The Characteristics of the Nature of Science (NOS)-based Instruments in Newton Law Using Rasch Model Analysis

Pratiwi Restu Murti*)
Universitas Sebelas Maret, Surakarta, Indonesia.
E-mail: restumurti94@gmail.com

Nonoh Siti Aminah
Universitas Sebelas Maret, Surakarta, Indonesia.
E-mail: nonoh_nst@yahoo.com

Harjana
Universitas Sebelas Maret, Surakarta, Indonesia.
E-mail: harjana59@gmail.com

*) Corresponding Author

Abstract: The objective of this research is to find out the characteristic of each test item with NOS-based instruments which is analysed based on validity and reliability level using Rasch model (RM). A test using NOS-based instruments was used to collect the data. This research is categorized as descriptive with quantitative using statistic from RM through QUEST program. There are 104 students participated as the subject of this research. The result, which is based on validity of each test item using RM analysis, shows that 25 test items are considered fit or accepted. Based on the estimate of item reliability, NOS-based instrument has reliability coefficiency at 0.96. Based on difficulty items 2 test items declared not good, namely test item number 19 and item test number 4. Based on the analysis according to Classic Test Theory of RM, the NOS instrument can be used to measure senior high school students' science literacy on Newton Law.

How to Cite: Murti, P., Aminah, N., & Harjana, H. (2020). The Characteristics of the Nature of Science (NOS)-based Instruments in Newton Law Using Rasch Model Analysis. *Al-Ta Lim Journal, 26*(3). doi:http://dx.doi.org/10.15548/jt.v26i3.493

INTRODUCTION

Nature of Science (NOS) is a prominent subject in science education as seen in several research used to prove the importance of NOS in science development (Holbrook & Rannikmae, 2009; Laugksch, 2000; Roberts, 2013; Wenning, 2006). NOS is considered to be an important element as several aspects play a prominent role in science development (Neumann et al., 2011a; Taber, 2018; Wilkin & Castleman, 2003). Science literacy is a multidimensional skill which includes knowledge (vocabulary, fact, and concept), processing skill (skilled and intellectual), disposition (behavior & attitude), well connection between science-technology-people nearby, and students' possession on science history and fact (Lehrer & Schauble, 2007).

There are more than 25 NOS instruments developed in the last 50 years. The actions taken are mostly centered on designing, developing, and evaluating NOS in different sample and population(Al-Bouti, 2018; Choi & Lee, 2003; Faikhamta, 2013; Khery et al., 2019; Lee, 2013). There is already an open respond NOS instrument development to measure high school students' literacy in science such as VNOS series (Holbrook & Rannikmae, 2009) and Nature of Science Literacy Test (NOSLiT) instrument. However, there is no NOS instrument that can be used to measure students' science literacy in certain subjects. Based on how it works,

science literacy indicates that students must be familiar with "the most basic principal of science" in basic physics law such as Newton Law about power and actions, Thermodynamics Law about energy and entropy, the equal relation between electricity and magnetism, and material atom structure (Gess-Newsome, 2002; Wilujeng & Suryadarma, 2017) .

Researcher develops a NOS-based instrument to measure high school students' literacy on Newton Law. NOS-based instrument consists of 25 multiple choices with 4 alternative answers. It adopts a framework from NOSLiT instrument modified on Newton Law (Rosana, 2018; Temel et al., 2017). Multiple-choice tests are more widely used than other test forms. This is because multiple-choice tests have advantages, including: (1) tested materials can cover most of the learning materials, (2) student answers can be corrected easily and quickly, (3) the answer to each question is definitely true or false , so that the objective assessment [6]. An assessment can not be relied upon if it contains too many items where the proportion of students can not answer correctly (Van De Watering & Van Der Rijt, 2006). The purpose of this study was to determine the characteristics of NOS-based instruments used to measure the science literacy of high school students in Newton's Law. The characteristics of the instrument include the validity of test items, the reliability of test items, and the degree of difficulty of the test items. In order to find out whether an instrument is considered as valid and reliable is by validating the instrument through several analysis techniques such as Rasch Model (RM). Several researchers also use RM to find out the validation and reliability from the implemented instrument.

RM is a fundamental measurement that is often used to develop and validate any instruments. Two basic assumptions for applying Rasch Measurement Theory are local independence and dimension. The dimensions of the tests associated with classifying items are classified as Content-Based Statistics Dimensions and Dimensions.

Item response theory (IRT) is general statistics about items (question) and scale (questionnaire) on performance and how it is related with the factors used for measuring items on scale. Rasch model one or best known as one parameter (1-PL) is the simplest model of logistic. This is because such model only has one parameter item to influence performance of subject. Thus, such model recognizes that all items in test have the same judgment power. In classical test theory it is assumed that the scores observed on the assessment are the sum of the components of the true score and the component of measurement error. RM uses probability data to answer questions for each individual but placed on each level of item difficulty. In this model, each individual and item is on separated location. RM assumes that a probability of certain individual will have a specific way of responding a certain item. This is considered as logistic function from relative distance between item location and individual. Rasch analysis brings several benefits such as; a) readable and understandable result, b) parameter estimation for each individual, c) comparison between individuals are very independent towards the instruments, d) comparison between items are independent from individual samples. Based on the explanation above, there is a need to conduct a research to find out the quality and characteristics of NOS instrument which is analyzed based on validity level of each test item and the reliability of item test using Rasch Model (RM). The analysis use QUEST program so the NOS-instrument is considered valid and reliable to measure high school students' science literacy, especially on Newton Law.

METHOD

This research was conducted in SMA Negeri 1 Karanganyar (high rank), SMA Negeri 2 Karanganyar (intermediate rank), and MA Negeri Karanganyar (low rank). Subjects were determined by analyzing the results of physics test in national exam for three consecutive years (Istiyono et al., 2014). It is conducted by compiling data from high schools in Karanganyar using PAMER UN

application. Trials are conducted on the basis of the school's rank (low, medium, and high) according to the value of UN Physics. Based on the data collected by the researcher, three schools are then selected to fulfill the high, intermediate, and low rank. There are 104 students as the samples in this research which consist of students from class X selected from their own class from each school. Experts state that the sample for RM analysis must consist of 30 to 300 individuals.

The data was collected by conducting a test using NOS-based instrument. This is a descriptive research with quantitative approach. The data was analyzed using RM statistic through QUEST program. In RM, the characteristic of the items is only indicated by the statistics of items' level of difficulty while the quality of the instrument is indicated by the validity and reliability of the test items. Any items are considered as fit using RM in QUEST program based on the value of INFT Mean of Square (INFT MNSQ) as well as its standard deviation or the value of INFT Mean of every test's INFT t. An item is declared fit

or compatible with the RM model if it has a MNSQ INFIT range range from 0.77 to 1.30 . The analysis was conducted using error limitation at 5% so the value of INFIT t will be ±1.96 or rounded to ±2,0. An item is considered as not fit if the value is in the range of <-2.0 or> +2.0 according to RM and it has to be omitted. Item is said to be good if the index of difficulty is more than -2.0 or less than 2.0.

RESULTS AND DISSCUSSION

The NOS-based instrument consists of 25 multiple-choice test items with 4 alternative answer options, adopting the framework of the modified NOSLiT instrument on Newton's Law. The NOS-based instrument is used to measure the science literacy of high school students on Newton's Law. Table 1 shows the differences between NOSLiT and NOS instruments developed by researchers. NOSLit does not contain general science knowledge, whereas NOS instruments contain only Newton's existing laws on Physics subjects.

Table. 1 Example of developing NOS instruments on Newton's Law materials

No	NOSLiT	Test Based on Nature of Science (NOS) in Newton Law
1	A teacher asks students, "What do you think will happen next?" The teacher is asking for a(n): a. Hypothesis b. Explanation c. Principle d. prediction	An object is placed on a piece of paper, then a teacher asks his students "What happens if the paper is pulled quickly and slowly?". The sentence indicates that the teacher is asking about. . a. Hypothesis b. Explanation c. Principle d. Assumption
2	The relationship between density, volume, and mass can be stated as follows: density = mass/volume Which of the following is a proper conclusion based on this relationship? a. if the mass of an object increases, its density will increase regardless of volum b. if the volume of an object increases, its density will also increase c. if more matter is packed more tightly into a fixed volume, the density of that matter will increase d. if more matter is packed more tightly into a fixed volume, the density of that matter will decrease	How do you think the conceptual relationship between force, mass, and acceleration? a. If the mass increases, the force increases by ignoring the acceleration. b. If acceleration increases, the force decreases. c. The larger the mass of the object will cause the acceleration to decrease, so the more difficult it is to change its state to move. d. The greater the mass of the object will cause acceleration to increase, so the object more easily change the situation to move.

No	NOSLiT	Test Based on Nature of Science (NOS) in Newton Law

| 3 | A lunatic runs through the street screaming repeatedly, "The moon is made of Swiss cheese." Is such a statement scientific?
a. Yes, even though the statement is wrong.
b. Yes, because the moon is white and has holes.
c. No, because the statement is wrong | When 2 children push the table in an unknown style of magnitude but the opposite direction, it turns out the table does not move a bit. One of the children said "this table is lazy to move". Is such a statement scientific?
a. Yes, because the statement is true
b. Yes, even if the statement is false |
| 4 | Billy thinks that winter is caused by geese flying south during the autumn. He also thinks that summer is caused by geese flying north during the spring. He claims, "If one event comes before another, the first event causes the second event. It's always this way." What, if anything, is wrong with the claim that if one
event follows the other, the first causes the second?
a. Nothing, this claim of cause and effect is perfectly correct.
b. Cause has nothing to do with effect according to most scientist; some things just randomly occur.
c. While effect must follow cause, it is important that the connection between the two be explained.
d. Cause does not always have an effect in the everyday world as scientists see it. | A student thinks that the eagle that flies up (condition 2) is caused because the bird moves its wings down (condition 1). Is there a cause-and-effect relationship of what the student is thinking?
a. Nothing, the statement is not a cause of effect, but the action of reaction because it occurs on two different objects of birds and air
b. There, the statement of cause and effect is very true, and not including reaction action because it occurs on only one thing ie birds.
c. Nothing, the cause has nothing to do with the result. because according to most scientists, some events just happen randomly.
d. There, due to having to follow the cause. It is important that the relationship between the two can be explained. |
| 5 | A well-known and highly respected scientist claims to have accurate knowledge of future events given to him by space aliens, and has predicted certain events in the not-too-distant future. How should other scientists respond to these predictions?
a. accept them because the scientist is well-known and highly respected
b. reject them, being certain to tell the general public that this man is a fraud
c. caution the public and wait to see if predictions by the scientist turn out to be true
d. entirely ignore the man and his predictions | A well-known and highly respected scientist predicts "if the reaction action force takes place on a single object, there will certainly never be accelerated motion because the total force on each object is zero." How should scientists respond to these predictions?
a. They accept, because the prediction is disanpaikan by scientists who are famous and highly respected so that the prediction is true.
b. They refused, because the statement was wrong.
c. They are waiting for a proof to see if the scientist's predictions are true or false.
d. They conclude themselves according to their own knowledge of the predictions of the scientist. |

Researcher checks the functional item using RM analysis to find out the quality of NOS-based instrument test. It is analyzed based on the level of validity and reliability. Any items are considered as fit using RM in QUEST program based on the value of INFT Mean of Square (INFT MNSQ) as well as its standard deviation or the value of INFT Mean of every test's INFT t. The same is stated by that testing the goodness of fit for the overall test and the overall test / case / person) based on the mean value of INFIT Mean of Square (Mean INFITMNSQ) and its standard deviation or observing the average value of INFIT t (Mean INFIT t) and its standard deviation. Quest program rules that an item test or person is considered as fit if the INFT/ OUTFT MNSQ is in range of 0.77 to 1.30.

Picture 1 shows that 25 items are considered as fit with acceptance limit on ≥ 0,77 to ≤ 1,30. The analysis was conducted using error limitation at 5% so the value of INFIT t will be ±1,96 or rounded to ±2,0. An item is considered as not fit if the value is in the range of <-2.0 or> +2.0 according to RM and it has to be omitted. Based on item validity test using RM analysis with INFT t and OUTFT t limit, the result is 25 item tests are considered fit or acceptable because the value of INFT t and OUTFT t is in the range of -2 until +2. Based on INFIT t and OUTFT t limit, all 25 test items are eligible to be used and there is no omission. Specifically, the NOS instrument can be 100% valid without any test items being eliminated based on two different limits ie INFT / OUTFT t and INFT /OUTFT MNSQ. However, if outfit and INFT

MNSQ are accepted, the INFT/OUTFT t index can be ignored.

Fit items show how far the consistency in using the items by how sample responds to other items. If the value of INFT/OUTFT MNSQ is more than 1,30, the test is confusing. If MNSQ value is lower than 0.77, it is too easy for respondents. Data from Picture 1 shows that all 25 test items have INFT/ OUTFT MNSQ value in the range of 0.77 to 1.30. It can be stated all of the NOS-based test instrument are not confusing or not too easy for respondents.
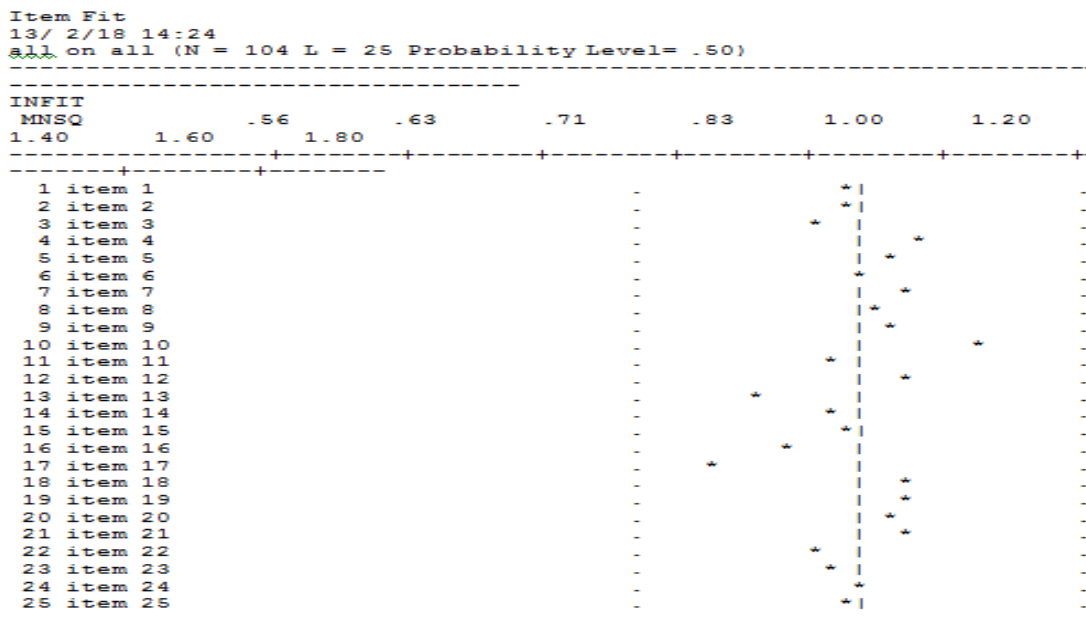
Picture 1. Item fit

Difficulty Items

RM analysis can identify the misunderstanding between item and respondent. For example, a very bright student must have answered the questions easily. This method can identify the difficulty level of items and the respondents' ability. Picture 3 shows distribution sample on the left and distribution item on the right. Sample with the same position with the item is 50% likely to answer questions correctly. For example, item number 8 is answered by one person with 50 % chance and number 18 is answered by 18 persons with the same percentage of chance. Sample with a higher position over the item has bigger chance to answer correctly because the item is usually too easy for them. Item test with the similar level of difficulty is in the same place on logit scale. In this test, it is on number 9 with 15 and number 2 with 11.
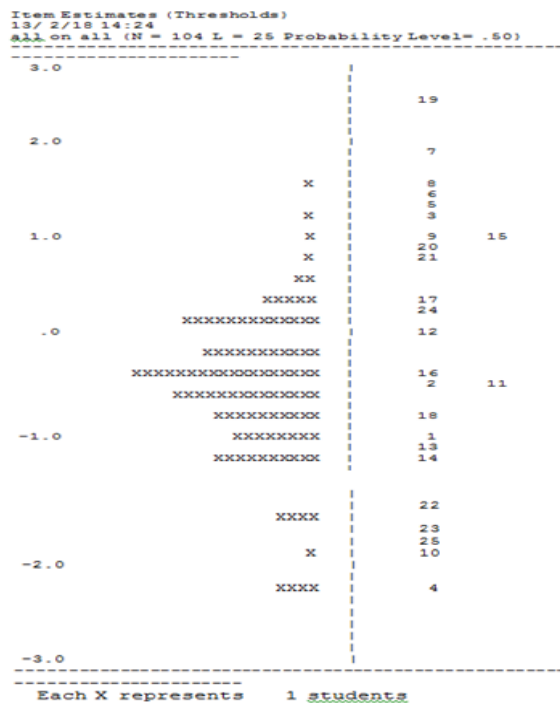

Picture 2. Item estimates

Figure 2 shows that the test item number 19 lies at the top, which means the test item number 19 is the most difficult test item.

Figure 3 shows the number of testee data that can answer correctly from each test item, stating that the test item number 19 is a test item with the correct answerer at least, that is, only 5 testees have a chance to answer correctly. The test item number 4 is located at the bottom shown in Figure 2. It states that test item number 4 is the easiest test item. A total of 87 testees have the opportunity to answer correctly as shown in Figure 3. RM analysis can identify misunderstandings of items and respondents. For example, a very smart student should be able to answer questions easily. This method can identify the difficulty level of items and the ability of the respondents. Sources of error are also factors that affect the performance of the testee, such as emotional motivation and tension, and errors due to accidental elements of certain test items such as guessing. Assessment difficulties, or some items in the assessment, may degrade assessment reliability in two ways. First, if assessments are more difficult than students expect, this can lead to confusion, decreased motivation, loss of concentration, uncertainty, anxiety, etc. and as a consequence, this means more mistakes. Second, especially in multiple-choice assessment format, there is a possibility of guessing. If the item is more difficult, this means more students will guess and this adds a random error to the variance of the scoring score.
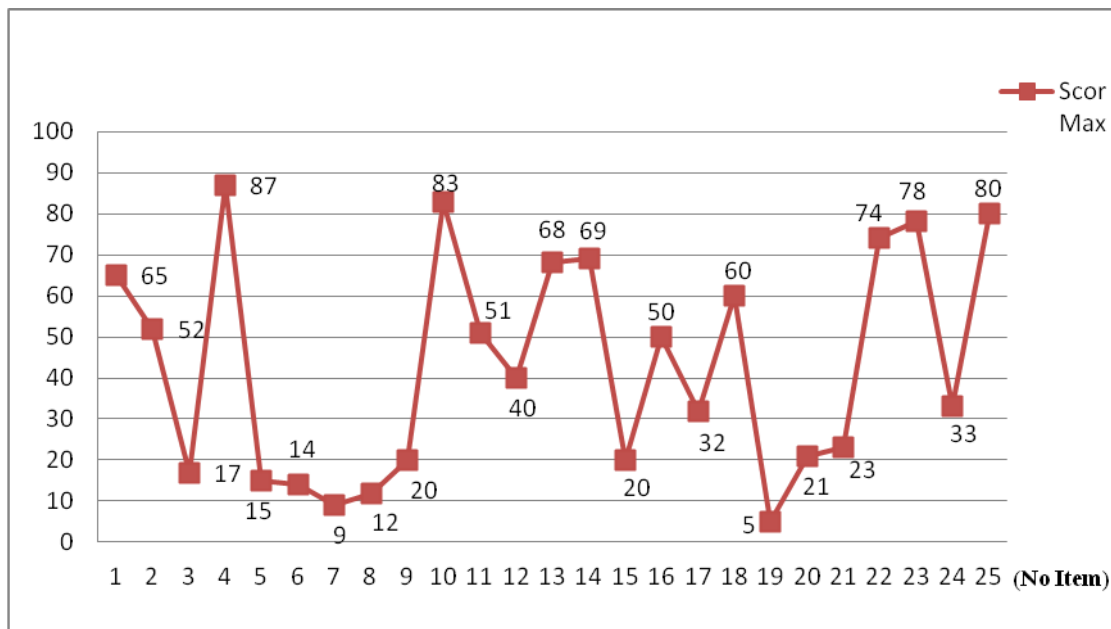
Picture 3. Maximum score of each item

Based on the analysis, difficulty items (threshold value) lie between -2.23 to 2.59. Item is said to be good if the index of difficulty is more than -2.0 or less than 2.0 (-2.0 <b <2.0). Based on difficulty items, 2 test items are not good test item 19 and test item number 4. Item test number 19 with a threshold value of -2.23 test item is declared too difficult, while item test number 4 with a value of 2.59 is declared too easy. So there are 23 test items that are either based on difficulty items. In his research conducted anlisis with limit of the same difficulty items (-2.0 <b <2.0) obtained as many as 44 items test otherwise good.

Reliability of Estimates

Picture 4. Reliability of item estimates

Picture 2 shows the estimate reliability of NOS-based instrument with 0.96 coefficient reliability. Similar research was conducted (Neumann et al., 2011) by analyzing NOS-instrument using technique developed by Lombrozo. The result shows that Cronbach α (reliable co-efficiency) is at 0.81, this value indicates that test items are reliable enough. If an instrument is analyzed with RM approach and use WINSTEPS application, the reliability value is at 0.93, this value can be chategorized as very good chategory. Based on the analysis results obtained the reliability of the instrument set (PhysTHOTS) of 0.95, this reliability value belongs to high category. The result of this research is taken from Cronbach α (reliable co-efficiency) at 0.96. Thus, the value shows that the instruments are in good and effective condition with high level of consistency it can be used in the real research.

Reliability and Separation Items and Respondents based on the RM approach, which Cronbach's Alpha α can accept is between 0.71- 0.99 where it is at the best level Based on the description it can be said that the NOS instrument developed by researchers with a reliability value of 0.96 has a high level of reliability. The NOS instrument shows excellent condition and effectiveness for measuring the science literacy of high school students in Newton Law material.

CONCLUSION AND RECOMMENDATION

Based on examination of validity and reliability of NOS instruments using RM through Quest program, it can be concluded that the NOS instrument developed by the researcher is stated fit or received 100% without any test items thrown away. The NOS instrument developed by researchers has a high degree of reliability, which shows excellent conditions and effectiveness for measuring the science literacy of high school students in Newton Law material. Based on difficulty items, 2 test items are not good test item 19 and test item number 4. This research can be used as a reference for the development of instruments used to measure the science literacy of high school students.

Researcher would like to express her gratitude towards Carl J. Wenning for his Nature of Science Literacy Test (NOSLiT) instrument as a reference in the development of a Test Based on Nature of Science (NOS) in Newton Law.

REFERENCES

Al-Bouti, E. (2018). *The UAE High School Students' Conceptions of the Nature of Science (NOS)*.

Choi, S.-H., & Lee, E.-A. (2003). Scientifically Gifted Students' Conceptions of Nature of Science. *Journal of the Korean Earth Science Society*, *24*(2), 100–107.

Faikhamta, C. (2013). The development of in-service science teachers' understandings of and orientations to teaching the nature of science within a PCK-based NOS course. *Research in Science Education*, *43*(2), 847–869.

Gess-Newsome, J. (2002). The use and impact of explicit instruction about the nature of science and science inquiry in an elementary science methods course. *Science & Education*, *11*(1), 55–67.

Holbrook, J., & Rannikmae, M. (2009a). The meaning of scientific literacy. *International Journal of Environmental and Science Education*, *4*(3), 275–288.

Holbrook, J., & Rannikmae, M. (2009b). The meaning of scientific literacy. *International Journal of Environmental and Science Education*, *4*(3), 275–288. https://doi.org/10.1080/095006906010 07549

Istiyono, E., Mardapi, D., & Suparno, S. (2014). Pengembangan tes kemampuan berpikir tingkat tinggi

fisika (pysthots) peserta didik SMA. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *18*(1), 1–12.

Khery, Y., Nufida, B. A., Suryati, S., Rahayu, S., & Budiasih, E. (2019). Mobile Learning with Oriented Nature of Science (NOS): Does undergraduate school need it? *3rd Asian Education Symposium (AES 2018)*.

Laugksch, R. C. (2000). Scientific literacy: A conceptual overview. *Science Education*, *84*(1), 71–94.

Lee, Y. H. (2013). A proposal of inclusive framework of the nature of science (NOS) based on the 4 themes of scientific literacy for K-12 school science. *Journal of the Korean Association for Science Education*, *33*(3), 553–568.

Lehrer, R., & Schauble, L. (2007). Scientific thinking and science literacy. *Handbook of Child Psychology*, *4*.

Neumann, I., Neumann, K., & Nehm, R. (2011a). Evaluating instrument quality in science education: Rasch-based analyses of a nature of science test. *International Journal of Science Education*, *33*(10), 1373–1405.

Neumann, I., Neumann, K., & Nehm, R. (2011b). Evaluating instrument quality in science education: Rasch-based analyses of a nature of science test. *International Journal of Science Education*, *33*(10), 1373–1405. https://doi.org/10.1080/09500693.2010.511297

Roberts, D. A. (2013). Scientific literacy/science literacy. In *Handbook of research on science education* (pp. 743–794). Routledge.

Rosana, D. (2018). Desemination of Authentic Assessment in Local Content-Based Sciences Learning to Achieve The Learning Outcomes Based on Nature of Science. *Journal of Physics: Conference Series*, *1097*(1), 012034.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), 1273–1296.

Temel, S., Şen, Ş., & Özcan, Ö. (2017). Validity And Reliability Analyses For The Nature Of Science Instrument Secondary (NOSI-S). *Journal of Baltic Science Education*, *16*(3).

van de Watering, G., & van der Rijt, J. (2006). Teachers' and students' perceptions of assessments: A review and a study into the ability and accuracy of estimating the difficulty levels of assessment items. *Educational Research Review*, *1*(2), 133–147.

Wenning, C. J. (2006). Assessing nature-of-science literacy as one component of scientific literacy. *Journal of Physics Teacher Education Online*, *3*(4), 3–14.

Wilkin, C., & Castleman, T. (2003). Development of an instrument to evaluate the quality of delivered systems. *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of The*, 10–pp.

Wilujeng, I., & Suryadarma, I. G. P. (2017). Science learning based on local potential: Overview of the nature of science (NoS) achieved. *AIP Conference Proceedings*, *1868*(1), 080005.