Pegem Journal of Education and Instruction, Vol. 13, No. 1, 2022 (pp. 193-201)

RESEARCH ARTICLE WWW.PEGEGOG.NET

Developing Self-Assessment Instruments of Affective Domain on Belief and Morality (Aqidah Akhlak) Subject in *Madrasah Tsanawiyah*

Rohmad[1], Abu Dharin[2], Donny Khoirul Azis[3]

[1-3]State Islamic University (UIN) Prof. KH. Saifuddin Zuhri Purwokerto, Jl. A. Yani No. 40 A Purwokerto 53126, Central Java, Indonesia

Abstract

This study aimed to find a valid and reliable self-assessment procedure and instrument to measure the spiritual and social attitude domain of Belief and Morality (Aqidah Akhlak) subject in Madrasah Tsanawiyah by implementing Borg & Gall's research and development model. The instruments' content validity was analyzed using the Aiken formula, and the construct validity used Explanatory Factor Analysis (EFA), while the instrument reliability was conducted by using Alpha Cronbach. The results showed that the instrument's content validity and construct validity were valid. In addition, there was a positive relationship between morality to Allah and the Messenger of Allah towards the morality of parents, teachers, themselves, and others. The 2013 curriculum demands diverse evaluation instruments to measure affective, cognitive, and psychomotor learning outcomes comprehensively. The instruments developed in this study are expected to be used as a reference for teachers in developing instruments to evaluate learning outcomes in the domain of attitude (affective), both spiritual and social attitudes. The instruments can measure the achievement of learning outcomes regarding the affective domain in Madrasah Tsanawiyah or the equivalent.

Keywords: Instrument, Self-Assessment, Affective Domain, Aqidah Akhlak, Madrasah Tsanawiyah

Introduction

Aqidah (belief) and akhlak (morality) are the substance of Islamic law. Aqidah is the foundation and the base of faith; in addition, akhlak (morality) manifests faith in the behavior or good deeds. Akhlak (Moral) Education is a never-ending discourse of the learning process, using various names, such as ethical education (Curriculum 2013).

The main principle of curriculum development in 2013 is based on a competency-based curriculum model with standards of graduate's competence (SKL) set for one unit of education, education level, and educational program. In addition to having the main principles, the 2013 curriculum has three aspects of assessment, namely aspects of knowledge, aspects of skills, and aspects of attitudes and behaviors. The three elements of the assessment are then outlined in four core competencies (KI) such as core competence 1 (KI.1), namely the spiritual domain; core competence 2 (KI.2), such as the domain of social attitudes; core competence 3 (KI.3) such as the cognitive domain and core competence 4 (KI.4) such as the psychomotor domain.

As developed by Krathwohl et al. (1986), the structure of the affective domain is quite complicated; it means that the affective structure of elements is quite complex, and not all affective characteristics should be evaluated in schools. Some effective characteristics that need to be considered (measured and assessed) related to schools' Islamic education subjects, such as attitudes, interests, self-concepts, and values (Dikdasmen, 2003).

The main challenge is human development (human resource development) and natural resources. Qualified human resources are not only in the form of intelligence and skills but also attitude and mentality. The demands for the development of the evaluation system are strengthening in the group of subjects of Religious Education and Noble Morals. The subjects of Religious Education and Noble Morals are considered the main leading factors of other subject groups to help students internalize religious education, become accustomed to noble behavior, and be smart, skilled, and creative.

In the context of Islamic religious education, the development of learning evaluation is emphasized in the affective domain (Solichin, M, 2007), which is how the evaluation is directed to know the extent to which the passions, awards, and behaviors of learners have been in accordance with two main sources of Islam, namely in the Qur'an and as-Sunnah. Al-Abrasyi (1987) emphasized the aspect that moral education is very important for Islamic education objectives. Education and teaching process are not fulfilling the brains of students with all kinds of knowledge that they do not know, but educating their morals and souls, instilling a sense

Corresponding Author: rohmad@uinsaizu.ac.id

https://orcid.org: 0000-0002-1229-462X

How to cite this article: Rohmad, Dharin A, Azis DK (2022). Developing Self-Assessment Instruments of Affective Domain on Belief and Morality (Aqidah Akhlak) Subject in Madrasah Tsanawiyah. Pegem Journal of Education and Instruction, Vol. 13, No. 1, 2022, 193-201

Source of support: Nil.

Conflict of interest: None.

DOI: 10.47750/pegegog.13.01.21

Received : 29.04.2022

Accepted : 09.07.2022 **Published:** 01.11.2022

of fadhilah (virtue), familiarizing them with high modesty, preparing them for a holy life and having sincerity and honesty.

The main purpose of education is summarized in one word "fadhilah" (virtue), with the main mission sent by Muhammad as a prophet to complete morality (***Innamaa bu' itstu liutammima makaarim al-akhlaaq***) (al-Syaibani, 1979). But, in implementing the 2013 Curriculum and the assessment system, this cognitive goal has been prioritized in education in Indonesia, many educators are not concerned about other domains, and the teachers have not fully understood the assessment system (Retnawati et al., 2016) because of the assessment of the implementation of 2013 Curriculum. Based on the authors' observations in several Madrasah Tsanawiyah, there are still a lot of handbooks, or Student Worksheets (LKS) used as learning evaluation materials that only emphasize the cognitive aspects. The student performance related to psychomotor and affective competencies is often overlooked. Hall, R.A (2011) stated that affective assessment is often overlooked in learning practice, whereas educators can provide students with a complete and relevant educational experience and an attractive learning process. According to Suyanto (2010), ignoring the affective domain harms the development of learners, both individuals, and society as a whole. Students know a lot about something but lack attitude, interest, value system, and positive appreciation of what they know.

The learning of Aqidah Akhlak is inseparable from the implementation of evaluations conducted by the teachers. According to the objective conditions in the field during the study, Aqidah Akhlak does not emphasize the purpose of the assessment itself. One of the contributing factors is the lack of teachers to carry out the evaluations in a varied and continuous manner because they pursue the targets that must be achieved regardless of the expected quality of the material so that the students' ability in affective competencies is neglected.

Teachers realize that the affective domain (spiritual domain and the domain of social attitudes) is essential in learning Aqidah Akhlak. However, in fact, on its strength, most teachers do not assess the affective domain (the domain of spiritual and social attitudes), and the psychomotor domain implements relevant instruments. The assessment does not have an apparent reference and is considered an unstructured and planned instrument. The quality of educational evaluation in Madrasah Tsanawiyah, in general, is inseparable from the quality of the use of relevant assessment instruments for cognitive, affective, and psychomotor aspects. Speaking of assessment instruments certainly cannot be separated from how to develop assessment devices in the three aspects. Assessments conducted on these three aspects will result in the students who do master not only the knowledge but also have a commendable attitude and morality.

The reality of evaluation instruments that are currently widely used by teachers of Akhlak is evaluation in the cognitive domain; concrete manifestations of the achievement of student learning outcomes in a number of subjects are measured based on test results. Until now, the assessment instruments developed have not been able to measure the substance of student faith and behavior, so the instruments of evaluation of affective domains (spiritual domains and social attitude domains), moral subjects, and religions in madrasah Tsanawiyah need to be developed. The research aims to develop self-assessment instruments of an affective domain on belief and morality (aqidah akhlak) subject in madrasah tsanawiyah.

Literature Review

Self Evaluation

Self-evaluation refers to an individual's essential evaluation of abilities and competence (Peng et al., 2016; Bourke, 2017; Yan, 2019). In the context of an institution, self-evaluation is an evaluation conducted by the institution to collect the data, data analysis, and interpretation of results used for planning, development, improvement, and improvement of the institution's performance (Nuchron et al., 2013). At this point, the institution's improvement relates to the theories and research regarding the strategies for the positive changes in the educational system to achieve the goals (Mekhlafi & Osman, 2019). Self-evaluation is used to explore, understand, and be well aware of the institution's profile, including quality and current institutional conditions to be used as a basis for the institution to determine the target or desired future conditions (Nuchron et al., 2013). According to Hofman et al. (O'Brien et al., 2015), self-evaluation in institutions, for example, is seen in school self-evaluation, where the evaluation is an internal process used to ensure quality and improve the teaching and learning process and school performance. Self-evaluation is an essential requirement for school improvement (Smith &McCully, 2013). According to Bollaert et al. (Alzamil, 2014), self-evaluation in educational institutions is an essential need and tool in improving the performance of institutions to achieve a highly competitive position.

Self-evaluation has been stimulated to encourage sustainable quality improvement (Van der Bij et al., 2016; Carroll, 2020; To & Panadero, 2019). Self-evaluation is one of the essential aspects of the overall accreditation cycle with various roles and uses, including quality assurance (Sopingi et al., 2015). Self-evaluation becomes a milestone of a development, where development is a planned change, so it is necessary to understand how to conduct a comprehensive, structured, and systematic self-evaluation so that the results can be used as a basis in the planning process to achieve the desired goals, namely sustainable quality improvement (Luqman, 2017). The principle of implementation of goal-oriented self-evaluation refers to the criteria of success, the principle of benefits, and objectives; in other words, the purpose of self-evaluation is

implemented in reference to the goal to be achieved (Nuchron et al., 2013). Therefore, through self-evaluation, it can be seen the advantages and disadvantages that further, this deficiency becomes the purpose of improvement.

Affective Learning

According to Mosby's Medical Dictionary, affective learning is defined through learning skills, particularly as the acquisition of behaviors that express feelings in attitude, appreciation, and value (Bamidis, 2017). This is in line with Krathwohl et al.'s opinion that defines affective learning as an object that emphasizes the tone of feelings, emotions, or students' acceptance or rejection of subject matter (Gupta &Pandey, 2018; Kulkarni et al., 2018). Meanwhile, Hatfield et al. (Webster et al., 2013) define affective learning as the main proposed educational focus for guiding student approach/avoidance behaviors related to subject matter content. In other words, affective learning is learning that refers to learners' interests, attitudes, and motivations (Bamidis, 2017). In affective learning, the domain of learning, according to Mottet & Beebe (Webster et al., 2013), covers the underlying attitudes, beliefs, values, and emotions and is related to the knowledge and skills they are acquiring. Therefore, affective learning relates to the value that is difficult to measure because it relates to the person's awareness (Fatimah Kadir, 2015).

Krathwohl et al. regulate the domain of affection during the learning process in five levels, namely (1) accepting (described as a willingness to attend an event or event), (2) responding (learners are willing to react and participate in an activity), (3) assessing (the intention to accept or reject the event by positively or negatively), (4) organizing (when learners can determine which values are more dominant), (5) complex values (when learners decide to act following the accepted by respecting and making that practice or attitude part of their personality) (Phan, 2019; Fong & DeWitt, 2019). According to Duncan-Whitt (Omar et al., 2015), affective learning can evoke feelings such as threats or pleasure depending on the learner's learning experience. Therefore, according to Vermunt and Vermetten, affective learning outcomes are the feelings that arise during the learning process that create an emotional state that may positively, neutrally, or negatively affect the learning process (Duchatelet et al., 2018). Thus, affective learning is more related to attitudes that emphasize how a person chooses and sorts an action and considers whether the chosen one is beneficial or causes problems (Alifah, 2019).

Student Competence (Character)

Competence is the underlying characteristic of a person concerning the performance activities of an individual in his work or the basic character that has a causal relationship with the criteria referred to (Purwati, 2016). In the context of education, learning objectives are formulated based on indicators of the competency achievement, including cognitive, affective, and psychomotor competencies (Diani, 2015; Patmawati et al., 2020; Amalia & Suwatno, 2016) so that all subjects taught and studied must contribute to the formation of core competencies that are expected through the teaching and learning process. In other words, student competence is defined as each individual's ability to integrate knowledge, skills, and attitudes in accepting and practicing learning (Baartman &De Bruijn, 2011; Lapisa et al., 2017) so that the learning process can help the students establish a higher level of thinking (Nadarajan et al., 2020). Students' competence is obtained after experiencing treatment that includes aspects of knowledge, attitudes, and skills (Ayu et al., 2016). Furthermore, the students have an important role in fostering their learning process (Nasri et al., 2020). Thus, during the learning process, there must be efforts to nurture and increase students' motivation in learning so that the students can be encouraged to improve their abilities (Nur, 2018).

Competency supports the creation of character development of each student to improve the competence, skills, and understanding of the learning process (Baartman & De Bruijn, 2011). One of the indicators of competency is to create character development which is seen in educational objectives that not only focus on skills and knowledge that boil down to the creativity and competence of students in understanding science and science but also focus on instilling morality and ethics into students who bear fruit on good moral attitudes in the midst of a society in the future (Sutjipto, 2014). In line with this, competence in character formation is also seen in the domain of affective learning, in which, according to Supardi, competencies include characteristics of responsibility, cooperation, discipline, commitment, confidence, honesty in respecting the opinions of others, as well as the ability to control themselves where the form of assessment in the affective domain can use non-test instruments (Alifah, 2019). Therefore, to realize this goal, there needs to be a harmonious synergy between students, teachers, and school managers (Permatasari et al., 2013).

Methods

This research used a type of research and development (R&D). The research and development method is used to produce the product of affective self-assessment instrument (spiritual domain and social attitude domain) and tests the product's effectiveness. Borg and Gall (1983) explained, "The purpose of R & D is to bridge the gap between educational research and practice."

Field Testing

The design of the field testings to develop self-assessment instruments in the spiritual domain and the domain of social attitudes of Aqidah Akhlak Subjects in Madrasah Tsanawiyah in this study used five main steps they are (1) drafting the preliminary instrument (phase one), (2) conducting limited field testings in three Madrasah Tsanawiyah, (3) evaluating and revising phase one, (4) conducting wider field testings in fifteen Madrasah Tsanawiyah, and to (5) evaluating and revising phase 2. The five main steps are visualized as follows.

Research Design

This research conducted a research and development model (R&D) using Borg and Gall (1983) with ten development steps. They are research and information collecting, planning, developing a preliminary form of product, preliminary field testing, main product revision, main field testing, operational product revision, operational field testing, final product revision, and dissemination & implementation (Fig. 1).

Population and Sample/ Study Group/ Participants

The research population is students and teachers of madrasah tsanawiyah (MTs) in the Central Java province and Yogyakarta Special Region (DIY). Sample selection is done randomly. Central Java Province sample selected MTs in Banyumas, Tegal, and Kendal districts. DIY province sample selected MTs in Yogyakarta City and Bantul Regency. The first trial stage is selected 3 MTs. The 2nd trial stage selected 15 MTs. The sample list is presented in the table 1.

Data Collection Tools

The data collected consisted of quantitative data supported by qualitative data. Quantitative data was data from instrument test results in guidelines/observation guides for teachers and questionnaires for students (Reflection Sheet). At the same time, qualitative data is in the form of a collection of information in the form of words supporting data in this study.

Data Collection

Researchers conducted an in-depth exploration of the madrasah Tsanawiyah curriculum focused on Aqidah Akhlak subjects at the planning stage. Afterward, the theories of affective domain evaluation (spiritual and social attitudes) were investigated. At this planning stage, the main data collection technique implemented was documentation. At the test phase in the field, both in the first and second stage of the instruments tested ware the scale of attitudes in the form of observation guidelines and questionnaires. Observation guidelines were intended for teachers to make observations on student attitudes. At the same time, questionnaires were used to measure student attitudes based on student perception and recognition or, in other terms, self, report, and teacher assessment. The researchers observed teachers and students and randomly interviewed teachers or students to determine

Table 1: List of instrument test subjects

No	*School Name*	*Description*
1	MTs N 1 Banyumas	Field Testing I
2	MTs NU Purwokerto	Field Testing I
3	MTs Muhammadiyah Purwokerto	Field Testing I
4	MTs N 1 Kendal	Field Testing II
5	MTs NU 01 Cepiring Kendal	Field Testing II
6	MTs Muhammadiyah 1 Weleri Kendal	Field Testing II
7	MTs N 1 Tegal	Field Testing II
8	MTs NU Wakhid Hasyim Talang Tegal	Field Testing II
9	MTs Muhammadiyah Dukuhturi Tegal	Field Testing II
10	MTs N 1 Banyumas	Field Testing II
11	MTs Muhammadiyah 1 Purwokerto	Field Testing II
12	MTs NU 1 Purwokerto	Field Testing II
13	MTs N 1 Kota Yogyakarta	Field Testing II
14	MTs Nurul Ummah Kota Gede Yogyakarta	Field Testing II
15	MTs Muhammadiyah 1 Kota Yogyakarta	Field Testing II
16	MTs N 1 Bantul	Field Testing II
17	MTs NU 1 Bantul	Field Testing II
18	MTs Muhammadiyah 1 Bantul	Field Testing II

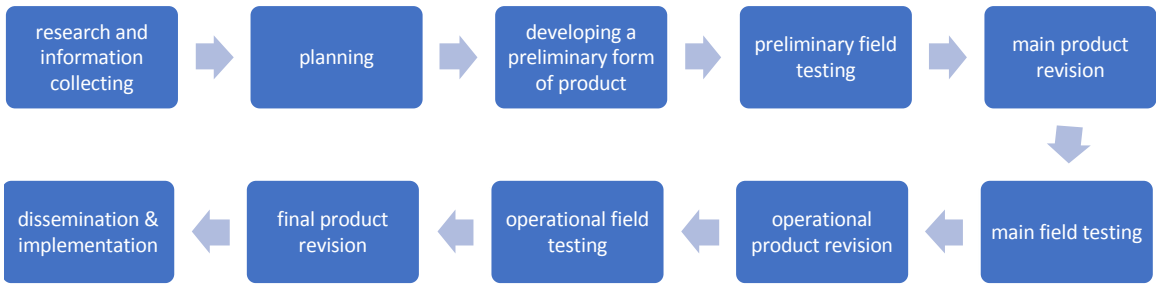

Fig. 1: Research Design

what they thought about the instruments being developed. The instrument was equipped with minutes of the meeting that the teacher must fill during the field testing.

Data Analysis

Instrument analysis was performed twice; they were analysis of field testing in phase I and the analysis of field testing in phase II. The purpose of the first field testing, which was also called a preliminary tryout (Hadari &Martin, 1995), was directed to: know the face validity, check the possibility of an instrument that is not clear, and examine the case of foreign words or terms so that it is not understood, examine the possibility of instruments that were too shallow in revealing indicators of achievement, and examining the possibility of instruments that were not relevant to the information they wanted to disclose as indicators of learning outcomes.

Empirically, testing the suitability of items with the grid (construct) could be tested with a total-item correlation analysis to see the contribution of items to the total variable. Calculating the correlation between attribute groups can obtain the discriminant validity (Munir, Abdul Razak, 2005). Items that measure the same attribute are expected to have a high correlation index, while items that measure different attributes have a low correlation index. Instruments that can make measurements carefully with a low level of measurement error will provide relatively consistent and stable measurement results (Purwanto, 2010).

Empirical validity testing is conducted with factor analysis to determine the construct through The Explanatory Factor Analysis (EFA) conducted through SPSS 23, in which if the KMO value is more than 0.50 and the anti-image correlation is more than 0.30; then, it can be said that the instrument has met the construct (Munir, Abdul Razak, 2005). While the reliability coefficient is done using Alpha Cronbach.

Result and Discussion

Preliminary Product Design

There are four steps in the preparation of the initial design, namely: identification of objectives or measuring areas, operationalization of concepts into behavioral indicators, writing of statement items, and review of statement items.

Field Testing and The Results of Phase I Testing

The field testing was conducted in 3 madrasah tsanawiyah, namely MTs Negeri 1 Banyumas, MTs Muhammadiyah Purwokerto, and MTs Ma'arif NU 01 Purwokerto. The instruments tested consisted of self-assessment of the spiritual domain and the domain of social attitudes and assessing the teacher's observation of the spiritual domain and social attitudes (affective domain). Before the field testing, the author coordinated and conducted simulations

Table 2: Test results of phase I: Instrument of self-assessment on spiritual domain and the domain of social attitudes Grade VII Madrasah Tsanawiyah (MTs)

Assessment Type	*Reliability*	*Number of items*	*Correlation items – Minimum Total of 0.3*
Self-assessment of Spiritual Domain	.83	93 items	85 items
Self-assessment of Social Attitudes Domain	.96	92 items	85 items

with all teachers of moral and religious subjects in 3 Madrasah Tsanawiyah on the technique of using instruments. The technique of using instruments for grade VII, VIII, and grade IX students is self-assessment spiritual domain and self-assessment in the domain of social attitudes, every ten students at each grade level so that in each school, there are 30 students multiplied by 3 Madrasah Tsanawiyah so that there are 90 students. The teacher explains and guides students in choosing instrument answers according to the actual conditions of students spontaneously. In terms of observation assessment instruments given to teachers of Aqidah Akhlak in MTs N 1 Banyumas, they were 3 people, teachers of Aqidah Akhlak in MTs Ma'arif NU 1 Purwokerto Barat and 1 teacher at MTs Muhammadiyah Purwokerto.

Based on the overall points of the question, the instrument of spiritual self-assessment with a total of 93 items is valid and reliable for 85 items because all the index of alpha reliability is above 0.3. Meanwhile, regarding the invalid and reliable items, there are eight items, namely items number: 18, 27, 39, 43, 57, 62, 71, and 83, because all alpha reliability indexes are below 0.3. At the same time, the self-assessment instrument of the domain of social attitudes showed that 92 items were valid and reliable for 85 items because all indexes of alpha readability were above 0.3. Meanwhile, in terms of the invalid and reliable items, there are seven items, namely items number: 22, 34, 45, 59, 67, 75, and 88, because all indexes of Alpha reliability are below 0.3. The overall reliability average of both types of assessment was 0.896. By using the minimum correlation criteria of each item with a total score of 0.3, then the points of the question that correlate less than 0.3 must be issued. By repeatedly conducting reliability testing by removing items with an index below 0.30, we obtained reliable instrument items while improving the overall reliability of the instrument. The items are excluded because they are below 0.3; then, they need to be revised before the next field-testing phase. The test results of the grade VII instrument self-assessment spiritual domain have reliability of 0.834 then, and the instrument belongs to a very high category. While the results of the grade VII field testing of social attitudes self-assessment instrument have a reliability of 0.825, then the instrument belongs to the category of very high.

Table 3: Test results of phase I: Instrument of self-assessment on spiritual domain and the domain of social attitudes, Grade VIII Madrasah Tsanawiyah (MTs)

Assessment Type	*Reliability*	*Number of items*	*Correlation items – Minimum Total of 0.3*
Self-assessment of Spiritual Domain	.98	120 items	100 items
Self-assessment of Social Attitudes Domain	.96	119 items	100 items

From the results of the evaluation instrument on-field testing using spiritual domain self-assessment and self-assessment of the domain of social attitudes, grade VIII MTs in phase I, the overall reliability of the instrument is as follows.

According to the overall points of the question, the instrument of spiritual self-assessment of the spiritual domain consists of 120 items that are valid and reliable 100 items because all the indexes of alpha reliability are above 0.3. In comparison, the invalid and reliable there are 20 items, namely items number 3, 7, 11, 18, 23, 28, 32, 39, 43, 46, 51, 57, 63, 69, 71, 77, 83, 96, 113 and 118 because all Alpha reliability indexes are below 0.3. As a self-assessment instrument in the domain of social attitudes amounting to 119 items that are valid and reliable 100 items because all the index of alpha reliability above 0.3, while the invalid and reliable there are 19 items namely point number 6, 12, 22, 27, 31, 34, 45, 48, 51, 59, 60, 67.72, 75, 85, 92, 103, 112 and 115, as all Alpha reliability indexes are below 0.3. The overall reliability average of both types of assessment was 0.964. Using the minimum correlation criteria of each item with a total score of 0.3, the points of the question that correlate less than 0.3 must be excluded by conducting reliability repeatedly testing by removing items that have an index below 0.30 obtained reliable instrument items while improving the overall reliability of the instrument. The items are excluded because below 0.3 need to be revised before the next phase of field testing. The results of the grade VIII field testing of spiritual domain self-assessment have reliability of 0.976 then. The instrument belongs to a very high category. The results of the grade VIII field testing of social attitudes self-assessment have a reliability of 0.951; then, the instrument belongs to a very high category.

From the results of the test instrument evaluation of the affective domain by means of spiritual self-assessment and self-assessment of the domain of social attitudes, in grade IX MTs in stage 1, the overall reliability of the instrument is as in Table 4.

Investigating the overall points of the question, the instrument of spiritual domain self-assessment consists of 92 items that are valid and reliable for 85 points because all alpha reliability index is above 0.3. In comparison, the invalid and reliable are seven items, namely items number 17, 22, 35, 47, 64, 73, and 89, because all alpha reliability indexes are below 0.3. At the same time, the self-assessment instrument of the domain of social attitudes consists of 102 valid and reliable items for 90 items because all the indexes of alpha reliability are above 0.3. In comparison, the invalid and reliable items are 12 items, namely item number: 35, 39, 42, 47, 52, 57, 61, 63, 75, 85, 91, and 112, because all the alpha reliability index is below 0.3. The overall reliability average of both types of assessments was 0.864. By using the minimum correlation criteria of each item with a total score of 0.3, then the points of the question that have a correlation of less than 0.3 must be excluded. By conducting reliability testing and repeatedly removing items with an index below 0.30, it obtained reliable instrument items and, at the same time, improved the overall reliability of the instrument. The items are excluded because below 0.3 need to be revised before the next phase of field testing. The results of the grade IX field testing assessment of the spiritual domain have a reliability of 0.773. Then, the instrument belongs to the high category. The results of the grade IX field testing of self-assessment in the domain of social attitudes have reliability of 0.955. Then, the instrument belongs to a very high category.

Table 4: Test results of phase I: Instrument of self-assessment on spiritual domain and the domain of social attitudes Grade IX MTs

Assessment Type	*Reliability*	*Number of items*	*Correlation items – Minimum Total of 0.3*
Self-assessment of Spiritual Domain	.77	92 items	85 items
Self-assessment of Social Attitudes Domain	.96	102 items	90 items

Revising Preliminary Design

Based on the test results of grade VII, VIII, and IX instruments, several items need to be revised before the phase II field testing. The initial design revision was conducted by removing instrument items with an index below 0.30 so that reliable instrument items are obtained while improving the overall reliability of the instrument.

Product Design Phase II

This section describes the instrument made in the second stage, which includes the steps of preparation and the details of the questionnaire/scale. The first step is to review the objectives and measurement area by referring to the instrument's preparation on stage one, namely the syllabus of moral and religious subjects issued by the Ministry of Religious Affairs. The second step is the instrument's preparation by considering the results of the preparation of the instrument as well as the first stage of field testing. In the second phase, several changes and revisions were made by studying the undeveloped content,

and the items of the instrument that fell in phase one were revised.

Results of the Testing on Phase II

The implementation of the second phase of the field testing was expanded. The expansion was carried out by increasing the number of madrasahs used and the population taking of each madrasah. Instrument validity testing is conducted with factor analysis. Before factor analysis, KMO and Bartlett's tests were conducted to determine the instrument's feasibility. The validity of the structure of the grade VII spiritual domain self-assessment instrument is done using EFA with a value of KMO = 0.649, self-assessment of the domain of social attitudes with a value of KMO = 0.706 and each item has an anti-image coefficient of more than 0.5 which means it has met the requirements for factor analysis. The validity of the structure of the grade VIII spiritual domain self-assessment instrument is done using EFA with a value of KMO= 0.819, self-assessment of the domain of social attitudes with a value of KMO = 0.674, and each item has an anti-image coefficient more than 0.5 which means it has met the requirements for factor analysis. The validity of the structure of the grade IX spiritual domain self-assessment instrument is done using EFA with a value of KMO= 0.629, self-assessment of the domain of social attitudes with a value of KMO = 0.695, and each item has an anti-image coefficient of more than 0.5 which means it has met the requirements for factor analysis. With the help of SPSS Version 23.0 application, for the reliability of the spiritual domain self-assessment instrument grade VII of 480 respondents Cronbach's Alpha value of 0.834, while the reliability of the instrument of self-assessment in the domain of social attitudes of 480 respondents Cronbach's Alpha value of 0.959. For the reliability of the instrument of self-assessment of the spiritual domain grade VIII of 450 respondents, Cronbach's Alpha value of 0.976, while the reliability of the instrument of self-assessment in the domain of social attitudes of 450 respondents values Cronbach's Alpha 0.951. For the reliability of the instrument of spiritual self-assessment grade IX from 420 respondents, Cronbach's Alpha value of 0.773, while the reliability of the instrument of self-assessment in the domain of social attitudes of 420 respondents Cronbach's Alpha value of 0.955.

Based on the data above, when the reliability value of spiritual self-assessment instruments is recapitulated from grades VII, VIII and IX obtained an average of 0.861, the calculation result of 0.834 + 0.976 + 0.773 = 2.583 divided by 3. While the average reliability value of the social attitude self-assessment instrument is 0.955, the calculation result of 0.959+0.951+0.955= 2.865 is divided by 3.

Thus, the product in the form of an instrument of evaluation of the affective domain of Aqidah Akhlak Subject, both spiritual self-assessment, and self-assessment of the domain of social attitudes, has been proven and successfully developed. The content of Aqidah Akhlak in MTs includes the pillars of faith, namely faith in God, faith in angels, faith in the book, faith in the Messenger of God, and faith in the Day of Resurrection, and faith in ***qadha*** and ***qadar***. This aspect of religion becomes the basis for giving rise to morality for others. Morality to Allah and morality to the Messenger of Allah is the basis of faith that from both aspects gives rise to morality to the other. In other words, morality to parents, teachers, themselves, and others manifests morality to God and His Messenger. By associating the aspects of morality to God and morality to the Messenger of God with four other aspects of morality, namely to parents, teachers, themselves, and others, the results of analysis for the answer to self-assessment of the spiritual domain can be visualized as follows.

Discussion

The development of assessment instruments must be tailored to the domain of learning objectives. The test assessment instrument is precisely used to measure the learning outcomes of the cognitive and psychomotor domains. While the affective domain is more appropriate to use non-test assessment instruments (Chen et al., 2021; Gabriel et al., 2021; Karimova & Csapó, 2021). There are several non-test assessment techniques that can be used, including; observation, questionnaire, peer judgment, and self-assessment (Nygaard et al., 2020). Spiritual attitude is related to man's attitude towards God. The technique of observation assessment, questionnaire, and peer assessment is only able to observe the manifestation of man's relationship with God in observable aspects. While the spiritual aspect that does not seem can only be done with self-assessment techniques. Self-assessment uses the assumption that the one who knows the most about a person's spiritual condition is himself. The limitation of self-assessment is related to a person's "honesty" to assess his spiritual condition. The development of self-assessment instruments that are able to honestly reveal one's spiritual condition needs to continue to be developed (Rohmad, 2019; Zamora-Polo & Sánchez-Martín, 2019).

Conclusion

Based on the study results, it can be concluded that in developing the instrument of self-assessment evaluation of the affective domain (spiritual domain and the domain of social attitudes) in the subjects Aqidah Akhlak in Madrasah Tsanawiyah includes instruments of self-assessment spiritual domain and self-assessment area of social attitudes. The development of affective domain evaluation instruments is carried out in 10 steps, namely: (a) the study of Aqidah Akhlak subject literature and the development of affective domain evaluation instruments in Madrasah Tsanawiyah; (b) review of the competence standards of graduates (SKL),

core competencies (KI) and essential competencies (KD) of Aqidah Akhlak subjects; (c) development of learning outcome indicators; (d) selection of affective domain evaluation instruments; (e) drafting affective domain evaluation instruments; (f) phase one: field testing at 3 Madrasah Tsanawiyah; (g) evaluation and improvement of phase one test results; (h) phase two: field testing at 7 Madrasah; (i) analysis of the field testing on phase two; (j) improvement of the field testing results on phase two. There are two affective domain evaluation instruments developed, namely the instrument of spiritual domain self-assessment intended for madrasah Tsanawiyah students according to their level, namely grade VII, VIII, and IX, and the instrument of self-assessment in the domain of social attitudes intended specifically for students of grade VII, VIII and IX. Instruments prepared by each grade are designed by considering competence and content in accordance with the grade level. After a phase, I and II field testing, overall instruments from grades VII, VIII, and IX have met the KMO-MSA (Kaiser-Meyer-Olkin and Measure of Sampling Adequacy) test, as the results are more than half and with a significant value (sig) or the chance (p) is less than a half. As such, the items analyzed in the Factor Analysis are eligible for implementation.

Based on the data of validity and reliability of affective self-assessment (spiritual domain self-assessment and self-assessment of the domain of social attitudes) above, the instrument is worthy of being used and disseminated. In the 2013 Curriculum implementation, it demands multi techniques used in measuring comprehensive learning outcomes, and the resulting instrument can be used as a reference and consideration for teachers in developing learning outcome evaluation instruments in the affective domain of spiritual and social attitudes, especially to measure the achievement of learning outcomes in the domain of social attitudes. The product instruments can be used to measure the achievement of affective domain learning outcomes in Madrasah Tsanawiyah. In its implementation, measurement needs to be done several times. The person in charge of the educational institution, namely the Head of Madrasah Tsanawiyah and the teachers, can coordinate the right implementation time to take measurements. The content and scope of the instrument can be adjusted in line with the implementation of the content of the 2013 curriculum.

References

Athiyah, A.M. (1987). *Dasar-dasar Pokok Pendidikan Islam* [The Principles of Islamic Education] (B.A. Gani & D. Bahry, Trans.). Jakarta: Bulan Bintang.

Al-Toumy, A.O.M. (1979). *Falsafah Pendidikan Islam* [The Philosophy of Islamic Education], (H. Langgulung, Trans.). Jakarta: Bulan Bintang.

Alifah, F.N. (2019). Pengembangan strategi pembelajaran afektif [The Development of Affective Learning Strategy]. *Tadrib: Jurnal Pendidikan Agama Islam*, *14*(2): 11-26.

Alzamil, Z.A. (2014). Quality improvement of technical education in Saudi Arabia: Self-evaluation perspective. *Quality Assurance in Education*, *22*(2): 125-144.

Ayu, T., Asrizal, A., & Kamus, Z. (2016). Pengaruh LKS IPA terpadu berbasis WEB dengan mengintegrasikan nilai karakter pada materi GLSTSGPS terhadap kompetensi siswa kelas VIII SMPN 8 Padang [The Effect of WEB-based Integrated Science Worksheets by Integrating Character Values in the GLSTSGPS Material on the Students' Competence of Grade VIII in SMPN 8 Padang]. *Pillar of Physics Education*, *7*(April), 193–200.

Baartman, L.K.J., & Bruijn, E.D. (2011). Integrating knowledge, skills and attitudes: Conceptualising learning processes towards vocational competence. *Educational Research Review*, *6*(2):125-134.

Bamidis, P.D. (2017). Affective learning: Principles, technologies, practice. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-67615-9_1

Bourke, R. (2017). Self-assessment to incite learning in higher education: Developing ontological awareness. *Assessment & Evaluation in Higher Education*, 10512: 1-13.

Carroll, D. (2020): Observations of student accuracy in criteria-based self-assessment. *Assessment & Evaluation in Higher Education*, *45*(8): 1088-1105.

Diani, R. (2015). Pengembangan Perangkat Pembelajaran Fisika Berbasis Pendidikan Karakter dengan Model Problem Based Instruction [Development of Character Education-Based Physics Learning Tools with Problem-Based Instruction Model]. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, *4*(2): 241-253.

Dikdasmen. (2003). *Kompeteni guru sekolah lanjutan tingkat pertama (bahan rujukan pelatihan terintegrasi berbasis kompetensi)* [The competence of junior high school teachers (competency-based integrated training reference material)]. Jakarta: Direktorat Pendidikan Lanjutan Pertama.

Duchatelet, D., Bursens, P., Donche, V., Gijbels, D., & Spooren, P. (2018). Student diversity in a cross-continental EU simulation exploring variation in affective learning outcomes among political science students. *European Political Science, 17*: 601-620

Kadir, F. (2015). Strategi Pembelajaran Afektif Untuk Meningkatkan Pendidikan Masa Depan [Affective Learning Strategies To Improve Future Education]. *Al-Ta'dip*, *8*(2), 147.

Fong, C.S., & DeWitt, D. (2019). Developing Intercultural communicative competence: Formative assessment tools for Mandarin as a foreign language. Malaysian *Journal of Learning and Instruction, 16*(2), 97-123.

Borg, W.R. & Gall, M.D (1983). *Educational Research: An Introduction* (Fourth edition). Boston, New York: Person Education. Inc.

Borg, W.R. & Gall, M.D (2003). *Educational Research: An Introduction* (Seventh edition). Boston, New York: Person Education. Inc.

Chen, F. R., Fung, A. L. C., & Raine, A. (2021). The cognitive, affective, and somatic empathy scales (CASES): Cross-cultural replication and specificity to different forms of aggression and victimization. *Journal of Personality Assessment*, *103*(1), 80–91. https://doi.org/10.1080/00223891.2019.1677246

Gabriel, E. T., Oberger, R., Schmoeger, M., Deckert, M., Vockh, S., Auff, E., & Willinger, U. (2021). Cognitive and affective Theory of Mind in adolescence: developmental aspects and associated neuropsychological variables. *Psychological Research*, *85*(2), 533–553. https://doi.org/10.1007/s00426-019-01263-6

Gupta, M., & Pandey, J. (2018). Impact of Student Engagement on Affective Learning: Evidence from a Large Indian University. *Current Psychology*, *37*: 414-421.

Hall, R.A. (2011). Affective assessment: The missing piece of the educational reform puzzle. *Delta Kappa Gamma Bulletin*, 77(2). Retrieved from http://www.questia.com/library/1P3-2257394971.

Karimova, K., & Csapó, B. (2021). The Relationship Between Cognitive and Affective Dimensions of Reading Self-Concept With Reading Achievement in English and Russian. *Journal of Advanced Academics*, *32*(3), 324–353. https://doi.org/10.1177/1932202X21995978

Krathwohl, G.R. (1986). *Instrument Development in the Affective Domain*. Boston: Kluwer-Nijhoff Publishing.

Lapisa, R., Basri, I.Y., Arif, A., & Saputra, H.D. (2017). Peningkatan kompetensi siswa melalui pelatihan auto cad [Increasing student competence through auto cad training]. *Invotek: Jurnal Inovasi Vokasional dan Teknologi, 17*(2): 119-126.

Luqman, A. (2017). Posisi Evaluasi Diri dalam Sistem Penjaminan Mutu Pendidikan Tinggi [Positioning Self Evaluation in Higher Education Quality Assurance System]. *Cendekia: Journal of Education and Society*, *15*(2): 37.

Mekhlafi, A.M.A., & Osman, M.E.T. (2019). The effect of a holistic school improvement model in enhancing school effectiveness in Oman. *Malaysian Journal of Learning and Instruction, 16*(2), 187-200.

Munir, A.R. (2005). Aplikasi Analisis Jalur (Path Analisis) dengan Menggunakan SPSS Versi 12 [Path Analysis Application (Path Analysis) Using SPSS Version 12]. *Journal Laboratorium Kompetensi Jurusan Manajemen Universitas Hasanuddin*.

Nadarajan, T., Khairi, H.O.M., & Banu, M.N.N. (2020). Using forum theatre to develop various levels of thinking skills among moral education students in secondary school. *Malaysian Journal of Learning and Instruction, 17*(2), 167-194.

Nasri, N.M, Halim, L., & Talib, M.A.A. (2020). Self-directed learning curriculum: Students' perspectives of university learning experiences. *Malaysian Journal of Learning and Instruction, 17*(2), 227-251.

Nuchron, N., Soenarto, S., & Sudarsono, F. (2013). Model evaluasi diri sekolah menengah kejuruan di Daerah Isitimewa Yogyakarta [Vocational high school self-evaluation model in Yogyakarta Special Region]. *Jurnal Pendidikan Vokasi*, *3*(1): 80-89.

Nur, T. (2018). Upaya peningkatan pembelajaran Bahasa Arab dengan penerapan manajemen berbasis sekolah (MBS) [Efforts to improve Arabic learning by implementing school-based management (MBS)]. *Shaut Al Arabiyyah*, *6*(1): 68-84.

Nygaard, J., Colli, M., & Wæhrens, B. V. (2020). A self-assessment framework for supporting continuous improvement through IoT integration. *Procedia Manufacturing*, *42*(2019), 344–350. https://doi.org/10.1016/j.promfg.2020.02.079

O'Brien, S., McNamara, G., & O'Hara, J. (2015). Supporting the consistent implementation of self-evaluation in Irish post-primary schools. *Educational Assessment, Evaluation, and Accountability*, *27*: 377-393.

Omar, N. N., Peng, L. Y., & Norishah, T. P. (2015). Educational online games: Implementing MMO approach to enhance users' affective learning experiences. *ACM International Conference Proceeding Series*, *21*: 1.

Patmawati, Nurhayati, & Latjuba, A. Y. (2020). Internalisasi aplikasi mind map pada media pembelajaran guru untuk meningkatkan kompetensi [Internalization of the mind map application on teacher learning media to improve competence]. *Jurnal Ilmu Budaya*, *8*, 302–312.

Peng, J. I. A. X. I., Li, D., Zhang, Z., Tian, Y. U., Miao, D., Xiao, W. E. I., & Zhang, J. (2016). How can core self-evaluations influence job burnout? The key roles of organizational commitment and job satisfaction. *Journal of Health Psychology*, *21*(1): 50-59.

Phan, V.T. (2019). Affective learning objectives in online courses. *ACM International Conference Proceeding Series*.

Purwati. (2016). Peningkatan kompetensi dan hasil belajar bahasa indonesia materi teks percakapan melalui metode pqrst siswa kelas VI SDN Ngastorejo [Improving the competence and learning outcomes of Indonesian language conversational text material through the method of pqrst students of class VI SDN Ngastorejo]. *Jurnal Ilmiah Pendidikan Dasar, III*(1), 43–52.

Retnawati, H., Hadi, S., & Nugraha, A. C. (2016). Vocational high school teachers' difficulties in implementing the assessment in curriculum 2013 in Yogyakarta Province of Indonesia. *International Journal of Instruction*, *9*(1), 33–48.

Rohmad, R. (2019). Authentic Assessment of Affective Domain in Islamic Education and Character Building at A Curriculum 2013 Pilot Project Elementary School in Banyumas Regency Central Java Indonesia. *International Conference of Moslem Society*, *3*, 69–79. https://doi.org/10.24090/icms.2019.3028

Smith, R., & McCully, A. (2013). Organizational self-evaluation and teacher education for community relations in a transforming society? In *European Journal of Teacher Education*, *36*(3): 320-345.

Solichin, Muchlis M. (2007). Pengembangan evaluasi pendidikan agama Islam berbasis ranah afektif [Development of an evaluation of Islamic religious education based on the affective domain]. *Tadris*, 2(1): 76-91.

Sopingi, S., Utami, E., & Amborowati, A. (2015). Prototype executive information system untuk mendukung evaluasi diri perguruan tinggi [Prototype of executive information system to support higher education self-evaluation]. *Data Manajemen Dan Teknologi Informasi (DASI)*.

Sutjipto. (2014). Dampak pengimplementasian Kurikulum 2013 terhadap performa siswa Sekolah Menengah Pertama [The impact of implementing the 2013 Curriculum on the performance of junior high school students]. *Jurnal Pendidikan Dan Kebudayaan*, *20*(2): 173-186.

Suyanto, (2010). *Refleksi dan Reformasi Pendidikan di Indonesia Memasuki Millenium III* [Reflections on and Educational Reform in Indonesia Entering the Third Millennium]. Yogyakarta: Adicita Karya Nusa.

To, J. & Panadero, E. (2019). Peer assessment effects on the self-assessment process of first-year undergraduates. *Assessment & Evaluation in Higher Education*, *44*(6): 920-932.

Van der Bij, T., Geijsel, F. P., & ten Dam, G. T. M. (2016). Improving the quality of education through self-evaluation in Dutch secondary schools. *Studies in Educational Evaluation*, *49*: 42-50.

Webster, C., Mîndrilă, D., & Weaver, G. (2013). Affective learning profiles in compulsory high school physical education: An instructional communication perspective. *Journal of Teaching in Physical Education*, *32*(1): 78-99.

Yan, Z. (2019): Self-assessment in the process of self-regulated learning and its relationship with academic achievement. *Assessment & Evaluation in Higher Education*, *112*(4):545–557

Zamora-Polo, F., & Sánchez-Martín, J. (2019). Teaching for a better world. Sustainability and Sustainable Development Goals in the construction of a change-maker university. *Sustainability (Switzerland)*, *11*(15). https://doi.org/10.3390/su11154224