (Anderson & Bourke, 2013). Communicative means whether t

he instrument can be understood by the respondent, the steps for working on the instrument must be clear and systematic. Objective means the extent to which the rater's subjectivity affects the instrument's points. Reliable means whether the instrument used produces consistent data from the subject being assessed if the assessment is repeated at the same time. And easy to interpret means whether the instrument produces data that can be interpreted and then concluded as a result of the assessment. There are many types of scales that can be used to measure assessments such as the Likert Scale, Thurstone Scale, Guttman Scale, Semantic Differentiation, and Sign Scheme, but researchers agree that the Likert Scale is the best scale for affective assessment (Anderson & Bourke, 2013; Dunn, Morgan, OReilly, & Parry, 2004; McCoach et al., 2013).

Anderson (Anderson & Bourke, 2013) describes the stages of the strategy of designing the instrument using a Likert scale, namely: 1) select a statement based on the affective characteristics studied; 2) feasibility test of the selected statement to the reviewer; 3) provide backup statements to replace invalid statements; 4) determine the required response scale interval; 5) present the instrument equipped with working steps; 6) instrument feasibility test on the sample; 7) analysis of test results to conclude that the instrument is suitable for use.

As explained above, in netnographic research, self-assessment instruments need to be complemented by observations. The most relevant type of observation is participant observation. This method is conducted by observing students' use of words/sentences in online classes and observing directly by participant observers. Observer participants in question such as family members, neighbors, and close friends of students. Szurawitzki formulated a technique to analyze the results of observations in students' online classes, namely 1) determining the size, length, and amount of text used by students; 2) determine the type of language used; 3) grapheme analysis, namely the use of capital letters; 4) analysis of semiotic elements such as the use of emoticons, stickers, symbols, slang, abbreviations, dialects; 5) analysis of syntax elements and punctuation (Szurawitzki, 2012).

Furthermore, to ensure the quality of the instruments compiled, it is necessary to review the degree of validity and reliability. The instrument is declared valid if it can be used to measure what is to be measured, while the instrument is declared reliable if the instrument is used several times on the same object and will produce the same data (Sugiyono, 2017). In the preparing the instrument, the validity test is divided into two, namely internal validity and external validity. Internal validity is only limited to construct validity, while construct and content validity are used for qualitative research data collection instruments

(Bakker, 2018; Sugiyono, 2017). Construct validity can only be done by experts. Reviews by experts can be done face-to-face non-face to face (Tessmer, 1993). Le Comte in Cohen describes the main things in internal validity, namely trust in data, data authenticity, data strength, research design accuracy, data credibility, data auditability, data dependence, and data confirmability(Cohen, Manion, & Morrison, 2000). While the instrument reliability test is divided into three, namely stability, equivalence, and internal consistency (Cohen et al., 2000; Gall et al., 2002). The stability was measured by testing the instrument on the same subject repeatedly at different times. Equivalence is measured by testing an instrument that has been paraphrased without changing its meaning. Internal consistency is measured using the helper table and the moment product.

METHOD

This study uses the Thiagarajan development model (Sivasailam, Semmel, & Semmel, 1974), namely *Four*-D (Define, Design, Develop, Dissemination) with a description of the activities outlined in the following flowchart. The instrument compiled is a self-assessment instrument using a Likert scale response. Meanwhile, the observation and interview formats were not designed in this study. Thus, it is expected that there will be further research that focuses on the preparation of observation sheets and interviews for affective assessment.
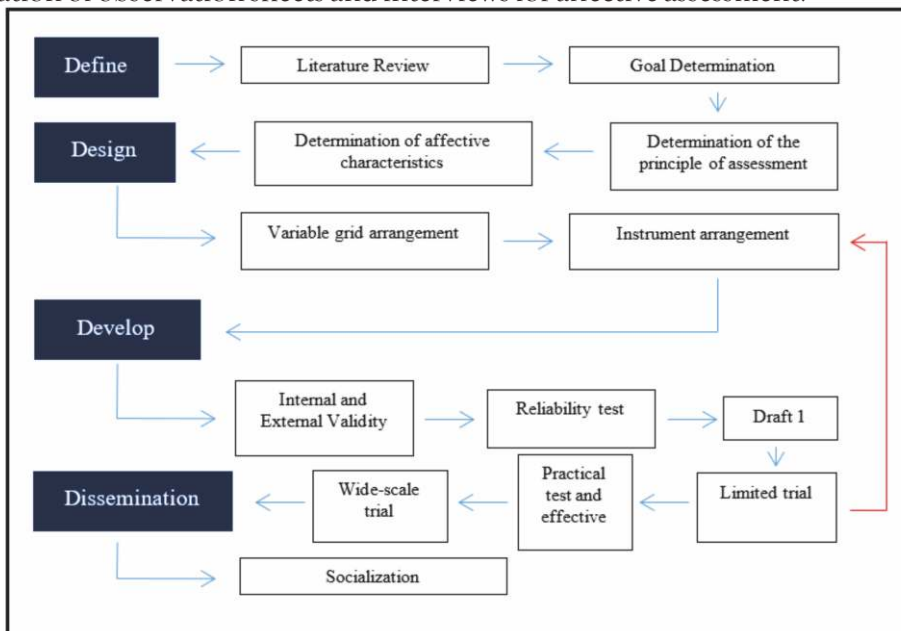

Fig. 1 - Thiagarajan 4-D Model Development Research Flow

The instrument needs to be tested for the degree of validity and reliability. The validity of this instrument was tested through internal validity and external validity. Internal validity was tested for construct validity and content validity. The constructs were validated by analyzing relevant theories and consulting experts. The content is validated by calculating the coefficient of the assessment results from the experts for each question item, and then entered in the following Aiken's V formula with a score conversion range of 0-1 for each item that is declared valid.

$$V = \frac{\sum s}{[n(c-1)]} \qquad (1)$$

Meanwhile, the external validity was tested by comparing the existing standards/criteria in the field. The reliability of the instrument was tested for stability, equivalence, and internal consistency. The subject of this product development trial was carried out on 15 teachers from various subject backgrounds. The way to measure the stability of the instrument is by test-retesting the same group at different times. Then measured the correlation coefficient of the first experiment with the next experiment, if it is positive and significant then the instrument is declared reliable. The measurement of the correlation coefficient uses the formula and the auxiliary table described by Sugiyono, then the results of the r count are compared with the r table of product moment (Sugiyono, 2017). The t-count formula as follow:

$$r_i = \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{[n \sum X_i^2 - (\sum X_i)^2][n \sum Y_i^2 - (\sum Y_i)^2]}} \qquad (2)$$

The way to measure the equivalence of the instrument is by making a comparison instrument with the same meaning but in a different language. The way to measure the internal consistency using the Cronbach Alpha formula.

$$r_i = \frac{k}{(k-1)}\left\{1 - \frac{\sum s_i^2}{s_t^2}\right\} \qquad (3)$$

To measure the level of practicality and effectiveness of the instrument, the researcher asked to the respondents to observe and fill out a questionnaire. Then the researcher conducted random interviews with the respondents to determine the verbal response to this instrument.

RESULTS AND DISCUSSION

At the product development stage, the researcher applied the 4-D Thiagarajan model. The researcher took the define stage where the researcher conducted a literature study and pre-research field survey. The results of this survey obtained data that many respondents answered that there were no affective assessment instruments that could be used in learning the new normal era. The survey results are strengthened by the results of observations of several assessment instrument documents used so far. Researchers also conducted random interviews to get more accurate data. Then, the researcher determines the development objectives and pays attention to the principles of assessment in accordance with the literature review. The researcher also formulates what affective characteristics are the scope of the instrument, namely attitudes and values. This character is in accordance with the strengthening of government character education that must be mastered by students.

At the design stage, the researcher compiled a variables grid, indicators for each variable, and the number of statements representing each variable. For each variable, 2-3 statements are provided which aim to obtain the consistency of the respondents' answers. At the develop stage, the instrument is arranged in a Google Form to make it easier for teachers to assess students online. Responses to questions were selected on a Likert scale with 5 intervals in order to be able to provide more accurate and specific answers. The weight of the response scores between 1-5 consecutively represented the options of not answering, never, sometimes, often, always.

The construct validity test was conducted by analyzing theories from various sources such as books, national and international journals, scientific articles, opinions of figures and experts. Experts provide recommendations for improvement by filling out the validation sheet shown in the image below.

Fig. 2 - Validation sheet and consultation with experts

The content validity test obtained the following data after being processed into the Aiken's V formula. The range of values calculated as valid content was 0-1.

Table2 - Content Validity Test Results

N	V	N	V	N	V	N	V	N	V	N	V	N	V	N	V	N	V	N	V
1	0.63	6	0.75	11	0.50	16	0.75	21	0.75	26	0.38	31	0.75	36	0.50	41	0.63	46	0.50
2	0.63	7	0.63	12	0.75	17	0.75	22	0.63	27	0.63	32	0.63	37	0.50	42	0.75	47	0.63
3	0.63	8	0.63	13	0.63	18	0.75	23	0.63	28	0.50	33	0.75	38	0.63	43	0.75	48	0.63
4	0.63	9	0.38	14	0.75	19	0.63	24	0.63	29	0.75	34	0.63	39	0.75	44	0.63	49	0.63
5	0.63	10	0.50	15	0.50	20	0.75	25	0.63	30	0.50	35	0.63	40	0.63	45	0.50	50	0.75

The external validity test was obtained by comparing the existing instruments in the Student Assessment Guide book published by the Ministry of Education and Culture. A snippet of the instrument model is presented in the image below.
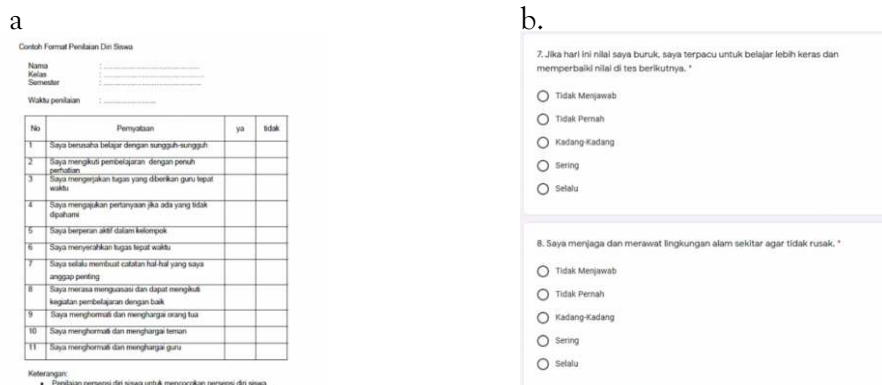
a

b.

Fig. 2 – Figure Description (a) Snippets of instrument statement items issued by the Ministry of Education and Culture; (b) Draft Instrument developed

The reliability test is done by testing its stability, equivalence, and internal consistency. The instrument stability test obtained the following data, then entered into the moment product formula.

Table 3 - Helper table of experimental results of 5 respondents at different times

N	X_1	X_2	$X_1{}^2$	$X_2{}^2$	X_1X_2
1	195	192	38025	36864	37440
2	199	205	39601	42025	40795
3	210	215	44100	46225	45150
4	195	202	38025	40804	39390
5	202	205	40804	42025	41410
Total	1001	1019	200555	207943	204185

$$r_i = \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{[n \sum X_i^2 - (\sum X_i)^2][n \sum Y_i^2 - (\sum Y_i)^2]}}$$

$$= \frac{5\,(204185) - (1001)(1019)}{\sqrt{[5(200555) - (1001)^2][5(207943) - (1019)^2]}}$$

$$= \frac{1020925 - 1020019}{\sqrt{[1002775 - 1002001][1039715 - 1038361]}}$$

$$= \frac{906}{1023,7}$$

$$= 0.885$$

Based on the above calculations, it can be obtained that the value of r count is 0.885. Furthermore, the calculated r value must be compared with the table r price. If the calculated r value is greater than the r table price, the instrument can be declared reliable. Based on the r table, it can be obtained prices with a significant level of 5% of 0.878 and with a significant level of 1% of 0.959. Thus the position of the calculated r price is 0.959 > 0.885 > 0.878, it can be concluded that the instrument with a significant level of 1% is declared less reliable, but with a significant level of 5% the instrument is declared reliable.

Furthermore, the level of equivalence is measured by making a comparison instrument with the same meaning but in a different language. The results obtained are that there is no significant change in the assessment results from the two instrument formats. Meanwhile, to measure the internal consistency using the Cronbach Alpha formula.

Table 4 - Alpha Cronbach formula helper table

Item No.	Respondent Number					Total per item	Squared per item
	1	2	3	4	5		
X_t	195	199	210	195	202	2276	115038
X_t^2	38025	39601	44100	38025	40804	$\sum X_t^2$: 200555	

$$s_t^2 = \frac{\sum X_t^2}{n} - \frac{(\sum X_t)^2}{n^2}$$
$$= \frac{2276}{5} - \frac{(1001)^2}{25}$$
$$= 455,2 - 40080,04$$
$$= -39624,84$$

$$s_i^2 = \frac{JKi}{n} - \frac{JKs}{n^2}$$
$$= \frac{2276}{5} - \frac{115038}{25}$$
$$= -4146,32$$

$$r_i = \frac{k}{(k-1)}\left\{1 - \frac{\sum s_i^2}{s_t^2}\right\}$$
$$= \frac{50}{(50-1)}\left\{1 - \frac{-4146,32}{-39624,84}\right\}$$
$$= 1,02\ \{0,9\}$$

$$= 0,92$$

The researcher then converts the results of these calculations so that the appropriate conversion criteria are obtained at the value above, namely 0.90 ri < 1.00 with very high information. Thus it can be concluded that the internal coefficient score is 0.92 with a very high category.

Furthermore, the researchers conducted a practicality and effectiveness test by distributing questionnaires to experimental subjects with comparison subjects. The results obtained are the score of the Experimental Subject group of 154. While the score of the comparison Subject group is 149. The results of the conversion of these scores conclude that the scores obtained by the two groups of subjects are Practical and Effective.

The researcher continued the test by conducting interviews with randomly selected experimental subject respondents. The results of the interview stated that this instrument was quite practical, although there were a lot of question points, but it could describe the actual condition of the students (Ridlo, 2021).

CONCLUSION

The phenomenon of the Covid-19 pandemic has caused the affective assessment to be modified in such a way which is it can be implemented in online learning. A valid, reliable, practical and effective affective assessment instrument is needed by the teacher. The results of this development research produce affective assessment instruments that can be used by teachers in the New Normal era. This instrument is declared valid after being tested for internal and external validity. Then tested the level of reliability through instrument stability, equivalence, and internal coefficients. Then also tested the level of practicality and effectiveness by filling out a questionnaire and interviewing the subject. The results of this study are expected to encourage other researchers to develop affective assessment instruments with the observation method.

REFERENCES

Anderson, L. W., & Bourke, S. F. (2013). Assessing affective characteristics in the schools, second edition. Assessing Affective Characteristics in the Schools, Second Edition. doi:10.4324/9781410605443

Bakker, A. (2018). Design Research in Education: A Practical Guide for Early Career Researchers. New York: Routledge.

Boellstorff, T., Nardi, B., Pearce, C., & Taylor, T. L. (2012). Ethnography and Virtual Worlds: A Handbook of Method. New Jersey: Princenton University Press.

Cohen, L., Manion, L., & Morrison, K. (2000). Research Methods in Education (5th ed.). London: Routledge Falmer.

Dunn, L., Morgan, C., OReilly, M., & Parry, S. (2004). The Student Assessment Handbook. The Student Assessment Handbook. London: Routledge Falmer.doi:10.4324/ 9780203416518

Gall, M. D., Gall, J. P., & Borg, W. R. (2002). Educational Research: An Introduction. (A. E. Burvikovs, Ed.) (7th ed.). Boston: Library of Congress Cataloging-in-Publication Data.

Gibbs, G., & Simpson, C. (2003). Measuring The Response of Students to Assessment: The Assessment Experience Questionnaire. Milton Keynes. Retrieved from docplayer. net/12543680-Measuring-the-response-of-students-to-assessment-the-assessment-experience-questionnaire.html

Hall, R. A. (2011). Affective Assessment: The Missing Piece of The Educational Reform Puzzle. International Journal for Professional Education, 772(1), 710. Retrieved from http://www.deltakappagamma.org/NH/dkgbulletinwinter2011.PDF

Hetland, P., & Mørch, A. I. (2016). Ethnography for Investigating the Internet. Seminar.Net, 12(1), 114.

Imania, K. A. N., & Bariah, S. K. (2019). Rancangan Pengembangan Instrumen Penilaian Pembelajaran Berbasis Daring. PETIK, 5(1), 3146. Retrieved from https://journal.institutpendidikan.ac.id/index.php/petik/article/viewFile/445/4 52

Imtihan, N., & Ahyar. (2018). Analysis of Problem and Need on The Affective Assessment Implementation for Learners in Madrasah Aliyah. In Internasinonal Seminar on Education and Development of Asia (pp. 102112). Mataram. Retrieved from https://core.ac.uk/reader/234035660

Korucu, A. T., & Atun, H. (2016). Use of Social Media in Online Learning. In Handbook of Research on Innovative Pedagogies and Technologies for Online Learning in Higher Education (pp. 1547). USA: IGI Global. doi:https://doi.org/10.4018/978-1-5225-1851-8

Kozinets, R. V. (2010). Netnography: Doing Ethnographic Research Online. International Journal of Advertising (Vol. 29). London: SAGE. doi:10.2501/S026504871020118X

McCoach, D. B., Gable, R. K., Madura, J. P., McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). Defining, Measuring, and Scaling Affective Constructs. Instrument Development in the Affective Domain. London: Springer. doi:10.1007/978-1-4614-7135-6_2

Mendikbud RI. (2010). Pengembangan Pendidikan Budaya dan Karakter Bangsa. Jakarta: Badan Penelitian dan Pengembangan Pusat Kurikulum. Retrieved from http://new-indonesia.org/beranda/images/upload/dok/kurikulum/pengembangan-pendidikan-budaya-dan-karakter-bangsa.pdf

Mendikbud RI. (2016). Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 23 Tahun 2016. Jakarta. Retrieved from https://bsnp-indonesia.org/wp-content/uploads/2009/09/Permendikbud_Tahun2016_Nomor023.pdf

Mendikbud RI. (2017). Penguatan Pendidikan Karakter Jadi Pintu Masuk Pembenahan Pendidikan Nasional. Retrieved 21 November 2020, from https://www.kemdikbud.go.id/main/blog/2017/07/penguatan-pendidikan-karakter-jadi-pintu-masuk-pembenahan-pendidikan-nasional

National Advisory Board on Research Ethics. (2009). Ethical Principles of Research in The Humanities and Social and Behavioural Sciences and Proposals for Ethical Review (Vol. 2009). Helsinki. Retrieved from https://www.tenk.fi/sites/tenk.fi/files/ethicalprinciples.pdf

Näykki, P., Laru, J., Vuopala, E., Siklander, P., & Järvelä, S. (2019). Affective Learning in Digital EducationCase Studies of Social Networking Systems, Games for Learning, and Digital Fabrication. Frontiers in Education, 4(November), 114. doi:10.3389/feduc.2019.00128

Phan, W. T. (2019). Affective Learning Objectives in Online Course. In ICEMT 2019: Proceedings of the 2019 3rd International Conference on Education and Multimedia Technology (pp. 3336). Nagoya - Japan: Association for Computing

Ratumanan, T. G., & Rosmiati, I. (2019). Perencanaan Pembelajaran (1st ed.). Depok: Rajawali Pers.

Ridlo, A. (2021). Wawancara. Sidoarjo.

Riscaputantri, A., & Wening, S. (2018). Pengembangan instrumen penilaian afektif siswa kelas IV sekolah dasar di Kabupaten Klaten. Jurnal Penelitian Dan Evaluasi Pendidikan, 22(2), 231242. doi:10.21831/pep.v22i2.16885

Setiawan, H., & Tumardi, T. (2019). Pengembangan Instrumen Asesmen Kompetensi pada Ranah Afektif di Sekolah Dasar. Musamus Journal of Primary Education, 2(1), 112. doi:10.35724/musjpe.v2i1.1944

Sivasailam, T., Semmel, D. S., & Semmel, M. I. (1974). Instructional development for training teachers of exceptional children: A sourcebook. ERIC. Indiana: Indiana University Bloomington.

Sugiyono. (2017). Statistika Untuk Penelitian. Bandung: Alfabeta.

Szurawitzki, M. (2012). Analyzing the Language of Social Networking Sites - An Analysis Model. Proceedings of the 24th Scandinavian Conference of Linguistics, (December 2012), 355363.

Tessmer, M. (1993). Planning and Conducting Formative Evaluations. London: Kogan Page.

Velea, S., & Farca, S. (2013). Teachers Responsibility in Moral and Affective Education of Children. Procedia - Social and Behavioral Sciences, 76, 863867. doi:10.1016/j.sbspro.2013.04.221

Wong, S. L. (2020). Affective characteristics for 21st century learning environments: Do they matter? International Journal of Interactive Mobile Technologies, 14(12), 186194. doi:10.3991/IJIM.V14I12.15567

Yadegaridehkordi, E., Noor, N. F. B. M., Ayub, M. N. Bin, Affal, H. B., & Hussin, N. B. (2019). Affective computing in education: A systematic review and future research. Computers and Education, 142(March), 119. doi:10.1016/j.compedu.2019.103649

Yee, N., & Bailenson, J. (2007). The Proteus Effect: The Effect of Transformed Self-Representation on Behavior. Human Communication Research, 33(3), 7190. doi:https://doi.org/10.1111/j.1468-2958.2007.00299.x