Anatolian Journal of Education
e-ISSN: 2547-9652

October 2025 ● Vol.10, No.2
www.e-aje.net
pp. 71-84

Development and Validation of a Self-Reported Assessment Literacy Instrument for Teachers

Jetro Gardon
De La Salle University, Philippines, *jetro_gardon_a@dlsu.edu.ph*

Maricar Prudente
De La Salle University, Philippines, *maricar.prudente@dlsu.edu.ph*

Auxencia Limjap
De La Salle University, Philippines, *auxencia.limjap@dlsu.edu.ph*

Assessment literacy is currently gaining attention because of its relevance to teachers' instructional practices and students' performance. Previous studies showed that Filipino teachers have low to mid-level assessment literacy. Therefore, it is crucial to determine how teachers view their assessment literacy. This study aimed to develop and determine the validity and reliability of the Self-Reported Assessment Literacy Instrument. Five experts on science education validated the instrument. After revising it based on the validators' comments and suggestions, it was pilot-tested on 42 science teachers. After a series of minor revisions based on the pilot results, the instrument was administered to 107 teachers. The exploratory factor analysis of the results revealed that it almost followed the three-dimensional model of assessment literacy. The findings also showed that the instrument is reliable and valid for surveying self-perceived assessment literacy. Therefore, it is recommended for use in measuring self-reported assessment literacy.

Keywords: assessment, assessment literacy, assessment practices, conceptual knowledge dimension, praxeological dimension, socio-emotional dimension

INTRODUCTION

The teachers' assessment literacy construct has attracted considerable attention among educators, administrators, and researchers (Gotch & McLean, 2019). This is due to the perceived benefits of good assessment practices and the importance of assessment results in education reforms. Given the relevance of assessment in instruction, teachers' assessment literacy has been given importance. A good grasp of assessment helps teachers get information about students' learning, leading to an effective teaching strategy that responds to the student's learning needs (Pastore & Andrade, 2019). Various studies showed Filipino teachers' low to mid-level assessment literacy (Napanoy & Peckley, 2020; Clores & Reganit, 2019; Hailaya, 2014). Based on their search, the researchers also noticed that no instrument measures teachers' self-reported assessment literacy based on its three dimensions formulated by Pastore and Andrade (2019). These findings make it interesting how teachers view their assessment literacy. This study aimed to develop and validate a self-reported assessment literacy instrument to help researchers describe how teachers view their assessment literacy.

Assessment literacy is an interrelated set of knowledge, skills, and dispositions a teacher uses in preparing and implementing assessment in teaching (Pastore & Andrade, 2019). It includes the knowledge of what's, why's, and how's of assessment (Stiggins, 1995; Ogan-Bekiroglu & Suzuk, 2014). Pastore and Andrade (2019) proposed three interrelated dimensions for assessment literacy from a socio-constructivist point of view:

Citation: Gardon, J., Prudente, M., & Limjap, A. (2025). Development and validation of a self-reported assessment literacy instrument for teachers. *Anatolian Journal of Education*, *10*(2), 71-84. https://doi.org/10.29333/aje.2025.1026a

a. Conceptual knowledge – refers to the teachers' knowledge of what assessment is and how assessment is designed and executed.
b. Socio-emotional – refers to the aspects of assessment as a social practice, especially in the classroom context.
c. Praxeological – refers to the actual practice of assessment. and the translation of assessment knowledge to assessment practices realized in the classroom.

Using the Assessment Literacy Inventory (ALI) developed by Mertler and Campbell (2005), various studies showed that participating teachers have low assessment literacy (Prizovskaya, 2017; Muhammad & Bardacki, 2019). Similar results were produced in the study of Napanoy and Peckley (2020) on in-service public elementary school teachers in Northern Philippines and Hailaya (2014) on elementary and secondary teachers in Tawi-Tawi, Philippines. These findings are alarming since low assessment literacy might result in inaccurate assessment practices, hindering students from reaching their full potential (Stiggins, 1995). Mellati and Khademi (2018) observed this when they saw differences in the assessment practices of teachers with high and low-level assessment literacy.

While inventories and other questionnaires on assessment literacy exist, the researchers could not find an instrument that measures this construct based on the teachers' perceptions. Moreover, the researchers needed to develop an assessment literacy instrument that targets its three dimensions based on the model developed by Pastore and Andrade (2019). To help establish baseline information on how teachers perceive their assessment literacy, the researchers sought to develop and validate the *Self-Reported Assessment Literacy Instrument* (SRALI). It was developed based on the three-dimensional model of assessment literacy proposed by Pastore and Andrade (2019) and the definition of assessment provided by the Philippine Department of Education (DepEd). Specifically, the study aimed to answer the following questions:

1. How was the *Self-Reported Assessment Literacy Instrument* developed?
2. What is the instrument's factor structure based on the data?
3. How reliable is the developed instrument?

The development of this new instrument can provide an additional tool for researchers to investigate teachers' self-perceived assessment literacy through its three-dimensional framework. The developed instrument can help researchers understand how teachers view themselves regarding the competencies in assessment. These results can serve as baseline information that policymakers can use to address issues related to assessment. Future researchers can also use the findings to conduct more studies about this construct.

This current study sought to develop the SRALI. The researchers asked five science education experts to validate the initial instrument. Then, a pilot run was performed. Afterward, the instrument was administered to teachers to determine its reliability and validity.

METHOD

Research Design and Procedures

The study used the survey method of research. This method allows the collection of information from individuals using their answers to the questions given by the researchers (Check & Schutt, 2012).

The study started with developing the items that will constitute the SRALI. The 45-item instrument underwent a stage of validation by five experts in physics education. After the necessary revisions, the instrument was pilot-tested on 42 science teachers to determine its reliability. Factor analysis was not performed since the sample used was not adequate.

After its pilot run, the SRALI underwent another series of minor revisions. After these revisions, the SRALI was again administered to 107 teachers from different schools and education levels in the Philippines. A question in the online form is given to seek the participants' consent to join the study. Those who consented to participate in the study and provided complete responses were included. Data was then analyzed to determine the instrument's factor structure, validity, and reliability.

Instrument

The SRALI is originally a 45-item instrument that seeks to describe teachers' self-reported assessment literacy. It has three sections corresponding to the three dimensions of assessment literacy by Pastore and Andrade (2019). The instrument went through a series of revisions throughout the study. An online version of the SRALI using Google Forms was administered to the participants. The form also includes questions that describe the participants' demographic profiles.

Initially, the instrument was made up of three sections parallel to the three dimensions of assessment literacy: Section I (Conceptual Knowledge of Assessment), Section II (Socio-emotional Aspects of Assessment), and Section III (Assessment Practices). Section I corresponds to the conceptual knowledge dimension of assessment literacy. It was initially made up of 15 items about assessment and its processes. The participants can signify their level of agreement on each statement using a five-point Likert scale. Meanwhile, Section II of the instrument corresponds to the socio-emotional dimension of assessment literacy. It originally comprised seven items related to assessment as a social practice. In Sections I and II, the participants can signify their level of agreement with each statement in this section using a five-point Likert scale. Section III is related to the praxeological dimension of assessment literacy. It is originally made up of 23 items related to various assessment practices across the different stages of assessment. The participants will answer the items using a five-point Likert scale to describe the frequency of using the strategies mentioned in each statement in their assessment practice.

Data Analysis

All statistical analyses were performed using IBM SPSS Statistics version 26. In the pilot run, the Cronbach alpha coefficient was used to determine the reliability of the SRALI. In its second run, data from the survey underwent exploratory factor analysis (EFA) to determine the instrument's factor structure. Afterward, the Cronbach alpha coefficient was again used to determine the instrument's reliability based on the factor structure produced in the EFA. While McDonald's omega is a more accurate measure of reliability, the researchers decided to use Cronbach alpha since the software used for statistical analysis doesn't compute the omega coefficient. Also, simulation studies performed by Malkewitz et al. (2023) and Orcan (2023) revealed that Cronbach's alpha and McDonald's omega computations yielded similar results except for a small number of items and sample size.

FINDINGS AND DISCUSSION

Development and Initial Validation of the SRALI

The conceptualization performed to prepare an instrument to measure self-reported assessment literacy allows the identification of the basis for developing it (Syahfitri et al., 2019). The developed SRALI is initially made up of three sections parallel to the three dimensions of assessment literacy by Pastore and Andrade (2019). Section I, called *Conceptual Knowledge of Assessment*, corresponds to the conceptual knowledge dimension. It comprises 15 items about what teachers know about the preparation and implementation of assessment and the processing, appreciation, and communication of assessment results. The participants responded to the items using a five-point Likert scale (1 – strongly disagree, 2 – disagree, 3 – undecided, 4 – agree, 5 – strongly agree) to describe their level of

agreement on each statement. Since items 5, 10, and 15 were originally negatively stated, the scale was reversed.

Section II of the instrument is called *Socio-emotional Aspects of Assessment*. It corresponds to the socio-emotional dimension. It comprises seven items related to assessment as a social practice. The participants answered the items using a five-point Likert scale (1 – strongly disagree, 2 – disagree, 3 – undecided, 4 – agree, 5 – strongly agree) to describe their level of agreement on each statement. Since item 5 was initially stated negatively, the scale has been reversed.

Section III of the instrument is called Assessment Practices. It corresponds to the praxeological dimension. It comprises 23 items related to various assessment practices across its different stages. The participants will answer the items using a five-point Likert scale (1 – not at all, 2 – a few times, 3 – sometimes, 4 – most of the time, 5 – all the time) to describe how often they practice each item.

Five experts in physics education were asked to validate the initially prepared instrument using the validation rating scale adapted from the work of Oducado (2020). It is a 13-item questionnaire that allows the validator to evaluate the proposed instrument for assessment literacy by describing their level of agreement with each statement using a five-point Likert scale. The mean for each item was determined. The validators were also asked for comments and suggestions on improving the instrument.

A summary of the ratings given by the validators per item in the validation rating scale can be seen in Table 1. The validators unanimously agreed strongly on statements 1, 2, and 3. As one of the validators added, "*The items on (the) conceptual knowledge of assessment, socio-emotional aspects of assessment, and assessment practices are certainly designed to elicit teachers' reflections on assessment*." The case is also the same with statements 5, 6, 7, 9, 11, 12, and 13, where all the validators gave a "5" rating. On the other hand, statements 4 and 10 got a mean rating of 4.80 from the validators.

Table 1
Validators' mean rating on the validity of the developed self-reported assessment literacy instrument for teachers

Statements	Mean Rating
1. The items in the instrument are relevant to answer the objectives of the study.	5.00
2. The items in the instrument can obtain depth to the constructs being measured.	5.00
3. The instrument has an appropriate sample of items for the construct being measured.	5.00
4. The items and their alternatives are neither too narrow nor limited in their content.	4.80
5. The items in the instrument are stated clearly.	5.00
6. The items on the instrument can elicit responses that are stable, definite, consistent, and not conflicting.	5.00
7. The terms adapted in the scale in the scale are culturally appropriate.	5.00
8. The layout or format of the instrument is technically sound.	4.40
9. The responses on the scale show a reasonable range of variation.	5.00
10. The instrument is not too short or long enough that the participants will be able to answer it within a given time.	4.80
11. The instrument is interesting such that participants will be induced to respond to it and accomplish it fully.	5.00
12. As a whole, the instrument could answer the basic purpose for which it is designed.	5.00
13. The instrument is culturally acceptable when administered in the local setting.	5.00
Mean	4.92

Meanwhile, the validators gave a mean rating of 4.40 for statement 8, the lowest among all the statements. Three of the validators suggested the same improvement to the layout of the printed version of the proposed instrument. As one of the validators mentioned, "*The survey is well-designed. However, the layout may be bothersome because it makes it difficult to understand the order of the format items. Put a line between them so the survey reader isn't confused*." Another validator added that the direction should specify what the respondent will do to accomplish the instrument. The researchers carried out all these suggestions.

Moreover, one of the validators asked if the researchers could reconsider item number 7 (*I administer my assessment activities using traditional methods (e.g., paper and pencil, etc*.)) given the conditions at the time of the validation. However, the researchers decided to retain the items since some schools still use traditional methods of administering assessment activities, and some schools have already returned to face-to-face classes during that time. In addition, the researchers specified in the direction for Section III that the participants will answer it based on their current assessment practices.

Overall, the developed self-reported assessment literacy instrument for teachers got a mean rating of 4.92 out of 5 from the validators. This confirms the face validity of the instrument developed by the researchers.

Pilot Run Results

The SRALI has undergone a pilot run on 42 science teachers. Cronbach's alpha coefficient for the whole instrument and per section was determined to find out if the instrument is reliable. For the whole instrument, Cronbach's alpha coefficient is 0.877. It means that the instrument has a very good internal consistency (Daud et al., 2018). Per section, Cronbach's alpha coefficients for Sections I, II, and III are 0.853, 0.701, and 0.793, respectively. All these values are considered acceptable (Daud et al., 2018; Nunnally, 1978). Based on these results, the instrument can be considered a reliable tool for determining the self-reported assessment literacy of the participants.

After the pilot run, a minor revision was made to the SRALI. All the statements stated negatively in Sections I and II were revised. They were restated positively to facilitate easier data collection using Google Forms.

Second Run: Demographic Profile of the Respondents

In the second administration of the SRALI, 107 teachers were included in this study. There are more female respondents (69%) than male respondents (31%). In addition, 66% of the participating teachers teach at the junior high school level, 20% at the elementary level, 8% at the senior high school level, and 6% at the college level. Meanwhile, more teachers hold an undergraduate degree (62%), while others hold a graduate degree. When asked if they took an assessment course during their undergraduate or graduate studies, most participants responded "Yes" (73%). The participants' ages ranged from 22 to 64 years, with a mean age of 39. Furthermore, the participants' teaching years range from 0 to 40 years, with a mean of 13.5 years.

Second Run: Exploratory Factor Analysis Results

In the second run of the SRALI, an exploratory factor analysis (EFA) was performed on the obtained data to determine the factor structure. EFA is used in item reduction in instrument development to identify related items and remove items with confusing or identical meanings (Suwono et al., 2021). Since Sections I and II used the same scale, they were analyzed together. On the other hand, Section III was separately analyzed since it used a different scale in the responses to its statements.

For the first two sections, there are 22 items, 15 for Section I and seven for Section II. To determine if the items in Sections I and II follow the conceptual knowledge and socio-emotional dimensions of

assessment literacy, EFA was performed using the principal components method in the responses. The correlation matrix for the items showed values exceeding 0.300. Tabachnick and Fidell (2013) mentioned that this result suggests that the principal components method is appropriate for the data.

Like the study of Arshad et al. (2022), the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test were performed to determine if EFA can be performed on the given data and proceed to further statistical operations. The KMO value for Sections I and II combined is .906. This measure is considered marvelous (Analysis INN, 2020) and is above the minimum acceptable value of .600 (Tabachnick & Fidell, 2013). This means that a strong partial correlation is present. Therefore, factor analysis can be performed. On the other hand, Bartlett's test produced significant test results (χ^2 (231) = 1957.783, $p < .000$). It means that the correlation matrix in the data from each section is not an identity matrix (Analysis INN, 2020; Tabachnick & Fidell, 2013). Therefore, the variables are related and can be subjected to factor analysis.

Based on the initial solution, three factors with an eigenvalue greater than one were generated, explaining 68.99% of the variance. Table 2 shows the factors generated and the total variance explained by each factor.

Table 2
Factors generated for sections I and II

	Initial eigenvalues			Rotation sums of squared loadings		
Factors	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
Factor 1	11.526	52.391	52.391	5.864	26.652	26.652
Factor 2	2.491	11.321	63.712	4.736	21.528	48.180
Factor 3	1.163	5.287	68.999	4.580	20.819	68.999

After determining the number of factors extracted, EFA was repeated using varimax rotation to interpret the results better. The percent of variance contributed by each factor after rotation can also be seen in Table 2. These values for each factor are considered substantial, suggesting that the factors generated were important in the instrument (Tabachnick & Fidell, 2013).

Factor loadings for each item in Sections I and II can be seen in Table 3. The items are ordered according to how they were arranged in the original SRALI, with items of Section I coded as A plus a number (e.g., A1 is item 1 of Section I) and items of Section II coded as B plus a number (e.g. B2 is item 2 of Section II). The cut-off used for the factor loadings is .5. The data shows all items loaded to one factor except items A5 and A12. Since they cross-loaded to two factors, they were not included in the subsequent analysis. These results showed that the items almost followed the two dimensions of assessment literacy by Pastore and Andrade (2019). Two factors are related to the conceptual knowledge dimension, while one is related to the socio-emotional dimension.

Factor 1 is named *Conceptual Knowledge of the Assessment Process* since items loaded in this factor describe the teachers' conceptual knowledge on how to perform assessment. Surprisingly, one item from Section II, which was expected to load with the factor related to the socio-emotional dimension, was loaded in this factor. It might be because the statement is more aligned with the conceptual knowledge dimension of assessment literacy. Another factor related to the conceptual knowledge dimension of assessment literacy is Factor 3. This factor is called *Conceptual Knowledge of the Nature of Assessment* since items that are loaded here describe the teachers' knowledge of the definition, principles, and types of assessment. On the other hand, Factor 2 is called the *Socio-emotional Aspects of Assessment*. All the items loaded in this factor are related to the socio-emotional dimension of assessment literacy.

Table 3
Factor loadings for section i and ii items

Items	Factor 1	Factor 2	Factor 3	Communality after extraction
A1	.249	.272	.680	.599
A2	.327	.418	.636	.685
A3	.138	.399	.671	.629
A4	.356	.331	.606	.603
A5	.529	.162	.662	.744
A6	.694	.229	.363	.666
A7	.743	.062	.484	.789
A8	.751	.067	.407	.734
A9	.820	.171	.198	.740
A10	.759	.131	.289	.677
A11	.481	.209	.606	.643
A12	.566	.228	.512	.635
A13	.441	.168	.691	.700
A14	.579	.195	.498	.622
A15	.690	.248	.237	.594
B1	.676	.427	.079	.645
B2	.402	.742	.205	.754
B3	.058	.808	.337	.770
B4	.117	.760	.231	.645
B5	.046	.849	.287	.805
B6	.216	.812	.304	.799
B7	.427	.721	-.031	.702

Table 3 also shows the communality values after extraction for each item. These values are reasonably high. This means that each item loading in each factor can be well predicted by their respective factors (Yildirim & Correia, 2015).

Meanwhile, EFA was also performed separately for the data obtained from Section III of the SRALI to determine its factor structure. The section has 23 items related to the praxeological dimension of assessment literacy. The principal components method was performed in the data. The correlation matrix for the items showed that most values exceed 0.300. This result suggests that the principal components method is appropriate for the data (Tabachnick & Fidell, 2013).

The KMO value for Section III is .863. This measure is considered meritorious (Analysis INN, 2020) and is also above the minimum acceptable value of .600 (Tabachnick & Fidell, 2013). This means that a strong partial correlation is also present. Therefore, factor analysis can also be performed. On the other hand, Bartlett's test also produced significant test results ($\chi 2$ (253) = 2001.878, $p < .000$). It means that the correlation matrix in the data from each section is not an identity matrix (Analysis INN, 2020; Tabachnick & Fidell, 2013). Therefore, the variables are related and can be subjected to factor analysis.

Based on the initial solution, five factors with an eigenvalue greater than one were generated, explaining 73.90% of the variance. After determining the number of factors extracted, EFA was repeated using varimax rotation to interpret the results better. However, only two items were loaded in Factor 4, and one loaded to Factor 5. Also, five items were loaded on two factors if a .5 cutoff was used. Factors with fewer than three items are considered weak and unstable (Costello & Osborne, 2005). Therefore, the researchers decided to repeat the analysis using varimax rotation and consider only the first three factors (with eigenvalues greater than 1.15) to determine if a better interpretation of the results could be obtained.

After the rerun of the analysis, the three factors with eigenvalues greater than 1.15 were generated, which explains 64.54% of the variance. The percent of variance contributed by each factor after rotation can be seen in Table 4. These values for each factor are also considered substantial, suggesting that the factors generated were important in the instrument (Tabachnick & Fidell, 2013).

Table 4
Factors generated for sections iii

	Initial eigenvalues			Rotation sums of squared loadings		
Factors	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
Factor 1	11.667	50.726	50.726	6.670	28.999	28.999
Factor 2	1.720	7.476	58.202	4.219	18.343	47.342
Factor 3	1.457	6.336	64.538	3.955	17.196	64.538

Factor loadings for each item in Section III can be seen in Table 5. The items are ordered according to how they were arranged in the original SRALI, with items coded as C plus a number (e. g. C1 is item 1 of Section III). The cut-off used for the factor loadings is .5. The data shows all items loaded to one factor except items C20 and C21. Since they cross-loaded to two factors, they were not included in the subsequent analysis. These results showed that the items follow a three-factor structure corresponding to different types of assessment practices.

For Section III, Factor 1 is named *Usual Assessment Practices* since items loaded in this factor include assessment practices teachers usually do, like considering comprehension, applying knowledge and higher-order thinking skills in preparing assessment tools, recording results, and observing the ethical principles of assessment. Factor 2 is called *Constructivist Assessment Practices* since the items loaded in this factor include practices like using alternative assessment methods, giving accurate and timely feedback, and considering the learners' effort to learn in grading them. Factor 3 is named *Traditional Assessment Practices* since the items loaded in this factor refer to traditional practices like using paper and pencil tests and considering learners' improvement, effort, and participation in grading them.

Table 5
Factor loadings for section iii items

Items	Factor 1	Factor 2	Factor 3	Communality after extraction
C1	.816	.253	.111	.743
C2	.694	.275	.182	.591
C3	.635	.284	.229	.536
C4	.141	.098	.843	.740
C5	.557	.488	.200	.588
C6	.288	.763	.119	.678
C7	.185	.015	.773	.632
C8	-.095	.532	.252	.356
C9	.444	.380	.533	.626
C10	.423	.316	.660	.715
C11	.319	.541	.481	.625
C12	.360	.286	.670	.660
C13	.407	.572	.387	.643
C14	.480	.670	.044	.681
C15	.741	.063	.370	.690
C16	.817	.027	.298	.757
C17	.533	.399	.434	.632
C18	.178	.717	.040	.547
C19	.529	.388	.448	.631
C20	.649	.528	.037	.701
C21	.537	.528	.248	.628
C22	.746	.273	.294	.717
C23	.762	.177	.338	.726

For Section III, Factor 1 is named *Usual Assessment Practices* since items loaded in this factor include assessment practices teachers usually do, like considering comprehension, applying knowledge and higher-order thinking skills in preparing assessment tools, recording results, and observing the ethical principles of assessment. Factor 2 is called *Constructivist Assessment Practices* since the items loaded in this factor include practices like using alternative assessment methods, giving accurate and timely feedback, and considering the learners' effort to learn in grading them. Factor 3 is named *Traditional Assessment Practices* since the items loaded in this factor refer to traditional practices like using paper and pencil tests and considering learners' improvement, effort, and participation in grading them.

Table 5 also shows the communality values after extraction for each item in Section III. These values are also reasonably high. This means that each item loading in each factor can be predicted well by their respective factors (Yildirim & Correia, 2015), similar to the items in Sections I and II.

Based on these results, the SRALI was reduced to 41 items, with two items being rejected from Section I and two from Section III. Moreover, one of the items in Section II is loaded with one of the factors in Section I. As a result, Section I gained one item from Section II, B1, and is now made up of 14 items divided into two factors. Section II, considered a single factor, has six items. Section III, made up of three factors, is made up of 21 items.

Second Run: Reliability of the SRALI

The reliability of the instrument was recomputed based on the resulting factor structure. The Cronbach's alpha coefficient was calculated for each factor for Sections I and II combined and Section III. Table 6 shows the reliability analysis for the items in Sections I and II.

Table 6
Reliability analysis for sections i and ii

Items	Corrected item-total correlation	Cronbach's alpha if item deleted	Cronbach's alpha
Factor 1 – Conceptual Knowledge of the Assessment Process			.925
Item A6	.762	.913	
Item A7	.820	.908	
Item A8	.789	.910	
Item A9	.772	.911	
Item A10	.750	.913	
Item A14	.710	.916	
Item A15	.691	.919	
Item B1	.666	.920	
Factor 2 – Socio-emotional Aspects of Assessment			.919
Item B2	.779	.900	
Item B3	.796	.898	
Item B4	.720	.908	
Item B5	.811	.896	
Item B6	.848	.890	
Item B7	.657	.919	
Factor 3 – Conceptual Knowledge of the Nature of Assessment			.882
Item A1	.678	.873	
Item A2	.779	.856	
Item A3	.666	.875	
Item A4	.725	.865	
Item A11	.686	.872	
Item A13	.690	.871	
Overall			.950

Table 6 shows the corrected item-total correlation and Cronbach's alpha if the item deleted for each item in Sections I and II to determine the reliability of each item. All values for the corrected item-total correlation are above .400. It means that each item correlates with the other items, excluding itself (Yildirim & Correia, 2015). Also, Cronbach's alpha if item deleted values show that no item increases the reliability if removed. Therefore, none of the items in Sections I and II are subject to further rejection. These findings result in the retention of the number of items for Sections I and II at 14 and 6, respectively.

Table 6 also shows that the overall reliability coefficient for Sections I and II is .950. It means that this part of the SRALI has an acceptable reliability (Nunnally, 1978). Daud et al. (2018) classify this reliability as very good. Each factor has good internal consistency, with coefficients fairly above the acceptable value of .700 (Nunnally, 1978).

Meanwhile, Table 7 shows the reliability analysis for Section III. It shows the corrected item-total correlation and Cronbach's alpha if the item deleted for each item in this section to determine the reliability of each item. For the items in Factors 1 and 3, all values for the corrected item-total correlation are above .400, which is acceptable (Yildirim & Correia, 2015). Also, Cronbach's alpha if item deleted values show that no item in Factors 1 and 3 increases the reliability if they are removed. Therefore, none of the items here are subject to further rejection.

Table 7
Initial reliability analysis for sections iii

Items	Corrected item-total correlation	Cronbach's alpha if item deleted	Cronbach's alpha
Factor 1 – Usual Assessment Practices			.934
Item C1	.772	.925	
Item C2	.713	.928	
Item C3	.680	.930	
Item C5	.647	.932	
Item C15	.750	.926	
Item C16	.783	.925	
Item C17	.713	.928	
Item C19	.711	.928	
Item C22	.796	.924	
Item C23	.812	.923	
Factor 2 – Constructivist Assessment Practices			.829
Item C6	.696	.745	
Item C8	.274	.857	
Item C11	.631	.765	
Item C13	.701	.753	
Item C14	.666	.758	
Item C18	.586	.772	
Factor 3 – Traditional Assessment Practices			.948
Item C4	.709	.827	
Item C7	.629	.854	
Item C9	.669	.837	
Item C10	.746	.822	
Item C12	.695	.831	
Overall			.948

However, an item in Factor 2, item C8, has a corrected item-total correlation of less than .400. Also, the reliability of Factor 2 increases significantly if it is removed. Due to this, the researchers decided to drop this item. This action reduces the number of items for Section III to 20.

The Cronbach's alpha coefficients for Factors 1 and 3 are fairly above the acceptable value of .700. It means that these factors also have good internal consistency (Nunnally, 1978). Daud et al. (2018) also classify this reliability as very good. Meanwhile, after dropping item C8, the reliability of Factor 2 increases to .857. This value is also considered acceptable (Nunnally, 1978; Daud et al., 2018). Overall, the reliability of Section III increases to .951 after rejecting item C8. This is also considered very good (Daud et al., 2018).

CONCLUSIONS AND RECOMMENDATIONS

The researchers developed and validated the SRALI. Five science education experts validated the initial instrument. Then, a pilot test was performed. After a series of revisions based on the results of the previous processes, the SRALI was administered to 107 teachers.

The EFA results showed that the SRALI almost followed the three-dimensional model of assessment literacy by Pastore and Andrade (2019). Sections I and II produced three factors, two relating to the conceptual knowledge dimension and one relating to the socio-emotional dimension. Meanwhile, Section III on assessment practices produced three factors. Moreover, the results revealed that the instrument is reliable based on the Cronbach alpha coefficients of each factor and overall.

The 40-item final version of the instrument is shown in Table 8. Sections I and II were combined to form the new Section I, called the *Conceptual Knowledge and Socio-Emotional Dimensions.* Meanwhile, the old Section III becomes the new Section II, called *Assessment Practices*. Each statement is preceded by its original item code from the initial draft of the SRALI.

Table 8

Final version of the SRALI

Section I. Conceptual Knowledge and Socio-Emotional Dimensions
Factor 1 – Conceptual Knowledge of the Assessment Process
A6. I am well-versed in choosing the appropriate assessment tool to be used.
A7. I have a comprehensive knowledge of developing assessment tools.
A8. I am very knowledgeable in implementing different assessment strategies.
A9. I have a deep understanding of the different strategies for monitoring learners' performance.
A10. I know the different mechanisms of providing feedback to learners.
A14. I am well-versed in the different strategies for communicating assessment results to learners.
A15. I am familiar with the different strategies for communicating assessment results to parents and/or guardians and other stakeholders.
B1. I effectively work with my fellow teachers and other stakeholders to make sense of my assessment practices.
Factor 2 – Socio-emotional Aspects of Assessment
B2. I greatly value my role and responsibilities as an assessor of learning.
B3. The learners' rights involved in assessment are very important to me.
B4. I always avoid doing different malpractices in assessment.
B5. I value ethical principles and standards of assessment.
B6. I am completely aware of the impact of assessment practices and results on learners.
B7. I am very familiar with the impacts of assessment practices and results on the parents/guardians and other stakeholders.
Factor 3 – Conceptual Knowledge of the Nature of Assessment
A1. I have extensive knowledge of the definition of assessment.
A2. I completely understand the principles of assessment.
A3. I am completely aware of the different types of assessment (e.g., diagnostic, formative, summative, etc.).
A4. I am very familiar with the different assessment tools (e.g., tests, performance-based assessment, etc.) used in teaching.
A11. I have a detailed understanding of how to score assessment activities.
A13. I have extensive knowledge of using assessment results in making decisions on students' learning progress.
Section II. Assessment Practices
Factor 1 – Usual Assessment Practices
C1. I consider learners' comprehension when preparing assessment tools.
C2. My assessment tools cover the learners' application of knowledge.
C3. In preparing my assessment tools, I consider the learners' higher-order thinking skills.
C5. I use performance-based assessments.
C15. I record the assessment results of my learners.
C16. Assessment records are used to track the learners' progress.
C17. I inform the learners of their progress based on their assessment results.
C19. Assessment results are used to adjust instruction.
C22. Ethical principles and standards of assessment are observed.
C23. The learners are involved in the assessment process.
Factor 2 – Constructivist Assessment Practices
C6. I use alternative assessment methods (e. g., self-assessment, portfolio assessment, etc.).
C11. The learners' adherence to the learning objectives is considered when I evaluate them.
C13. I give accurate feedback to my learners.
C14. Timely feedback is provided to learners.
C18. I communicate the assessment results of the learners to their parents and/or guardians.
Factor 3 – Traditional Assessment Practices
C4. I use traditional types of assessment (e. g. multiple-choice tests, true or false, etc.).
C7. I administer my assessment activities using traditional methods (e. g. paper and pencil).
C9. I evaluate learners based on their improvement.
C10. I consider the learners' effort to learn in assessing them.
C12. Class participation is considered in assessing learners.

Therefore, it can be concluded that the SRALI is a valid and reliable instrument that can be used to measure teachers' self-reported assessment literacy. The use of the instrument in studies about assessment literacy is strongly recommended. Also, the instrument can be readministered to teachers of other demographic backgrounds to confirm the instrument's factor structure revealed by the EFA performed in this study. Also, studies can be performed to determine the relationship between self-perceived assessment literacy, as measured using the SRALI, and the actual assessment literacy of teachers. Actual assessment literacy can be determined using conceptual tests in assessment, interviews, and actual classroom observations.

REFERENCES

Analysis INN. (2020). *KMO and Bartlett's test of sphericity*. https://www.analysisinn.com/post/kmo-and-bartlett-s-test-of-sphericity/

Arshad, S. S., Zaman, S., & Nazir, A. (2022). Development and Validation of Scale for Assessment of Followership among School Teachers. *International Journal of Instruction, 15*(3), 1031-1046.

Check, J. & Schutt, R. K. (2012). *Research Methods in Education*. SAGE Publication, Inc.

Clores, M. A. & Reganit, A. A. R. (2019). Investigating the Assessment Literacy of Teachers in Private Junior High Schools in the Philippines. *Humanities, Arts and Social Sciences Studies, 20*(2), 461-476.

Costello, A. B. & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation, 10*, doi: https://doi.org/10.7275/jyj1-4868

Daud, K.A.M., Khidzir, N.Z., Ismail, A.R. and Abdullah, F.A. (2018). Validity and reliability of instrument to measure social media skills among small and medium entrepreneurs at Pengkalan Datu River. *International Journal of Development and Sustainability, 7*(3), pp. 1026-1037.

Department of Education. (2015). *Policy Guidelines on Classroom Assessment for the K to 12 Basic Education Program.* Pasig, Metro Manila, Philippines: Department of Education.

Department of Education. (2017). *National Adaption and Implementation of the Philippine Professional Standards for Teachers*. Pasig, Metro Manila, Philippines: Department of Education.

Department of Education. (2020). *Interim Guidelines for Assessment and Grading in Light of the Basic Education Learning Continuity Plan*. Pasig, Metro Manila, Philippines: Department of Education.

Gotch, C. M., & McLean, C. (2019). Teacher outcomes from a statewide initiative to build assessment literacy. *Studies in Educational Evaluation, 62*, 30-36.

Hailaya, W. M. (2014). Teacher Assessment Literacy and Student Outcomes in the Province of Tawi-Tawi, Philippines (Doctoral thesis, The University of Adelaide, Australia). Retrieved from https://digital.library.adelaide.edu.au/ dspace/bitstream/2440/99098/2/02whole.pdf

Malkewitz, C. P., Schwall, P., Meesters, C., & Hardt, J. (2023). Estimating reliability: A comparison of Cronbach's α, McDonald's ω_t and the greatest lower bound. *Social Sciences & Humanities Open, 7*, 1-10.

Mellati, M., & Khademi, M. (2018). Exploring teachers' assessment literacy: Impact on learners' writing achievements and implications for teacher development. *Australian Journal of Teacher Education, 43*(6), 1-18.

Mertler, C. A., & Campbell, C. (2005). *Measuring teachers' knowledge & application of classroom assessment concepts: Development of the Assessment Literacy Inventory*. [Paper presentation]. American Educational Research Association, Montreal, Quebec, Canada.

Muhammad, F. H. N., & Bardakci, M. (2019). Iraqi EFL teachers' assessment literacy: Perceptions & practices. *Arab Society of English Language Studies, 10*(2), 431-442.

Napanoy, J. B. & Peckley, M. K. (2020). Assessment Literacy of Public Elementary School Teachers in the Indigenous Communities in Northern Philippines. *Universal Journal of Educational Research, 8* (11B), 5693-5703.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Oducado, R. F. (2020). Survey Instrument Validation Rating Scale. https://doi.org/10.13140/RG.2.2.25263.59040

Ogan-Bekiroglu, F., & Suzuk, E. (2014). Pre-service teachers' assessment literacy and its implementation into practice. *The Curriculum Journal, 25*(3), 344-371.

Orcan, F. (2023). Comparison of Cronbach's alpha and McDonald's omega for ordinal data: Are they different? *International Journal of Assessment Tools in Education, 10*(4), 709-722.

Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education, 84*, 128-138.

Prizovskaya, V. (2017*). Identifying New Jersey teachers' assessment literacy as precondition for implementing student growth objectives* (Doctoral dissertation). Available from ProQuest LLC. (Record No. 10287402)

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan, 77*(3), 238–245.

Suwono, H., Mauilidia, L., Saefi, M., Kusairi, S., & Yuenyong, C. (2021). The development and validation of an instrument of prospective science teachers' perceptions of scientific literacy. *EURASIA Journal of Mathematics, Science and Technology Education, 18*(1), 1-16.

Syahfitri, J., Firman, H., Redjeki, S. & Sriyati, S. (2019). Development and validation of critical thinking disposition test in biology. *International Journal of Instruction, 12* (4), 381-392.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. Boston: Pearson.

Yildirim, C. & Correia, A. P. (2015). Exploring the dimensions of nomophobia: Development and validation of a self-reported questionnaire. *Computers in Human Behavior, 49*, 130-137.