Educational Quality Mapping Instrument (IP-SNP)

Rina Mutaqinah[1]
Bambang Sumintono[2]
Yanti Triatna[3]

[1]BBPMP Provinsi Jawa Barat
[2]Universitas Islam Internasional Indonesia (UIII)

https://doi.org/10.37517/978-1-74286-697-0-10

Abstract

Decision-making must be based on valid data, as well as decision-making in the field of education. It is very important to have a valid instrument, therefore it is necessary to develop an educational quality mapping instrument that can measure the achievement of the quality of education. This instrument was developed based on educational quality standards, namely the competency quality of graduates, content, process, assessment, educators and education personnel, infrastructure, management, and financing. The purpose of this study was to obtain an instrument that has high validity. The educational quality mapping instrument (IP-SNP) uses a rating scale, collecting data through a survey of principals in 19,460 elementary schools (SD) in West Java Province. Construct validity was carried out by 3 experts in educational measurement and management, several items were revised based on expert input. The data processing and analysis method in this study uses the IRT (Item Response Theory) method with the Rasch Model Winstep application version 3.7. The results of the Rasch Model analysis show that the SD level IP-SNP instrument has a Cronbach Alpha score of 0.90 which means that the reliability of persons and items is high, while the reliability of 0.93 means that the instrument is very good. Item validity of 0.97 means that the item can measure the achievement of the quality of education, strengthened by the separation of 2.85 items. This question has a good distribution of responses. Educational quality mapping instruments can be used.

Background

Policymaking requires data that is valid, easy to read and up to date. The data is used as a material consideration in every decision-making. If a policy is formulated without data, it is certain that the policy will not be able to become a problem solver.

The monitoring dashboard for RKPD inspection on data-based planning in all local governments in Indonesia shows that education development planning for the 2023 fiscal year is 93.22% not in accordance with the performance indicators for affairs (IKU) that have been determined by the Ministry of Education and Culture based on the Education report card (Bagren-Datin Kemdikbudristek, 2022). The discrepancy between education quality data and education planning in the regions was mostly caused by a lack of coordination and cooperation between the Education Development Planning section and the Education Report Card Analysis Team, besides that there were difficulties for the Analyst Team in translating data on education report cards (BBPMP Provinsi Jawa Barat, 2022).

Anticipating education report card instruments originating from data on the results of AN (National Assessment), Dapodik and others, which can only be processed and analyzed by the Central Government, an instrument is needed that makes it easier for the Regional Government and/or Education units to measure the achievement of program implementation results for internal purposes and as comparative data on educational reports. The instrument developed is an instrument based on national education standards PP Number 4 of 2022 concerning Amendments to PP Number 57 of 2021 concerning National Education Standards in 2021. The standards used are Management Standards; Educators and Education Personnel; and Financing. Instruments for other standards based on the latest PP SNP (Competencies of Graduates, Content, Process, Assessment, Infrastructure) have not been developed.

Education Quality Standards

There is no universally accepted definition of quality. Quality of Education is defined in the context of the education system as a whole (including schools and related bodies, teaching-learning environment, policies, etc.) and the quality of what the system offers to students/learners (i.e., quality of teaching and learning processes, curriculum etc.). Terms such as efficiency, effectiveness, equity, and quality are often used synonymously. To improve the quality of education it is necessary to improve the quality of its components. According to our definition, all educational space is the component itself: standards and curricula; academic literature; teaching staff (their professional skills); monitoring of education; moral and patriotic education; scientific research in the sphere of education (Kousainov, 2016)

According to Hoy et al. (2005) quality is often in terms of outcomes to match a customer's satisfaction. This gives rise to the definition of quality as the extent to which the outcomes meet the customer's requirements. Competency-based quality results need to be supported by the capability to deliver the service on the part of everyone involved. The inevitable outcome of this system is to reward those who can deliver a quality product, and train those who aren't skilled.

Goddard & Leask (1992) state the definition of quality from different perceptions, as meeting customer needs. They have included different customers for education—parents, governments, students, teachers, employers, and institutions—who seek different quality characteristics. Education is a service and not a product, its quality cannot lie exclusively in the final output. Its quality should also be manifested in the delivery process. Quality of education should also take into account determinants such as the provision of teachers, building, curriculum, equipment, textbooks and teaching processes (Grisay & Mahlck, 1991).

For them, the quality of education has a three-dimensional approach consisting of the quality of human resources and materials available for teaching (input), teaching practice (process), and results (outcomes). Furthermore, according to them, there are several indicators—

repetition, dropout, promotion, and transition rates—that planners frequently visit to arrive at approximate quality measures. UNICEF also strongly emphasized the desirable dimensions of quality, as identified in the Dakar Framework. Its paper 'Defining Quality in Education' recognizes five dimensions of quality: learners, environment, content, processes, and outcomes, founded on 'the rights of the whole child, and all children, to survival, protection, development, and participation (UNICEF, 2000). The Communiqué of the World Conference on Higher Education 2009 states that 'Quality criteria must reflect the aim of cultivating in students critical and independent thought and the capacity of learning throughout life. They should encourage innovation and diversity (UNESCO, 2009).

Mapping the quality of education is one of the education evaluation processes by measuring the achievement of national education standards (SNP) listed in the Government Regulation of the Republic of Indonesia Number 32 of 2013 concerning Amendments to Government Regulation Number 19 of 2005 concerning National Education Standards. The Education Quality Assurance Institute (LPMP) of West Java Province is the Ministry of Education's Technical Implementation Unit (UPT) located in West Java Province according to Permendikbud 28 of 2016 concerning SPMP for primary and secondary education, has the main task of carrying out quality assurance of education in West Java Province including mapping quality of education, facilitation of quality improvement, assistance to the regions and monitoring of educational evaluations.

The results of quality mapping will serve as the basis and reference in determining program planning in the regions or in educational units because mapping the quality of education or measuring the achievement of the eight SNPs is an initial part of improving the quality of education. Measuring the quality of education must of course be supported by measurement instruments that can represent all the indicators contained in the eight SNP.

So far, the Ministry of Education and Culture through the Directorate of Primary and Secondary Education has measured the achievement of the SNP through e-EDS by completing the PMP instrument every year. The PMP instrument, which accommodates all educational units in Indonesia, certainly experiences many problems both in substance and technically, so the results of this national quality mapping cannot be quickly accessed and utilized by stakeholders in the regions or in educational units. Therefore, it is necessary to develop a similar alternative quality mapping instrument that can be quickly accessed by regional governments and education units.

LPMP West Java has developed this alternative quality mapping instrument called the SNP Measurement Instrument (IP SNP), which is expected to meet regional and educational unit interests in West Java Province. The IP SNP was developed based on the key indicators contained in the eight SNPs. It is hoped that the results of measuring SNP achievements through IP SNP will produce accurate data in the form of profiles, quality maps that show the achievements of the strengths and weaknesses of educational units and regions in achieving education quality indicators in the eight SNPs, which are then reported to relevant stakeholders as a basis for further policy making. Measuring the achievement of educational quality can be done using various data collection techniques such as rating scales, questionnaires, observations and interviews.

Attitude measurement data could be processed and analyzed using a model of 1) classic, 2) modern or IRT (Item Response Theory). IRT had various models including a model of 1 parameter, 2 parameters and 3 parameters and a model which was similar to the 1 parameter namely Rasch models. The Rasch Model was developed by Georg Rasch in the 1950s (Naga, 2013)

The Rasch Model was one of the IRT (Item Response Theory) models. It was the general framework of a specific mathematical function that described the interaction between the person (persons) and items (item test). IRT was not dependent on the sample or respondents chose in a test (free item and free person free). Therefore, it caused this pattern to be more precise measurement and calibration on items was carried out (Sumintono & Widhiarso, Aplikasi Rasch Model untuk Ilmu-ilmu Sosial, 2014).

Research Method

The research method was carried out using the instrument development method from Robert K. Gable, with the following development steps: (1) developing a conceptual definition, (2) developing an operational definition, (3) choosing a scaling technique, (4) reviewing/justifying items, (5) selecting the response format, (6) compiling instructions for the response, (7) preparing a draft instrument and conducting an initial trial (8) preparing the final instrument, (9) collecting final data, (10) analyzing trial data using factor analysis techniques, item analysis and reliability, (11) revise the instrument, (12) conduct the final trial, (13) produce the instrument, (14) perform additional validity and reliability analysis, and (15) prepare a test manual (Gable, 1986)

Data analysis was performed using the Rasch model approach through the Winsteps program. In the Rasch model approach, in addition to paying attention to item items, it also pays attention to the respondent's aspects and calculates the magnitude of the correlation. The results of the analysis shown are statistical summary, item accuracy index, respondent accuracy index, scalogram, unidimensionality, respondent item map, and rating scale analysis.

Based on the stages of development that have been described, the first trial was conducted on 3 education management experts and instrument development experts consisting of measurement lecturers and education quality management lecturers. Expert experts review the constructs, dimensions, indicators and items of educational quality instruments. The instrument was then tested on … 30 teachers and school principals to test the legibility of the instrument, which was then given to 16 thousand elementary schools in West Java. Then analyze the validity of the instrument using the RASCH model with the Winstep type 3.1 application.

This research applied Rasch Model analysis to the instrument with of the 38 items using Winsteps version 3.73 software (Linacre, 2013) This was conducted to investigate overall respondents' agreement to the items based on each school, and regarding the item difficulty estimates in the instrument. Rasch modelling is built on conjoint measurement which is a formulation that stipulates the relationship between a person (e.g., respondent) and an item based on a mutual latent trait (Andrich, 1988; Bond & Fox, 2007)Rasch modelling produces a measurement scale with equal interval units called logit (logarithm odd unit) that shows the level of difficulty of each item and the level of each person agrees to the items (Alagumalai & Curtis, 2005; Sumintono & Widhiarso, 2014). Rasch measurement models then can inform overall trends of items and respondents based on the logit distribution of item and persons.

Several instrument analysis indexes from Winsteps are used to know the quality of the instrument and its item. Fit statistics from the Rasch model for example, such as outfit of mean-square and z-standard and point measure correlation indicate how good is the item really measured based on respondents' responses (Bond & Fox, 2007). Other important issues regarding measurement are unidimensionality and different item functioning (DIF) which can detect if is there anything wrong at instrument and item levels (Boone, Staver, &Yale, 2014). Another important piece of information to know the quality of instruments and items is the Wright map or item-person map which can illustrate a comprehensive pattern of response ((Bond & Fox, 2007); Boone et al., 2014).

Result and Discussion

The dimensions or aspects and indicators that have been developed are validated by experts in instrument development, management, and quality assurance; the quality aspects of graduates are obtained; contents/ content; learning process; learning assessment; educators and education personnel; infrastructure; management and financing of Education. From this aspect, it is further developed into indicators and instrument items. The results of the empirical trials show: Out of all the items/indicators in the 'IP SNP' instrument, three items/indicators from the three constructs must be repaired. This is because the items/indicators are not in accordance with the model (misfit), namely in the honesty construct (4th item/indicator) and one item/indicator in the integrity construct (2nd item/indicator); whereas the 19th item/indicator (honesty-1) gives the same response pattern as the other items/indicators in the honesty construct (20th item/indicator), so it also needs to be rearranged.

Table 1. Value of Item Suitability Criteria

Criteria	***Value***
Outfit Mean Square (MNSQ) accepted	0.5<MNSQ<1.5
Outfit Z-Standard (ZSTD) accepted	-2.0<ZSTD<+2.0
Point Measure Correlation (Pt Measure Corr)	0.4<Pt Measure Corr<0.85

Based on the results of the analysis using the Rasch model, item information and participant responses were obtained in the scale trial (person). In this study, data analysis was performed several times to obtain a number of items that met the item-model accuracy index. The stages of analysis are summarized in Table 2 below. In the first stage of analysis, 63 respondents were identified as outliers or not quite right with the model. According to Boone, Staver, & Yale (2014), the parameters used to determine the accuracy or suitability of respondents include: first, the received outfit mean square (MNSQ) value: 0.5 < MNSQ < 1.5. Second, the value of the Z-standard outfit (ZSTD) received: -2.0 < ZSTD < +2.0. Third, value of Point Measure Correlation 0.4<Pt Measure Corr<0.85. The following table presents a summary of the data analysis results.

Table 2 Summary of Final Analysis Results

Summary of Final Analysis Results

Output		*Results*
Item	Reliability Item	1,00
	Index Separation (H)	64,77
	Strata (H)	86,69
	Highest Logit Value	2,72 Logit
	Lowest Logit Value	-1,70 Logit
Respondents	Respondent Average Value	1,56 Logit
	Respondent Reliability	0,89
	Index Separation	2,85
	Index Separation (H)	4,13
	Highest Logit Value	6,49 Logit
	Lowest Logit Value	-5,33 Logit
Instrument	Alpha Cronbach	0,93
	Raw Variance Explained by Measures	44,6%
	Unexplained Variance in 1st Contrast	9,1%
	Unexplained Variance in 2nd Contrast	5,1%

The results showed that the construct validity was performed on a panel of experts who had knowledge of the attitude scale of instrument development and or the environment. The items considered not 'good' were revised, particularly on item number 7, 8, 15, 26, 30, 37, and 38. The items processing used the Rasch model to get the item information and valid respondents, in accordance with the characteristics and the paradigm of the Rasch model. The function of the Rasch model was to provide directions and detect a problem with the instrument.

According to Boone et al. (2014), the values of outfit mean-square, outfit z-standard, and point measure correlation are the criteria used to see the level of item suitability. If the item does not meet the criteria, it is better if the item is repaired or replaced. Guidelines for assessing item suitability criteria according to Boone, 2014.

Table 3. Reliability Test Processing Results

Alpha Cronbach	Interpretation	Item Reliability	Interpretation	Person Reliability	Interpretation	Conclusion
0,93	Very Good	1.00	Excellent	0.89	Very Good	Reliable

Table 4. Validity Test Processing Results

Raw Variance Explained by Measure	Interpretation	Unexplained Variance 1st Contrast Eigenvalue	Observe	Interpretation
44.6%	Good	8.0	9.1%	Excellent

The item reliability value of 1.00 indicates that the quality of the items in this instrument is very high. In other words, the 47 items identified as having accuracy with the model are indeed quality items. Furthermore, the respondent's reliability value of 0.89 indicates that the consistency of the respondents' answers is high.

In other words, respondents answered all items seriously (not carelessly).

Unidimensionality Item to answer the question of whether the developed instrument is able to measure what it should measure. In the context of this study, the educational quality mapping instrument, from Table 4 it is known that the raw variance data measurement results are 44.6%. According to Sumintono and Widhiarso (2014), the minimum requirement for unidimensionality is 20%, and if the value is more than 40%, then even better, and the variance that cannot be explained by the instrument should ideally not exceed 15%. Based on this explanation, the raw data variance of 44.6% indicates that the minimum unidimensionality requirement of 20% can be met and is even classified as good because it is more than 40%. The results of the analysis of variance that cannot be explained by the instrument of 13.5% also meet the criteria, namely, not exceeding 15%.

Based on Table 4, the results of the raw variance values explained by measures indicate that the test items for mapping the quality of education are in the "good" category. Furthermore, based on the values observed in the unexplained variance 1 contrast, it shows that there is no trend of the discrepancy between the items so that they can be used, but the eigenvalues greater than 3 indicate that there are problematic items so that further analysis can be done with item fit order analysis to determine whether an item can be maintained or must be replaced.

Item fit referred to as item suitability can explain whether the items function normally to take measurements.

Outfit means-square, outfit z-standard, and point measure correlation are the criteria used to see the level of item fit (Item fit) (Boone, 2014). The criteria used to check the suitability of the items can be seen in Table 1.

Conclusions and Policy Proposal or Options

The results of the alpha reliability coefficient of 0.91 indicate that the self-efficacy scale in career decision-making has a high-reliability coefficient. That is, this scale produces a measurement score that is consistent and reliable. The reliability coefficients of the items and the respondents are also quite good, namely 0.91 and 0.91. This shows that these twenty items are quality items, and the group of respondents answered them seriously. These two results further strengthen and confirm that the self-efficacy scale in career decision-making is indeed a quality measurement tool because not only are the measurement results reliable but also the twenty items are quality items.

Instrument development begins with developing a blueprint and items, then validating the construct to the experts. Revise blueprints and items according to input from experts, then the results were analyzed using the Rasch Model. From 86 items, it was eliminated to 47 items. Item reliability in the very good category, Reliable. Item validity in the good category. based on the values observed in the unexplained variance 1 contrast, it shows that there is no trend of the discrepancy between the items so that they can be used, but the eigenvalues greater than 3 indicate that there are problematic items.

Based on the results of this study, the items for this quality mapping instrument can be used by all schools with the standard requirements and the dimensions measured are the same as this instrument.

References

Alagumalai, S., & Curtis, D. &. (2005). Applied Rasch Measurement: Book of exemplar papers in honour of John P Keeves. Springer.

Andrich, D. (1988). *Rasch models for measurement.* Sage Publications.

Bagren-Datin Kemdikbudristek. (2022, September 27). *Dashboard PBD* . Retrieved from Dashboar Monitoring Pemeriksaan RKPD: https://metabase.pauddasmen.id/dashboard/5374-iku-kabupaten-kota

BBPMP Provinsi Jawa Barat. (2022). Laporan Pendampingan Penyusunan Policy Brief dan Penyesuaian IKU Pemda Kab. Kota di Provinsi Jawa Barat. BBPMP Provinsi Jawa Barat.

Bond, T. G., & Fox, C. M. (2007). Applying the Rasch model: Fundamental measurement in the human sciences (2nd ed.). Lawrence Erlbaum Associates Publishers.

Boone, W. S. (2014). Rasch Analysis in the Human Sciences. Springer.

Gable, R. (1986). Instrument Development in the Affective Domain. Kluwer Nijhoff.

Goddard, D., & Leask, M. (1992). The search for quality: Planning for improvement and managing change. Paul Chapman Publishers.

Grisay, A., & Mahlck, L. (1991). A Preview of some research studies and policy documents: The quality of education in developing countries. IIEP.

Hoy, C., Baryne-Jardine, C., & Wood, M. (2005). *Improving Quality of Education.* Falmer Press.

Kousainov, A. K. (2016). The ways of improving the quality of the secondary education in the Republic of Kazakhstan. *EEIA-2016* (p. SHS Web of Conferences). Astana, Kazakhstan: EDP Sciences. https://doi.org/10.1051/shsconf/20162901036

Linacre, J. M. (2013). A user's guide to Facets: Rasch measurement computer program [Computer program manual]. MESA Press.

Naga, D. S. (2013). *Teori Sekor pada Pengukuran Mental.* PT Nagarani Citarasa.

Sumintono, B., & Widhiarso, W. (2013). *Aplikasi Rasch Model untuk Penelitian Ilmu Sosial.* Tim Komunika Publishing House.

Sumintono, B., & Widiarso, W. (2014). *Aplikasi Rasch Model untuk Ilmu-ilmu Sosial.* Trim Komunikata Publishing House.

UNICEF Programme Division, United Nations Children’s Fund (UNICEF). (2000). *Defining quality in education. A paper presented by UNICEF at the meeting of the international working Education Section.* Programme Division, United Nations Children’s Fund (UNICEF).