Conference Paper

Profesional Supervision Model: Development of Clinical Supervision Instruments for Teachers of Islamic Education Through a Multi-Faceted Rasch Model

Yuni Asdhiani[1], Ari Saptono[2], and Komarudin[3]

[1]Supervisor of Islamic Education at Bekasi Regency Junior High School
[2]Faculty economy UNJ, Jakarta
[3]State University of Jakarta, Jakarta

Corresponding Author:
Yuni Asdhiani
bunda.kaisharais@yahoo.co.id

Abstract

This study aims to develop (i) clinical supervision instruments for Islamic Religious Education teachers, and (ii) know the validity and reliability of clinical supervision instruments for Islamic Religious Education teachers. This research follows a quantitative non-experimental design. The subjects of this study were the Islamic Religious Education teachers in Bekasi Regency in the 2019/2020 school year. A survey method was used with a questionnaire using rubrics. The instrument development steps were: synthesizing theories; formulating constructs; developing dimensions and indicators; compiling instrument lattices; compiling and writing items; validating theory and empiricism as well as revising; empirical trials and readability; calculating instrument reliability; conducting repair and communicate to experts; reassemble the items of valid and reliable instruments to be made the final instrument. In the theoretical trial phase (the construct validity testing phase involving instrument development experts, clinical supervision experts, and linguists), data from the expert assessment results were analyzed with the help of the Rasch Multi-Rater Facet (MRFR) using facet software assistance

Keywords: clinical supervision, performance, multirater facet

1. Introduction

Education is the main key to creating quality human resources. In order to improve the quality of National Education in stages, planned and measured in accordance with the mandate of Law Number 20 Year 2003 concerning the National Education System and National Education Objectives, the correct process is required as stipulated. To guarantee the process, supervision is needed. Supervision needs to be done in order to guarantee the quality of education [1, 2]. Educational supervision is an important

instrument in the quality control of the educational process carried out in the education unit.

Educational Supervision is a scientific discipline that focuses on assessing the improvement of teaching and learning situations [2], activities that help, direct and inform teachers about what needs to be done or have been done and not only look for errors teaching teachers [3], but so that the implementation of the learning process conducted by teachers can be effective and efficient [4], can improve school quality [5], and develop guidance and professional processes for teachers [6].

Effective supervision is the main activity towards quality education. One of the main tasks that is closely related to developing teacher professionalism [6] and learning is academic supervision. Academic oversight is a supervisory function relating to the implementation of the tasks of coaching, monitoring, assessing, and guiding and training professional teachers, both in aspects of competence and its main tasks. Coaching activities in academic supervision of teachers concerning the ability of educators in managing the learning process to achieve learning objectives [12]

Teacher instructional competence is still considered low to date so that the impact results in poor academic achievement of students [8], here the need for supervisory actions which involves ensuring that teachers meet effective and efficient instructional responsibilities, teachers must demonstrate high academic standards through periodic checks to improve the quality of their work. One model of supervision that can help improve teacher professionalism is through clinical supervision. Clinical supervision is the process of facilitating a teacher's professional growth [4, 8–10], especially by observing teacher instructional practices, giving teachers feedback about classroom interactions and helping teachers use that feedback to make it more effective teaching [7], and improve teacher performance [3, 11, 13–15].

The practice of clinical supervision is still rarely performed by Supervisors because it is quite time consuming, there are still teachers who consider that clinical supervision is only to look for teacher weaknesses [3]. The results of the analysis of several journals on Clinical Supervision indicate that there is an effect of clinical supervision on teacher teaching performance [3, 8, 11, 13, 14, 16, 17]. Other findings suggest that clinical supervision has not been given enough by supervisors, because most of the time they are more focused on the administrative aspects [3] so that this is one of the obstacles to the implementation of clinical supervision in schools. The advantages of clinical supervision over academic supervision are in the process of how to improve learning, starting from before observation, when observing in class and the follow-up

process while academic supervision is enough to only see class observations writing the results and there is rarely follow-up improvement.

The position of the instrument in supervision has a very important function and role, especially in order to know the effectiveness of the learning process, knowing the extent to which the success of students in achieving learning objectives after they take the learning process. Requirements for a good measuring instrument are that it can provide accurate information to its users which must be valid, reliable and fair [18], meaning that the measuring instrument can distinguish the abilities of each teacher when the assessment process occurs in learning.

The better the quality of the instrument, the more useful, the greater the confidence in the score and the greater the confidence in making decisions based on these results. Therefore, it is very important to use high-quality instruments to do the assessment. Two main elements that explain the quality of the instrument are its validity and reliability [19–21].

The clinical supervision instruments possessed by the Supervisor of PAI according to the results of the observation do not yet have a standardized instrument according to the stages in clinical supervision namely the stage of pre-observation conference, implementation of observation and post-observation conference. The existing and implemented instrument is the learning observation stage, for the other stages the instruments do not yet have it. Based on interviews with several PAI Supervisors in provinsi jawa barat the implementation of clinical supervision is still rarely performed in schools, due to several reasons including the absence of standard instruments.

Performance appraisal by using the scoring scale allows the appraiser to provide a central value to master certain competencies because of the continuum value that gives more than two existing category values. The rating scale extends from not perfect to perfect. The scale, for example, is incompetent-somewhat competent-competent-highly competent. To minimize the subjectivity factor, the assessment needs to be done by more than one person, so that the assessment results are more accurate. There are three scale assessments: (1) numerical rank scales; (2) graphical rating scale; (3) Descriptive Graphic assessment scale [22].

Therefore, according to the researchers, urgent clinical supervision instruments were developed and validated so that the instrument can be used as a measuring instrument and a supervisory device in conducting clinical supervision of the teacher, supported by a valid and Reliable instrument that will support the improved learning process.

Based on the explanation above, the author is interested in conducting research on the development of professional supervision: development of clinical supervision

tools for teachers of Islamic education. Based on the background described earlier, the following problems can be formulated; (1) How to develop clinical surveillance instruments for PAI teachers?, (2) does PAI Teacher's clinical supervision equipment meet the requirements of validity and reliability? This study aims to answer the problem of formulation that has been stated above, namely; (1) Develop clinical supervision tools for PAI teachers; (2) Find out the point of the clinical supervision instrument from a PAI teacher that meets the validity and reliability.

2. Research Methodology

This research is a research development to produce empirical evidence from a non- test instrument (measuring instrument), as well as knowing the picture of professional supervision namely clinical supervision and teacher performanc The research method used in this study is a quantitative method of non-experimental design which is a research design that describes what has happened and examines the relationship without direct manipulation of the conditions experienced (Reseach Design and Reading Reseach Reports), which uses a survey method in the form of a questionnaire/questionnaire.

Population is the totality of objects or individuals where reference or inference is made based on the experience found in the sample [22]. Population is divided into two levels, namely the target population and available. The sample is defined as a number of respondents where the research data was obtained [23] Samples to be taken are PAI teachers who take part in teacher working groups in Bekasi Regency. Crocker and Algina (1986) recommend that the sample size required for parameter estimation based on item response theory varies between 200 to 1000 subjects [24]. Samples to be taken in this study are 300 samples.

Sampling uses a random sampling technique with the aim that the samples produced can be representative and use stratified random sampling. The use of sampling is not just for research samples but for trial samples of clinical supervision professional supervision instruments, the difference is the research samples are Islamic Religious Education teachers in Bekasi district through the teacher working group forum Teacher Working Groups at the Elementary School level, and Subject Teacher Deliberations at the Junior High School, Senior High School and Vocational High School.

The conceptual definition of clinical supervision is the coaching process through observing teacher performance in managing the learning process in the classroom

more effectively with several stages. The operational concept is to observe teacher performance in classroom learning which consists of pre-observation conference stages, learning observations, analysis sessions and post-observation conferences.

TABLE 1: Dimensions and Indicators of PAI Supervisor Clinical Supervision Implementation

No	Variable	Indicator	Number of questions
1	Pre-Observation Conference	1. Build a good relationship 2. Plan supervision objectives 3. Planning with the teacher 4. Planning observations	4
2	Observation Implementation	1. Observe the initial activities 2. Observe core activities 3. Observe the final activity	7
3	Post-observation conference	1. Analyze planning 2. Analyze preparation 3. Analyzing the results of observations 4. Give improvement suggestions 5. Determine follow-up	13

Broadly speaking, the steps for developing the instrument are as follows:

1. Synthesize theories that produce a concept of variables to be measured, then formulated the construct of these variables.

2. Based on this construct, the dimensions and indicators of each variable are developed.

3. In preparing the instrument lattice, it is contained in the form of a specification table consisting of dimensions, indicators, number of items and item options.

4. Compile and write test items (question) and non-test items (statements) along with answer options. These items can be in the form of negative sentences or in the form of positive sentences.

5. After the instrument has been prepared, it must go through a process of validation, both theory validation and empirical validation.

6. If the instrument still has shortcomings, the researcher makes improvements (revisions) and is reconciled back to the experts

7. After the revision, empirical and readability trials were conducted

8. Next calculate the reliability of the instrument. The range reliability coefficient (0-1) which is the magnitude of the consistency of the instrument's reliability value.

9. If the instrument still has shortcomings, the researcher makes improvements (revisions) and is re-communicated to the expert / expert / supervisor.

10. Re-assembly of valid and reliable instrument items to be made the final instrument [25].

Data Analysis Techniques by analyzing or describing measuring instruments to items as well as examining items in order to obtain information and can be used for various purposes. The function of item analysis is to find information about items that support the measurement reliability or not so that it can be maintained or replaced or revised, to find which items cause the reliability coefficient is low, as a process in the improvement of measuring instruments [26].

The method of testing the validity and reliability of the instrument is carried out in several stages, namely theoretical and empirical trials, at the theoretical trial stage, namely the construct validity testing phase involving experts who understand the instrument development procedures, experts in the field of clinical supervision, and linguists, expert judgment is very important in the development of instruments [27].

2.1. Validity

Validity testing at this stage uses a scale, to determine the validity of each question tested on a scale determined based on the median value of the experts. In carrying out instrument validation, it requires several experts or experts in the field so that the instruments

used during research can be said to be valid. According to Bajpai, S & Bajpai, R. (2013) Validity is a test of how well an instrument that is developed measures the particular concept it is intended to measure. In other words, validity concerned with the weather we measure the right concept or not [28]. Ideally, there should be a range of experts (also known as subject matter experts) on this panel at various professional levels. In content areas where it is difficult to find experts, the use of three experts is acceptable; normally, a panel of 5-10 experts is preferred. The use of> 10 experts is probably unnecessary [29].

If there are deficiencies in the instrument, the researcher must revise it. Then after revision, data from the results of expert assessments were analyzed with the help of Multi- Rater Facet Rasch (MRFR) because there was a rater in it, the research showed that assessors significantly varied in severity [30] through MFRM more effectively identifying the quality of work because many assessors were observing and powerful tool for handling politomous data in performance [31].

Rasch measurement using the help of facet software made by John Linacre. Expected value is the range is 0 to infinity. MnSq is said to have a suitable value between 0.5 and 1.5 productive for measurement, 1.5 to 2.0 is not produced but not destructive, values above 2.0 are distorted by some observers said to have no impact on the measurement (Zhu et al.., 2009), values less than 0.5 indicate slight variations, whereas fit MnSq values greater than 1.5 indicate more than variation [31].

2.2. Reliability

Reliability comes from the word rely, which means trust / reliable, while reliable, which means reliable. Reliability is related to accuracy and consistency. The results of a measurement that is able to produce data that has a high level of reliability is called a reliable measurement (reliable). According to Bajpai, S & Bajpai, R. (2013) If a measurement device or procedure consistently assigns the same score to individuals or objects with equal values, the instrument is considered reliable. In other words, the reliability of a measure indicates the extent to which it is without bias and hence insures consistent measurements of cross time and across the various items in the instruments as given [32]. Meanwhile, according to Obinne (2011) Reliability, that is a measure of the consistency of the application of an instrument to a particular population at aparticular [33].

3. Conclusion

Measurements or assessments that use non-test instruments to evaluate the affective aspect and motor skills. The form of assessment using the non-test apparatus/instrument is: The performance assessment, project/product assessment, and grading assessment, and attitude assessment. Assessment tools are classified as non-test techniques, among others: a) Questionnaire/poll, B) interview (interview), 3) matching list (check-list), 4) observation/observation, 5) assignment, 6) portfolios, 7) journals, 8) inventories, 9) Self- assessment, and 9) assessments by peer assessment.

In a test performance, it is usually done by telling participants to do something physical work (practice). This works form test is very suitable for conducting assessment in the practice/skills or practice lessons in the laboratory. The tools used to perform the assessment are generally a observation sheet (observation sheet). This form of action tests can generally be used to assess the process and outcome (product) of a practical activity.

Measuring is intended to give a quantitative form of an activity or capability that is in the form of numbers. The performance measurement used is the observation sheet. Performance measurement is used to match the suitability of knowledge about theory and skills in practice so that the evaluation results become clearer. The competency mastery assessment of skills or psychomotor that is owned by a person or student, there is only one form of the correct test that is action test (performance assessment). This means that people

who will be assessed their skills should display or perform their skills under the applicable employment requirements.

4. Discuss

Frequently encountered problems in the performance assessment. According to [34], there are three sources of error in the performance assessment, namely:

1. Scoring instrument flaws, the restriction manual is unclear so it is difficult to use by the assessment, generally because the components are difficult to observe (unobservable)

2. Procedural flaws, procedures used in the performance assessment are not good so it also affects the results of the Pensekoran.

3. Teachers personal-bias errors, penscorers tend to be difficult to eliminate personal bias problems, ie the possibility of a scoring has a problem generosity error, meaning rater tends to give a high-value, despite the fact that Actual results of a test participant's work is not good or vice versa. Another problem is the possibility of subjectivity of the penmanship so it is difficult for him to give objective value.

Applying scoring guidelines is a good way to provide an objective assessment. Performance assessment is a technique of data collection by systematically observing student behaviour about processes or products based on clear criteria, which serves as the basis for judgment. Performance observations need to be done in a variety of contexts to set the celebrity level of a particular ability.

To assess the ability to communicate where work for example, a diverse communication observation or observation is necessary, as: (1) external lines of communication are carried out openly, kindly, politely and sympathetic; (2) The language is used in a suitable intonation; (3) Body language is used naturally/nature is not feigned; (4) Sensitivity to cultural and social differences is shown; (5) Quality two-way communication is actively used. In this way, the participants ' ability to be more intact.

References

[1] Nneka, G. and Foundations, E. (2009). *Competencies Required by Primary School Supervisors: Basis for a Sustainable Universal Basic Education Programme*, pp, 393-406.

[2] Ozdemir, T. Y. and Yirci, R. (2015). A Situational Analysis of Educational Supervision in the Turkish Educational System. *Educational Process: International Journal*, vol. 4, issue 1, pp. 56–70, https://doi.org/10.12973/edupij.2015.412.5.

[3] Veloo, A., Komuji, M. M. A. and Khalid, R. (2013). The Effects of Clinical Supervision on the Teaching Performance of Secondary School Teachers. *Procedia - Social and Behavioral Sciences*, vol. 93, issue 2002, pp. 35–39, https://doi.org/10.1016/j.sbspro.2013.09.148.

[4] Situmorang, B. and Siburian, P. (2017). The Analysis of Headmaster of Sma's Supervision Competence Guidance By Supervisor in Aceh Tengah District. Vol. 7, issue 6, pp. 24–26, https://doi.org/10.9790/7388-0706022426.

[5] Dunaway, D. M., Bird, J. J. and Wang, C. (2014). Superintendents' Perceptions of the School Improvement Planning Process in the Southeastern USA. Vol. 42, issue 4, pp. 460–473, https://doi.org/10.1177/1741143213502186.

[6] Tyagi, R. S. (2009). Compare: A Journal of Comparative and International Education School – Based Instructional Supervision and the Effective Professional Development of Teachers. pp. 37–41, https://doi.org/10.1080/03057920902909485.

[7] Olibie, E. I. (2016). European Journal of Education Studies Curriculum Competences Manifested by Public And Private Secondary School Administrators For Clinical Supervision In Anambra State. p. *1111*.

[8] Okorji, P. N. and Ogbo, R. N. (2013). Effects of Modified Clinical Supervision on Teacher Instructional Performance. Vol. 4, issue 6, pp. 901–905.

[9] Taib, M. R. and Abdullah, Z. (2015). Clinical Supervision of Teaching Mara Junior Science College (MJSC), Malaysia, https://doi.org/10.1016/j.sbspro.2015.04.543.

[10] Burant, S. (2009).

[11] Sarfo, F. K. and Cudjoe, B. (2016). Supervisors' Knowledge and Use of Clinical Supervision to Promote Teacher Performance. Vol. 4, issue 1, pp. 87–100.

[12] Rahabav, P. (2016). The Effectiveness of Academic Supervision for Teachers Efektivitas Pengawasan Akademik untuk Guru. *Journal of Education and Practice Jurnal Pendidikan dan Praktik*, vol.7, issue 9.

[13] Chidi, N. and Victor, A. A. (2017). *Principals' Supervisory Techniques as Correlates of Teachers' Job Performance in Secondary Schools in Ebonyi State, Nigeria.*

[14] Bessong, F. E. and Ojong, F. (2008). Supervision as an Instrument of Teaching – Learning Effectiveness: Challenge for the Nigerian Practice, vol. 7, issue 1, pp. 15–20.

[15] Uludag, U., *et al*. (2016). Uludag University, vol. 41, issue 11.

[16] Fteiha, A. (2017). The Effectiveness of Clinical Supervision on Technology Teachers' Professional Development in Jerusalem a Case Study, vol. 7, pp. 11–23.

[17] Bulunuz, N., Goktalay, S. B. and Kesner, J. (2013). Clinical Supervision Model to Improve Supervisory Skills of Cooperating Teachers and University Supervisors during Teaching Practice Clinical Supervision Model to Improve Supervisory Skills of Cooperating Teachers and University Supervisors during Teaching.

[18] (Safari, 2018)

[19] Bank, D. and Covacevich, C. (2014). Instrument for Assessing how to Select an Instrument for Assessing Student Learning.

[20] Kimberlin, C. L. and Winterstein, A. G. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy*, vol. 65, issue 23, pp. 2276–2284, https://doi.org/10.2146/ajhp070364.

[21] Mardafi, D. (2008). *Teknik Penyusunan Instrumen Tes dan No Tes*. Yogyakarta: Mitra Cendikia Press.

[22] Grounlund, N. E. (1985). *Measurement and Evaluation in Resting (5th ed.)*. New York: Macmillan Publising Co, Inc.

[23] Kumaidi and Manfaat. (2016). *Pengantar Metode statistika Teori dan Penerapannya dalam Penelitian Bidang Pendidikan dan Psikolog*i. Cirebon JawaBarat: Eduvision.

[24] Crocker, L. and Algina, J. (1986). *Introduction to Classical and Modern Test Thoery*. Ohio: Holt, Rinehart, Winston, Inc.

[25] Djaali and Muljono, P. (2008). *Pengukuran Dalam Bidang Pendidikan*. Jakarta: PT. Grasindo.

[26] Naga, D. S. (2013). *Teori Sekor Pada Pengukuran Mental*. PT Nagarani Citrayasa.

[27] Zhu, W., Ennis, C. D. and Chen, A. (2009). Measurement in Physical Education and Exercise Science Many-Faceted Rasch Modeling Expert Judgment in Test Development, (August 2013), 37–41, https://doi.org/10.1207/s15327841mpee0201.

[28] Bajpai, S. and Bajpai, R. (2013). Goodness of Measurement: Reliability and Validity. International Journal of Medical Science and Public Health, vol 3, issue 2, **DOI: 10.5455/ijmsph.2013.191120133**.

[29] Gilbert, G. E. and Prion, S. (2016) Making Sense of Methods and Measurement: Lawshe's Content Validity Index. *Clinical Simulation in Nursing*, vol. 12, pp. 530-531. http://dx.doi.org/10.1016/j.ecns.2016.08.002.

[30] Myford, C. M. and Wolfe, E. W. (2004). Detecting and Measuring Rater Effects Using Many-Facet Rasch Measurement: Part I. *Journal of Applied Measurement*, vol. 5, issue 2, pp. 189–227.

[31] Basturk, R. (2008). Applying the Many-Facet Rasch Model to Evaluate Powerpoint Presentation Performance in Higher Education. *Assessment and Evaluation in Higher Education*, vol. 33, issue 4, pp. 431–444, https://doi.org/10.1080/02602930701562775.

[32] Till, H., Myford, C. and Dowell, J. (2013). Improving Student Selection Using Multiple Mini-Interviews with Multifaceted Rasch Modeling. *Academic Medicine*, vol. 88, issue 2, pp. 216–223, https://doi.org/10.1097/ACM.0b013e31827c0c5d.

[33] Obinne, A. D. E. (2011). A Psychometric Analysis of Two Major Examinations in Nigeria: Standard Error of Measurement. *Int J Edu Sci*, vol. 3, issue 2.

[34] Popham, J. W. (2011). Assessment Literacy Overlooked: A Teacher Educator's Confession. *The Teacher Educator*, vol. 46, issue 4, pp. 265-273, **DOI: 10.1080/08878730.2011.605048**.