



## **The Development of Instrument for Assessing Students' Affective Domain Using Self- and Peer-Assessment Models**

**Ari Setiawan**

University of Sarjanawiyata Tamansiswa, Indonesia, [arierobbani@yahoo.com](mailto:arierobbani@yahoo.com)

**Djemari Mardapi**

Yogyakarta State University, Indonesia, [djemarimardapi@yahoo.com](mailto:djemarimardapi@yahoo.com)

**Supriyoko**

Universitas of Sarjanawiyata Tamansiswa, Indonesia, [supriyoko@ustjogja.ac.id](mailto:supriyoko@ustjogja.ac.id)

**Dedek Andrian**

Riau Islamic University, Indonesia, [dedekandrian@edu.uir.ac.id](mailto:dedekandrian@edu.uir.ac.id)

This research aims to develop instrument for assessing student's affective domain using self-assessment (SA) and peer-assessment (PA) models employs the modified model of development by McCoach. The sampling technique used was cluster random sampling. Aiken, CFA second order and alpha cronbach were used as the data analysis techniques, aided by Excel program, Lisrel and SPSS. The results show that there were 6 constructs of instrument for assessing student's affective domain. The content validity was done using Aiken index and the result ranged from 0,750-1,00 and all of the items were valid. The construct validity of the instrument was done using CFA approach. A fit model of instrument for affective domain based on self-assessment was developed with the value of Chi-square ( $X^2$ ) = 267,22 df = 246, P-value = 0,168, RMSEA = 0,022 and the instrument based on peer-assessment has the value of Chi-Square ( $X^2$ ) = 151,55, df = 126, P-value = 0,060, RMSEA = 0,034. On the items estimation, the value of loading factor was 0,31-0,99 (>0,30) which means that the items in the instruments based on SA and PA are valid. The instruments reliability reached 0,788-0,886 which means that all the instruments developed were valid. The valid instrument is very important to get good information about self-assessment and peer-assessment.

Keywords: affective, self-assessment, peer-assessment, student, teaching

### **INTRODUCTION**

Many studies show that student learning is positively influenced by the assessment (Wen, Tsai, & Chang, 2006). Student learning is influenced by assessment (Foley, 2013).

**Citation:** Setiawan, A., Mardapi, D., Supriyoko, & Andrian, D. (2019). The Development of Instrument for Assessing Students' Affective Domain Using Self- and Peer-Assessment Models. *International Journal of Instruction*, 12(3), 425-438. <https://doi.org/10.29333/iji.2019.12326a>

Assessment informs students about their abilities, strengths and weaknesses of students and shows the strategies used for the learning process (Gullickson, 2007). Active involvement of students in the learning process is an important requirement in the assessment (van Gennip, Segers, & Tillema, 2010). Assessment is used as a means to obtain information about student learning progress, the learning process, and improve student learning outcomes (Pandora & Mardapi, 2017). Assessment is used to investigate what is already known and can be done by students and to make decisions about achieving the expected goals (Baird, Andrich, Hopfenbeck, & Stobart, 2017).

Student contributions in the learning process are assessed through self-assessment and peer assessment in addition to teacher assessment, which will then be combined as students' final grades (McCoach, Gable, & Madura, 1986). Thus, student-based assessment is very much needed in supporting the relevance of assessment. In peer assessments, students assess the abilities of other students, while students' self-assessments assess their own abilities. Peer assessment is usually used in evaluating projects and practical presentations (McCoach et al., 1986).

Primary education serves as the main foundation and social environment in building children's affective domain (social behaviors). In other words, the basic foundation to build such behavior is done in this stage. It is due to the quite long years of studying in primary school. This is also because social behavior is a part of learning result (McCoach et al., 1986). The instilling of the behavior can be done through the curriculum design, learning process and the appropriate assessment.

Affective domain is a tendency to evaluate social things in certain ways. This is an important part in children development stage, as it builds their perception towards the social environment and significantly influences their behavior (Dwyer, 1998). A child who starts to interact with social environment will immediately have social behavior, and it happens to primary students.

The evidences of children's behaviors these days are quite concerning (Kanioglou, Tsorbatzoudis, & Barkoukis, 2005). The primary students are now generally less disciplined, have low care and responsibility. It does not accord with the ideal affective development of primary students. Contents that there are some values related to social condition that should be instilled in primary students, they are: *polite, caring, cooperative, disciplined, humble, even-tempered, tolerant, independent, honest, confident, tough, positive, fair, peaceful, hard-working, creative, responsible and sincere* (Kritikos, Woulfe, Sukkar, & Saini, 2011).

Another idea of how to instill the above values comes from (Logan & Ed, 2015) who says that the values to be built in primary students cover: *cooperation, positive competition, generosity, social acceptance, sympathy, empathy, dependence, hospitality, imitation and closeness*. Those values are crucial in primary stage, thus, serious consideration and attention are definitely needed.

Affective domain is a component to be assessed in 2013 Curriculum, because it is a graduate competence standard (SKL). The existing assessment system is simple without sufficient indicators. The teachers have put more focus on the assessment of cognitive

aspect which has clearer construct and criteria, while the affective aspect has more complicated construct and the teachers have insufficient competence in designing the instruments of the assessment. Another obstacle is the fact that designing learning objectives in terms of affective aspect is more difficult than designing the cognitive and psychomotor aspects (Pandra & Mardapi, 2017). In other words, affective domain is difficult to define and assess because it is abstract.

Based on the information collected, the researcher found that the instrument of affective domain is limited. The existing instruments do not cover all the aspects suggested in the core competence of the curriculum. Besides, the assessment uses only one type of method or model.

Many teachers claim they find it difficult to develop such instrument of assessment because of the insufficient examples in the technical guideline (Hadi & Andrian, 2018). They also feel difficult in defining the indicators. Those obstacles must be solved remembering that teachers are responsible for assessing the students. Based on a research by Zhang & Burry-Stock (2003), a teacher is a person responsible for assessing the students, thus, teachers should improve their assessment skill. If they do not assess the students appropriately, the result of the assessment will definitely be less accurate.

From the existing instruments it is also found that the instruments have not been designed according to the guidelines. They are orientated to the teachers' observation, while such orientation only focuses on administrative demand and routines (Evans, Elwyn, & Edwards, 2004). An appropriate instrument should involve self- and peer-assessment. Self-assessment leads to the consistent result covering all components, tasks, and a relatively short time (Ross, 2006). Students will be able to assess themselves only if they have a clear understanding about what they are learning (Stiggins, Richard, 1999).

A complete form of such instrument will lead to comprehensive and meaningful assessment result, which then will be used as the graduate competence standard as one of the goal of 2013 Curriculum. The complexity of students' social behavior and the limited instruments for assessing them become the underlying reason for conducting review and development of more sufficient instruments. Since the existing instruments are focused only on one general substance and model, more sufficient instrument to get comprehensive assessment result is definitely needed.

Based on a survey in June, the recent instrument for assessing students' affective domain used by the teachers of primary schools in Yogyakarta Province are based on the assessment technical guideline for primary school year 2013 from Kemendikbud. The guidelines contain affective aspects to be assessed: *honest, disciplined, responsible, polite, caring, confident*, and other values relevant to the learning competence (Kemendikbud, 2013). The guidelines also provide examples of assessment though they are limited only to one affective component that is *honest* without elaborate indicators and assessment rubric. Besides, there are four types of model suggested which cover observation, self-assessment, peer-assessment and journal (Kemendikbud, 2013). However, the example is too simple and does not contain the complete affective aspects.

As a result of the available guideline which is insufficient, the primary teachers in Yogyakarta Province still have difficulties in revealing the students' affective domain. Although many of the teachers have already joined workshops on assessment based on 2013 Curriculum, they are still incompetent to develop such a good instrument. It very clear that assessment process which is only focused on the teacher and has incomprehensive instrument and non-authentic assessment technique cannot give accurate and elaborate information about affective domain such as *honest, disciplined, responsible, caring, polite* and *confident*. Based on the problem identification, the development of instrument for assessing students' affective domain using peer- and self-assessment is badly needed.

The research problem is limited to the development of instrument for assessing affective aspects covering *honest, discipline, responsible, polite, caring* and *confident* using self-assessment (SA) and peer-assessment (PA) models. The aim of the problem limitation to make the research focused and applicable based on the proposed limitation.

#### **METHOD**

This research aims to develop instrument for assessing primary students' affective domain using SA and PA models. The instrument developed was in non-test form. The development of the instrument covered some components; they are *honest, disciplined, responsible, polite, caring* and *confident*. In developing the instrument the researcher employed McCoach et al., (1986) modified procedures which cover 13 steps.

The research procedures can be seen from the figure 1 below.

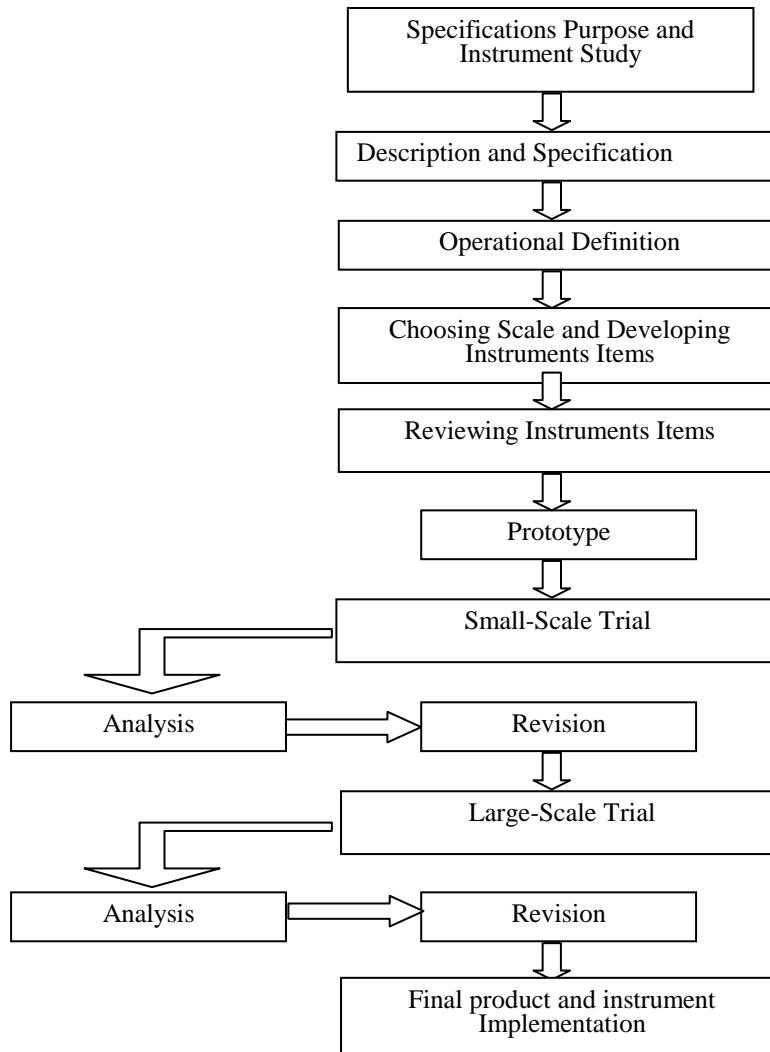


Figure 1  
Procedures for the Development of SA and PA Instruments

The procedures above follow the model from McCoach et al., (1986). Formulating the objectives of the instrument is an important early step. Developing instrument is a long and valuable process, thus, there is a need for making sure there has been no similar type of instrument. Therefore, the researcher conducted surveys at SD MuhammadiyahSapen and SD Karangmojo. In these schools, the researcher did analysis of the documents used

for assessing students' affective domain. The next step was defining the affective domain. The affective domain in this research is behavioral pattern related to the social situation which is conditioned at the school. This means the behavior is as the result of learning process.

Based on the behavior construction, the specification of the affective domain emphasizes on the students' activities at school. Besides, the developed component was based on the core competence 2 (KI 2) as stated in 2013 Curriculum. The components covered are *honest, disciplined, responsible, polite, caring* and *confident*.

The next step was comprehensive literature review. This aimed to define the characteristics of affective domain. The definition was derived from the relation between the affective aspects and the core competence which can lead to the *honest, disciplined, responsible, polite, caring, and confident* values. Operational definition then could be made which covered variable and components to be measured. This operational definition was used as the guideline to develop comprehensive and applicable instrument. An appropriate operational definition could help the researcher arrange instrument indicators which are relevant to the developed construct.

Table 1  
Operational Definition of Affective Domain Components

No.	Components	Operational Definition
1	Honest	Trying to be trusted person in speaking, acting and working at school
2	Disciplined	Obeying all the rules established at school
3	Responsible	Doing all the duties which should be done at school
4	Polite	Showing respect reflected in speaking and behaving to the teachers and peers
5	Caring	Giving appreciation and empathy to the teachers, peers and school environment
6	Confident	Such behavior which is based on mental or psychological condition which build self-certainty and bravery

The next step was defining the scale which involves social behavior scale. The scale developed employed summative rating scale and modified the *Likert* scale. Summative rating scale was used to develop the instrument based on self- and peer-assessment models. The options were *Often* (Of), *Sometimes* (So), *Rare* (Ra), *Never* (Ne) with the score range 4-3-2-1 for favorable statement, and the reverse for the unfavorable statement. The options used in the instruments were: *Accustom* (A), *Began to grow* (Bg), *began to look* (Bl) and, *have not seen* (Hs). The score range was 4-3-2-1 favorable statement and the reverse for the unfavorable statement.

After defining the scale, the step was constructing indicators used for the guideline for developing instrument items. The indicators chosen can be seen in table 2.

Table 2  
Affective Domain Components and the Indicators

No.	Components	Operational Definition
1	Honest	1. Doing test individually and independently 2. Sharing information as it is
2	Disciplined	1. Arriving school in time 2. Obeying rules established at school
3	Responsible	1. Doing individual task as told 2. Returning things borrowed
4	Polite	1. Greeting the students and peers 2. Asking for permission when entering or leaving class
5	Caring	1. Helping friends who need help 2. Keeping the school environment safe and clean
6	Confident	1. Being serious during the learning process 2. Sharing opinion during the learning process

After the indicators were defined, the researcher then design the instrument based on those indicators and applied them to the measured component and construct. The result of this stage is first draft of the instrument. This stage was then followed by formal validation. The validation process involved some experts. They reviewed the questions both qualitatively and quantitatively. The qualitative review was done using Delphi method (Harlod, 2002) involving primary education experts, evaluation and measurement experts, psychometrics expert, psychological counselor, teachers and some students. The validation involved items readability, the relevance of the items and the guideline, and items revision.

The quantitative review was done through content validity. The *Aiken* formula was used, which was more elaborated in data analysis stage. The readability of the items was also done by distributing them to the teachers and students of SD MuhammadiyahSapen.

After the validation stage was conducted, the researcher designed the final version of the instrument which was ready for a limited try-out. The revised instrument was tried-out to primary students. The try-out was done at SD PonjongGunungkidul, SD MuhammadiyahPapringanSlemanand SD Pakel Kota Yogyakarta. The limited try-outs involved 180 students. The aim of the try-outs was to reveal the quality of the instrument in terms of the validity and reliability.

#### ***Sample and Sampling Technique***

The sample used in this research was elementary student in Yogyakarta Province. The sampling technique used in this research was cluster random sampling. The sampling process done by taking some of district randomly, then taking some ofschools randomly and taking elementary student randomly.

#### ***Instrument***

The instrument used in this research is a questionnaire. The questionnaires that have been developed are then validated by experts. The questionnaires consists of Affective

Domain using self-assessment (SA) and peer-assessment (PA) models employs the modified model of development by Mccoach. There are six indicators or component in this development research namely; honest, disciplined, responsible, polite, caring, and confident. There are 24 items for peer-assessment instrument and 18 items for self-assessment. There two items for each component or indicators from self-assessment instrument and there are three items for each component or indicators from peer-assessment instrument.

#### **Data Analysis Technique**

The data containing students' affective domain were then analyzed. The analysis covered items validation using CFA and LISREL8.80(Jöreskog, Olsson, & Wallentin, 2016). This was done to get the amount of the valid items. To measure the reliability of the developed instrument, *cronbach alpha* was employed, while the reliability coefficient was measured using SPSS 20.0. The result showed that the instrument needed revision. The items redaction and the item itself were parts to be revised. Such revisions aimed to get valid items and scale which could yield stable and internally-consistent score for the long term. Besides, the scores should enable the researcher to construct or derive conclusion of the construction that can assess affective domain.

After revising the instrument, the researcher did larger try-outs. The try-outs were done at SD KarangmojoGunungkidul, SD SokowatenBantul, SD Berbah Sleman, SD Gedongkiwo Kota Jogjakarta, SD Sentolo III Kulonprogo, SD Kaliagung Kulonprogoand SD WojoBantul. There were 380 students involved. The data found were analyzed to estimate the validity and reliability. The analysis was done using CFA aided with LISREL 8.80 program (Joreskog, 1999), while the coefficient reliability was done using *alpha* from *Cronbach* aided by SPSS 20.0.

#### **FINDINGS AND DISCUSSION**

The development of instruments for assessing affective domain which employed Mccoach approach has resulted instrument construct for assessing students' affective domain, covering *honest, disciplined, responsible, polite, caring, and confident*. The result has served as the initial or early step for developing affective domain instrument. Questionnaires were used as the form of development, employing self- and peer-assessment models.

Questionnaire method is relevant to the assumption that direct questions can reveal students' affective domain, since the students are ones who know well about themselves. Another assumption is honesty; human will express what he or she feels(Dwyer, 1998). The direct expression meant is written down through single or double items. The scale for affective domain is a method revealing students' affective aspects in the form of self-report containing lists of questions to be answered. The answers or statements were the object of the affective domain; they are *honesty, being disciplined, responsibility, politeness, care, and confidence* dimensions.

The next method, observation, was already appropriate as contended by (Nunnally & Bernstein, 1994) that behavior assessment can be done through behavior observation.



Someone's behavior can be observed consistently, for example when he often comes to the class late, it can be inferred that he has low self-discipline. It then quite makes sense when behavior is interpreted based on the observable actions or routines. In other words, such actions or routines can be the indicators of individual's affective domain.

The developed model based on SA and PA in this research is considered to be appropriate for assessing affective domain of primary students. The application of the models could reveal the students' social behavior comprehensively. SA model is in line with the idea of (Sluijsmans & Moerkerke, 2001) who emphasizes that self-assessment is generally used for formative assessment with the aim that students can reflect their process and result of learning. It is the reflection towards process and result of learning (Noonan & Duncan, 2005). Therefore, it is very clear that employing SA model is needed to assess affective domain, as a result of thematic learning process.

The second model or PA is based on students' activities related to their friends. Peers are involved to increase the assessment quality and serve as learning media. The students' involvement during the assessment process could: (1) improve the quality of assessment decision, (2) increase the students' commitment to employ the best assessment, (3) reduce the students' resistance towards feedback and need of change, (4) enhance the students' achievement, (5) increase the students' motivation to learn and build positive behavior, and (6) improve the students' self-assessment. The students' involvement in the learning and assessment process plays an important role to train the students to be responsible for their own learning.

The construct and assessment models which were defined gave guidelines and early stage in developing the instrument for assessing affective domain. Based on the result there were 12 indicators and 24 items constructed accommodating SA and PA models. Summative rating scale with modified *Likert* scale was also used. The scale had four options Often, Sometimes, Rare, Never, while the scoring was Often = 4, Sometimes = 3, Rare = 2, and Never = 1 for the favorable items and the reverse for the unfavorable items.

The developed instrument was then estimated in terms of its validity and reliability. Content validity was gained from the experts' judgment through *Deplhi* method, followed by *Aiken* index measurement. The result showed all of the indicators and items had *Aiken* index which ranged from 0,750-1,000. The result showed that the value is  $>0,7$  which means the indicators and items proposed were all valid (Heri Retnawati, 2016). *Aiken* index was chosen because of its accuracy in revealing the content validity of an instrument. The construct validity was done using CFA *second order*, and the results are as follows:

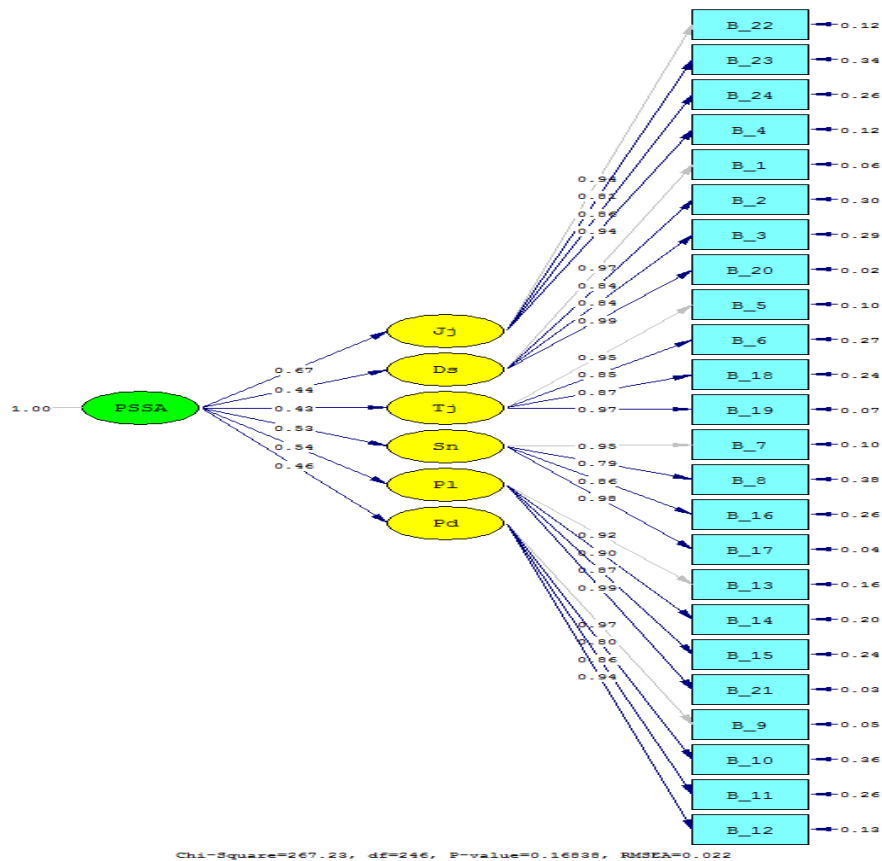


Figure 2  
The Result of SA-based Instrument Analysis Using CFA Second Order

Figure 2 shows that Chi-Square ( $\chi^2$ ) = 267,22 df = 246, P-value = 0,16838, RMSEA = 0,022. The result fulfilled the goodness of statistics. It can then be concluded that the assessment model of affective domain using SA accorded the data from the field.

Seen from the *loading factor* value of the indicator (all > 0,3) above, all the indicators constructing the components of assessment were valid. The items validity can be seen from the *loading factor* for each item with the *loading factor* value > 0,3, thus, item is considered to be valid.

On the PA-based model of instrument, the result is: *Chi-Square* ( $\chi^2$ ) = 151,55, df = 126, P-value = 0,06015, *Root RMSEA* = 0,034. The analysis result using CFA fulfilled the *goodness of statistics*. Thus, it could be said that the assessment model suits the data

from the field. In other words, the model developed was fit. A fit model means the model tried-out is supported by the data from field. The other requirements for a fit model are: the significance value is or  $p \geq 0.05$ , and  $RMSEA < 0.05$ . This means the model is fit (Nunnally & Bernstein, 1994).

The use of validity criteria seen from *>Loading factor* minimum 0,30 as the consideration is based on the idea from Nunnally & Bernstein (1994). The valid items were used in the instrument for assessing the students' affective domain. There is an item (item 7) which was invalid with the value of *loading factor* 0,05 ( $< 0,30$ ), as seen in the figure below.

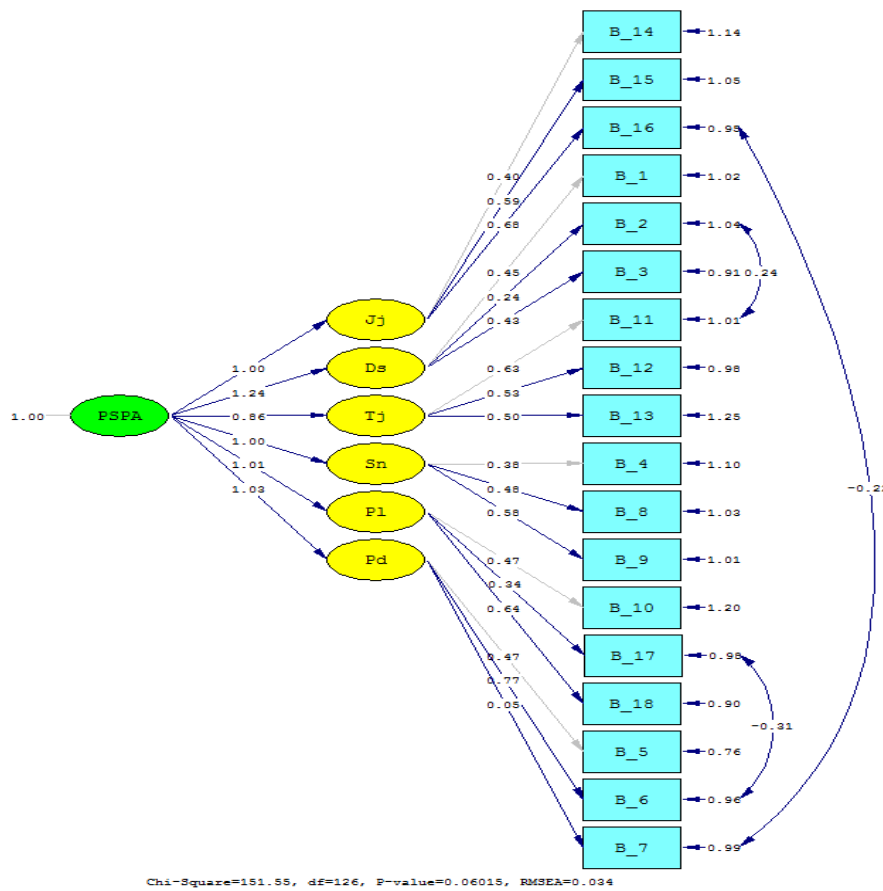


Figure 3  
The Result of PA-based Instrument Analysis Using CFA Second Order

The instrument was not estimated only from its validity but also from its stability through reliability estimation. Reliability estimation was conducted using *Alpha*

*Cronbach* and the result showed that the value of reliability ranged from 0,798-0,886 (>0,70). The result can be seen at Table 4.

Table 4  
Reliability Statistics

Instrument	Cronbach's Alpha	N of Items
Self-Assessment	0.886	24
Peer-Assessment	0.798	18

This is based on the idea that the criteria of instrument reliability are defined when the coefficient grouped items (alpha reliability) reaches 0,70 or above. Therefore, the instrument is said to be reliable (Nunnally & Bernstein, 1994; Gullickson, 2007).

The reliability of the assessment model shows the stability, consistency, and reliability to describe the indicators as they are (Nunnally & Bernstein, 1994; Mistar, 2011). Such assessment is reliable if it is treated to the same subjects it will yield the same results, although the time and condition are different. Though reliability has other terms such as trust, consistency, and so forth, the main idea of the concept is the same: how far the result of measurement is trusted. Based on figure 2 obtained all of items were valid and reliable so that these items can give information accurately about component to be measured. Instrument that valid and reliable can give information holistically about education programs (Andrian, Kartowagiran, & Hadi, 2018).

To get results maximum in doing the research, the researchers require a valid and reliable instrument since valid instruments can provide accurate information on the weaknesses and deficiencies a education programs. Instruments must be valid and reliable in content and construct because it's an important part of the development of the instrument (Wynd, Schmidt, & Schaefer, 2003). the good and bad information obtained in the field is very dependent on valid and reliable instruments the validity and reliability of the instrument will ensure that measurements will go well and provide information as expected (Wright & Craig, 2011). Reliability and validity are part of important component or the entities required in instrumental development (Burton & Mazerolle, 2011). Researchers who develop instruments in the field of education, psychology, evaluation, measurement, assessment and other social need to pay attention to the importance of the concept of validity and reliability so that information needed to support research can be obtained accurately (Andrian, Kartowagiran, & Hadi., 2018) Effective instrument is valid and reliable instrument. Valid and reliable instrument will make it easier for users to access information that obstacle a program's success (Tooth, Nielsen, & Armstrong, 2013).

## CONCLUSION

Based on the discussion, it can be concluded as follows: construct of the assessment instrument for affective domain which covered *honest, disciplined, responsible, polite, caring, and confident* was developed in 2 models: *self assessment (SA)* and *peer assessment (PA)*. There were 12 indicators and 24 items constructed based on SA model. The developed instrument covered scale of social behavior and modified the *Likert*

scale. Evidence of the content validity was strong; it was supported by the *Aiken* index of the indicators and items measurement which ranged from 0,750 to 1,000. Evidence of the construct validity was done using CFA *second order*. It yielded a fit model for the instrument for assessing children's affective domain, with the values of *loading factor* > 0,30. The instrument reliability developed was considered to be good, as seen from the coefficient *Alpha Cronbach* value which reached 0,788 to 0,886. The further research is expected to be able to find the new indicator about peer-assessment and self-assessment through more in-dept research using qualitative research.

#### REFERENCES

- Andrian, D., Kartowagiran, B., & Hadi, S. (2018). The Instrument Development to Evaluate Local Curriculum in Indonesia. *International Journal of Instruction*, 11(4), 922–934. <https://doi.org/10.12973/iji.2016.9115a>
- Baird, J. A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: fields apart? *Assessment in Education: Principles, Policy and Practice*, 24(3), 317–350. <https://doi.org/10.1080/0969594X.2017.1319337>
- Burton, L. J., & Mazerolle, S. M. (2011). Survey Instrument Validity Part I: Principles of Survey Instrument Development and Validation in Athletic Training Education Research. *Journal of Athletic Training Education*, 6(1), 27–35.
- Dwyer, C. A. (1998). *Assessment and Classroom Learning: theory and practice. Assessment in Education: Principles, Policy & Practice* (Vol. 5). <https://doi.org/10.1080/0969595980050109>
- Evans, R., Elwyn, G., & Edwards, A. (2004). Learning in practice Review of instruments for peer assessment of physicians, 328(May), 1–5.
- Foley, S. (2013). Student views of peer assessment at the International School of Lausanne. *Journal of Research in International Education*, 12(3), 201–213. <https://doi.org/10.1177/1475240913509766>
- Gullickson, A. M. (2007). Review of Practical Assessment, Research, & Evaluation, Volume 10. *Journal of MultiDisciplinary Evaluation*, 3(4), 199–203.
- Hadi, S., & Andrian, D. (2018). 2018. *The New Educational Review*, 53(3), 250–260.
- Jöreskog, K. G., Olsson, U. H., & Wallentin, F. Y. (2016). *Multivariate Analysis with LISREL*. Switzerland: Springer. <https://doi.org/10.1080/00210860108702002>
- Kanioglou, A., Tsorbatzoudis, H., & Barkoukis, V. (2005). Socialization and Behavioral Problems of Elementary School Pupils With Developmental Coordination Disorder. *Perceptual and Motor Skills*, 101, 163–173.
- Kritikos, V. S., Woulfe, J., Sukkar, M. B., & Saini, B. (2011). Intergroup peer assessment in problem-based learning tutorials for undergraduate Pharmacy students. *American Journal of Pharmaceutical Education*, 75(4). <https://doi.org/10.5688/ajpe75473>
- Logan, B., & Ed, D. (2015). Reviewing the value of self-assessments: Do they matter in the classroom? *Research in Higher Education Journal*, 29(September), 1–11.

- McCoach, D. B., Gable, R. K., & Madura, J. P. (1986). *Instrument Development in the Affective Domain*. New York: Springer.
- Mistar, J. (2011). A Study of the Validity and Reliability of Self-Assessment. *TEFLIN Journal*, 22(1), 45–58. <https://doi.org/10.15639/TEFLINJOURNAL.V22I1/45-58>
- Noonan, B., & Duncan, C. R. (2005). Peer and Self-Assessment in High Schools. *Practical Assessment, Research and Evaluation*, 10(17), 1–8. <https://doi.org/https://doi.org/10.1016/j.biopha.2017.09.089>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory*. New York: Tata McGraw-Hill Publishing Co. Ltd.
- Pandra, V., & Mardapi, D. (2017). Development of Mathematics Achievement Test for Third Grade Students at Elementary School in Indonesia. *International Elektronik Journal of Mathematics Education*, 12(8), 769–776.
- Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research & Evaluation*, 11(10), 1–13.
- Sluijsmans\*, D. M. A., & George Moerkerke\*, J. J. G. van M. and F. J. R. C. D. (2001). Studies in Educational Evaluation IN PROBLEM BASED LEARNING. *Studies in Educational Evaluation*, 27(1), 153–173.
- Stiggins, Richard, J. (1999). Assessment , Studen Confidence , and Scho. *The Phi Delta Kappan*, 81(3), 191–198.
- Tooth, J. A., Nielsen, S., & Armstrong, H. (2013). Coaching effectiveness survey instruments: Taking stock of measuring the immeasurable. *Coaching*, 6(2), 137–151. <https://doi.org/10.1080/17521882.2013.802365>
- van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal variables and conceptions. *Learning and Instruction*, 20(4), 280–290. <https://doi.org/10.1016/j.learninstruc.2009.08.010>
- Wen, M. L., Tsai, C. C., & Chang, C. Y. (2006). Attitudes towards peer assessment: A comparison of the perspectives of preservice and inservice teachers. *Innovations in Education and Teaching International*, 43(1), 83–92. <https://doi.org/10.1080/14703290500467640>
- Wright, P. M., & Craig, M. W. (2011). Tool for assessing responsibility-based education (TARE): Instrument development, content validity, and inter-rater reliability. *Measurement in Physical Education and Exercise Science*, 15(3), 204–219. <https://doi.org/10.1080/1091367X.2011.590084>
- Wynd, C. A., Schmidt, B., & Schaefer, M. A. (2003). Two quantitative approaches for estimating content validity. *Western Journal of Nursing Research*, 25(5), 508–518. <https://doi.org/10.1177/0193945903252998>
- Zhang, Z., & Burry-Stock, J. A. (2003). Classroom Assessment Practices and Teachers' Self-Perceived Assessment Skills. *Applied Measurement in Education*, 16(4), 323–342. [https://doi.org/10.1207/S15324818AME1604\\_4](https://doi.org/10.1207/S15324818AME1604_4)