

## Development and Validation of Teachers' Practices on Formative Assessment Scale (TPFAS): A Measure Using Feedback Loop Model

Faith Celeste B. Ole

*Carlos Hilado Memorial State College, Talisay City, Negros Occidental, Philippines*

faith.celeste29@gmail.com

First draft received: 16 May 2020    Date Accepted: 1 Aug 2020    Final proof received: 28 Aug 2020

### Abstract

The practices on formative assessment are recognized as an essential aspect of teaching and learning, yet the deluge of data, both formal and informal, gathered by teachers through classroom assessments inept them on how to analyze and respond to. A data-driven approach called the Feedback Loop Model is designed to enable the teachers to interpret these data for determining the next steps in the process of teaching and learning. For this reason, a study was conducted to develop and validate an instrument concerning teachers' practices on formative assessment utilizing the elements of the Feedback Loop Model. An instrument called Teachers' Practices on Formative Assessment Scale (TPFAS) anchoring on the elements of the model was pilot tested to 157 science teachers in the Philippines. Teachers' responses were analyzed using Cronbach's alpha reliability coefficient on the Feedback Loop constructs and Confirmatory factor analysis for the entire instrument. Findings suggest the deletion of 10 survey items from the initial 44 items in which the scale provided a valid and highly reliable measure in determining teachers' practices on formative assessment. The TPFAS instrument exhibited an overall reliability coefficient consistency of 0.93 indicating an acceptable standard for an instrument used. Moreover, reliability analysis was conducted within each subscale which exhibited internal consistency reliability (alpha) ranging from .819 to .884 for the four subscales or constructs.

**Keywords:** Feedback loop; practices on formative assessment; TPFAS; validation

### To cite this paper (in APA style):

Ole, F. C. B. (2020). Development and validation of teachers' practices on formative assessment scale (TPFAS): A measure using feedback loop model. *International Journal of Education*, 13(1), 53-62. doi: 10.17509/ije.v13i1.24715

### INTRODUCTION

Assessment is an integral and essential part of effective teaching. This systematic basis for making inferences about learners' development allows them to explore their abilities and determine their performance in a class. This also helps teachers assess their performance and identify their strengths in delivering effective learning to students. According to Laurillard (1994, p. 6), "Teachers need to discern more than just their subject. They need to identify the various ways it can be understood, the ways it can be misunderstood, and what counts as understanding; they need to know how individuals experience the subject." Reflecting on these would make us realize the importance of strengthening the use of formative assessments. This type of assessment is more "diagnostic" than the evaluative aspect because this allows the teachers to inform their ongoing instruction using the knowledge of student understandings (Black, 1993 as cited in Ruiz-Primo & Furtak, 2007). Based on considerable studies, classroom-based formative assessments when used

appropriately can positively affect substantial learning (Ash & Levitt, 2003). The value of formative assessment in promoting learning has immensely proven to be an effective method of monitoring students' learning. The greater gains made through formative assessment revealed a greater impact on the overall performance of a group (Black et al., 2003).

Even though formative assessment is recognized as an important practice which teachers need to be knowledgeable and skillful at, some researchers argued (e.g., Athanases & Achinstein, 2003; Black et al., 2003; Schneider & Randel, 2009 as cited in Albuquerque, 2014) that many teachers have insufficient levels of assessment literacy and lack expertise in sound formative assessment practices. Few teachers understand the pedagogical implications of scaffolding learning through utilizing formative assessments (Buck & Trauth-Nare, 2009) and its underlying importance in optimizing teaching practice that support student learning. Hence, an opportunity

for other researchers to explore the potential needs for improvement in its practices at schools.

Although multiple instruments exist to evaluate the teacher classroom practices (Danielson, 2013; Pianta, La Paro, & Hamre, 2008 as cited in Marshall, et al., 2016), most studies have focused on one aspect of assessment, either attitudes or practices or both teachers' attitudes and practices regarding assessment. One particular study which used both was conducted by Yan and Cheng (2015); this study aimed to explore the relationships among teachers' attitudes, intentions, and practices regarding formative assessment under the framework of the Theory of Planned Behaviour (TPB). TPB has been successful in providing a better interpretation of diverse behaviours in the western and Hong Kong settings (Yan & Cheng, 2015). Ajzen's TPB (1991) is a rigorous theoretical framework which has the potential to provide prediction and explanation of teachers' intentions and practices of formative assessment. Using this framework, an instrument called "A Teacher's Conceptions and Practices of Formative Assessment Questionnaire" containing seven scales was developed to assess the five components in the TPB framework regarding formative assessment. The study of Yan and Cheng (2015) has significantly contributed in building a structural understanding of teachers' attitudes, intentions, and practices on formative assessment. However, the TPB components have not been effective predictors of teachers' formative assessment practices. The proposed TPB-based model was not able to thoroughly explain the teachers' reported formative assessment practices, citing more evidences on the external and contextual factors when examining their formative assessment practices.

As a science teacher, it is significant to be abreast in exploring and considering various instructional approaches that may help us grow in our practice of assessment and to improve student learning. Science teachers are subjected to a surge of data in the course of their work, both formal and informal - such as written works in the form of tests, laboratory reports, worksheets, - and even student expressions and emotions (Furtak et al., 2016). Issues related to these data-driven approaches are classroom assessments conducted by teachers to ascertain what students know and are able to do. Yet the toughest thing about all these information is on how one must analyze and respond to as a science teacher, and able to learn to determine what it means. Generally, formative assessment is characterized as a three-step process in which a teacher sets learning goals, determines what students currently know, and provides feedback to support students in meeting the goals (NRC, 2001). In the new framework, called the Feedback Loop (Furtak et al., 2016), the three steps—designing and selecting tools, collecting data, and making inference—are the ways of explicating and determining what students currently know. The piece that is not represented as a stand-alone step in the Feedback Loop is the final element of formative

assessment, which is providing helpful feedback to move students toward learning goals. This step has the final arrow connecting inferences and learning goals. It is the feedback that connects what has been inferred on what students know and are able to do with the goals for student learning – a process that identifies the gaps. This gap, consequently paved the way for the researcher to use the Feedback Loop Model in helping science teachers to go beyond thinking about the pieces of data in isolation into reorienting them as part of a larger system that teachers can design and act on (Furtak et al., 2016).

Therefore, it is with this perspective that the researcher would like to develop and validate an instrument that is designed to determine the Science teachers' practices in conducting formative assessment. Specifically, this instrument is designed for science teachers in the Philippines. Any field of specialization in Science can utilize this instrument since the four-main steps in the process or elements such as *setting the goals, designing, selecting and adapting tools, collecting data, and making inference* of the Feedback Loop Model are applicable to any discipline.

## METHOD

According to Furtak et al., (2016), there are four elements used in the feedback loop model: goal, tool, data, and inference (Figure 1). The goal which is the cornerstone of the feedback loop is the first step. Building on a goal is the guiding principle that underlies what teachers are asking students to do. The second element refers to the common instruments teachers use to collect data about student learning, such as worksheets, classroom assessments, and observation protocol or any tool used to record a lesson. It can also anything that may not be written down or handed out but includes a plan to get students' ideas, as long as it is aligned with the goal the teacher intends to assess. The third element in the loop is the data where all the bits of information indicate the students' knowledge. These are yielded from the tools teachers created and used in the classroom. It can be in a form of quantitative, qualitative, formal or informal data from students. Formal types of data are usually the result of tools planned in advanced. Informal data include student's responses to questions asked on the fly, their expressions, and their participation in the class. And lastly - the arguably most important element in the loop is on how teachers make sense of the data that has been collected -- the process of making inference. This aspect is essential so as to inform teachers of the next step for their instruction.

Once teachers have gone through each of the four main steps in the process, the idea is to connect the inferences that were made back to the goals. This process of closing the loop is often called feedback in the formative assessment literature (Black & William, 1998) or, put more simply, using the information gained through the loop to move students forward in

their learning. The term feedback can be something that effective teachers use every day when they are being responsive to information about student thinking.

**Figure 1**  
*Feedback Loop Model*



### Instrument Development

After reviewing literatures related to formative assessments and introducing new framework called the Feedback Loop model, a Table of Specification (TOS), as presented in Table 1 was formulated to see whether all constructs of the framework were addressed. This TOS showing some examples of items found in the questionnaire also served as guidelines to ensure completeness of the instrument. Since instrument construction follows an iterative process, revisions in drafting the items and other important features were expected before sending it out to the experts for the pre-test.

**Table 1**  
*Specification for instrument construction and sample questions per construct*

Constructs and Item Placements	Operational Definition of Constructs and Indicators	Item Number and Sample questions
<p><b>Setting Goals</b>  (Items 1 to 12)</p>	<p>These are well-defined learning goals that assist teachers in adjusting their instruction and help students take more control of their learning because of the defined anticipated outcomes of lessons. Teachers know what they want students to know and be able to do.</p>	<p>Item 1: I see to it that my instruction for the activity is specific and easy to understand.</p> <p>Item 2: I see to it that the learning activity is clear and doable.</p>
<p><b>Designing, selecting, and adapting tools</b>  (Items 13-23)</p>	<p>These are multiple tools used by teachers in generating information relative to the learning goals.</p>	<p>Item 13: I prepare worksheets or tests that are aligned to the objectives of my lesson.</p> <p>Item 14: I find it easy to develop an activity or task related to the topic.</p>
<p><b>Collecting data</b>  (Items 24 -33)</p>	<p>This information could be in quantitative, qualitative, formal or informal form provided by the students. Formal types of data are usually the result of tools planned in advanced. Informal data include student's responses to questions asked on the fly, their expressions, and their participation in the class.</p>	<p>Item 26: I collect their outputs after doing the learning activity.</p> <p>Item 32: I capture students' thinking by collecting their written work or jotting down their ideas on a piece of paper.</p>
<p><b>Making inference</b>  (Items 34- 44)</p>	<p>Teachers can identify trends and patterns in data about what students know and are able to do. It can serve as guide for giving students feedback.</p>	<p>Item 41: I engage my students to do peer assessment.</p> <p>Item 43: I encourage my students to write feedback in helping other students improve their tasks.</p>

The initial draft prior to pre-testing constituted a total of 44-items employing a Likert scale type that measured frequency of practice according to these scales:

- Never (0% of the time) = 1
- Rarely ((25% of the time) = 2
- Sometimes (50% of the time) = 3
- Frequently (75% of the time) = 4
- Always (100% of the time) = 5

The choice of format items and response sets were deemed essential and appropriate in order for the researcher to determine their practices in conducting formative assessments during their Science classes.

Three content experts and five target users were given the initial pool of the 44-items to evaluate for content and face validity of the instrument. They were requested to comment and suggest the acceptability of the items. Such items must be clear, comprehensive and manifest representativeness of the construct as operationally defined and described in the TOS. These experts were also asked to rate on the relevance and appropriateness of the items relative to the constructs of Feedback Loop framework. Attached in the validation form was the statement of purpose and table of specification (TOS) so that experts would have an overview of what the instrument was all about. One content expert was a PhD in Science Education (Physics), and currently the Director of the Publication Office of a state university. Another expert was a senior high school instructor of a state university with a PhD degree in Physics Education and another expert was a Senior Education Program Specialist in the Bureau of Learning Delivery of the Department of Education – Central Office.

A 5-point category was used to validate the items, with 5 denoting a very relevant item and 1 indicating no relevance at all. After the revisions have been made that included deletion of some items and replacing them with new ones which were duly approved and recommended by content experts, the ratings were subsequently used for the calculation of the content validity coefficient of the items.

On the other hand, the same initial instrument was also given to five target users who served as potential respondents. They were all science teachers that have been requested by the researcher to answer the instrument, and gave valuable comments on the extent of how they understood the items. Three of them were PhD graduate students teaching Biology and Physics subjects in high school, and two of these three PhD students, were familiar with the Feedback Loop model); another expert was an Ed. D graduate whose research expertise has brought significant contributions to her affiliated state university and lastly, a junior high school master teacher teaching Physics for 12 years. With all the feedback from experts and target users, the TPFAS instrument was then prepared via an online survey form for the pilot testing.

### **Research Context and Participants**

With the advantage of using technology, the researcher was able to collect data from science teachers across the country via email and Facebook messenger. A total of 157 respondents participated through a convenience sampling method. This method was employed in order to administer and gather data outside the province of Negros Occidental. The inclusion criteria for the respondents could be any secondary and tertiary science teachers from public and private schools in the Philippines. Any field of specialization in Science was considered in this study provided that these teachers were currently teaching in the country; they were considered target users. The researcher also requested some respondents to share the form to their colleagues for wider dissemination of the survey form. Likewise, those non-science teachers acquainted to the researcher were also requested to assist her in sharing the form to their respective science teachers of the school. Upon accessing the survey form, each respondent was presented with an informed consent document that transcribed the purpose of the survey and how their responses could provide relevant information to the research endeavor. A timestamp feature of the google form registered the length of time a respondent could finish the survey. Most of them completed the online form for approximately 10-15 minutes. With the use of SPSS software, the responses were statistically analyzed and interpreted.

The science teachers who participated in this study consisted of 105 (67%) females and 52 (33%) males. With regards to teaching experience, 114 (72%) teachers had no more than 10 years of experience; 37 (24%) had 11-20 years of experience, and 6 (4%) were experienced teachers with more than 20 years of teaching experience. Of the 157 respondents, most are Biology majors 41 (26.1%), followed by General Science majors 38 (24.2%), then Chemistry 34 (21.7%) and the rest were Physics majors 32 (20.4%), Physical Sciences 9 (5.7%), and Earth Science 3 (1.8%) majors.

Among the respondents, 85 science teachers (54.14%) were handling one grade level only with majority of them teaching Grade 9 Science class. Another set of teachers (40) that constituted the 25.48% of the respondents handled two grade levels in which majority of them were senior high school teachers, teaching Grade 11 & Grade 12 classes; while the rest of the 32 respondents (20.38%) were found to be handling three to six grade levels in their respective schools and institutions. The multiple handling of grade levels is a common practice in the high school level.

Meanwhile, with regards to the distribution of respondents across the country, majority of the regions in the country was represented except for the following regions: Region I (Ilocos), Region IV-B (Mimaropa), Cordillera Autonomous Region (CAR), Region 11 (Davao Region), and Region 12

(Soccskargen). Majority of the respondents came from Negros Occidental, a province under the Region VI (Western Visayas), which was expected because the researcher came from this region and had registered 82 (52.23%) respondents. This was followed by National Capital Region (NCR) with 26 respondents (16.56%), then Region III (Central Luzon) with 18 (11.46%), Region IV-A (CALABARZON) with 13 (8.28%) and 6 (3.82%) from Region V or BICOL Region. The rest of the regions such as Regions II, VII, VIII, & X accounted the remaining 7.64% wherein there were only 1 to 5 respondents who answered the questionnaire.

### Data Analysis

For this section, quantitative research methods were used to establish the extent of the validity and reliability of the instrument, particularly on the 44-items of the TPFAS instrument that used Likert scale. Although the "practicality" of the instrument is not considered psychometric properties, its relevance is be discussed in terms of factors that determine practicality of a research instrument.

#### Content Validity

To check for the content validity of the instrument, a content validity calculation was employed using Aiken's V (1985) content-validity coefficient. The content validity calculation method proposed by Aiken is only applicable for sequential evaluation data such as the Likert rating scale (Yang, 2011). In this study, a Likert scale that measures frequency (Always, Frequently, Sometimes, Rarely, Never) was used.

#### Construct Validity

Construct validation is usually done empirically by factor analytic techniques and to perform such techniques, basic assumptions must be satisfied (Langub, 2019). Given a sample size of 157 which according to Cattell's (1978) recommendation (3:1-6:1 standard number of cases per variable ratio) is considered enough, the Kaiser-Meyer-Olkin (KMO) result of .866 that assured a sampling adequacy because 0.50 is considered suitable for factor analysis (Williams et al, 2010), and the Bartlett's Test of Sphericity is significant ( $p < .001$ ); it is therefore safe to conclude that a factor analysis is suitable to use in this study. In establishing the dimensionality of the 44-items from TPFAS, the extraction method utilized was the principal components factor analysis, while fixing the number of factors to extract at four (4) as based on the elements of the Feedback Loop Model. This analysis specifically examined the manner in which constructs were delineated within the instrument in relation to the pre-determined indicators. Additionally, a varimax rotation method with Kaiser normalization as its rotation method was also employed, suppressing small coefficients below to 0.4.

#### Reliability

The data were examined for internal consistency using Cronbach's alpha ( $\alpha$ ) and the effect of specific items on overall scale reliability. Internal consistency was measured for the entire TPFAS and for the separate factors identified using a principal components factor analysis. A reliability coefficient of .70 or higher, in which considered "acceptable" in most educational research situations (Cortina, 1993) was the basis for identifying the reliability of TPFAS.

#### Face Validity

Although face validity in principle is not considered as validity as far as measurement principles are concerned (Anastasi & Urbina, 1997), since it is more focused on the appearance and attractiveness of an instrument; the researcher used the experts' and prospect respondents' viewpoints in improving the face validity in terms of the relevance, simplicity and incomprehensibility of the meaning of words used in the items.

#### Practicality

The "heart" of research instrumentation is primarily based on one's purpose and if the developed tool addressed the intended purpose of the study, this instrument has attained the practicality or usability of the tool. According to Asaad and Hailaya (2004), to determine the practicality of an instrument, the following factors must be achieved:

- Ease of administration
- Ease of scoring/coding and decoding of results
- Ease of interpretation and application
- Low cost or economical
- Proper mechanical make-up

## RESULTS AND DISCUSSION

This section will discuss the results of the validity and reliability of TPFAS. This also presents the final items based on the data analysis and likewise, the practicality of the instrument according to the feedback shared by the respondents.

#### Content Validity

According to Retnawati (2016), a questionnaire is proved to be valid if the expert believes that the instrument measures the identified constructs. The degree of agreement among the experts regarding the importance of the item content was quantified into one coefficient (V value).

$$V = \frac{\sum s}{n(c - 1)}$$

The s pertains to the scores assigned by each rater minus the lowest score in the used category ( $s = r - l_o$ , with  $r$  = rating by an expert and  $l_o$  = the lowest possible validity rating);  $n$  is the number of raters/experts; and  $c$  is the number of categories that

raters can choose. The V value ranges from 0 to 1, and accordingly, the closer an item to 1, the better it is, because it is more relevant to the indicator.

For this study, three experts were invited to engage in the content validity testing. After checking the table recommended by Aiken (1985) (particularly on 3 number of experts/raters and using 5 number of rating categories -- see table V, Aiken, 1985), it shows that the content validity coefficient (V value) of each tested item has to be falling at 0.92 or higher to effectively reach significant standard and be considered a relevant item. The calculation results shown in Table 2 suggests that if an item did not reach a V value equal to 0.92, the item must be deleted. Consequently, the content validity coefficients (V value) of the items in this study were found to be either at 0.92 or 1.00. This means that the experts agreed for each item to be relevant and significant to the identified constructs. Therefore, the TPFAS instrument suggests a good content validity, indicating the scale is an effective measurement tool.

**Table 2**  
*Results of content validity coefficient (V value) per item using Aiken's formula*

Items	V	Items	V
1	1.00	23	1.00
2	1.00	24	1.00
3	1.00	25	1.00
4	0.92	26	1.00
5	0.92	27	0.92
6	0.92	28	1.00
7	1.00	29	1.00
8	0.92	30	1.00
9	1.00	31	1.00
10	0.92	32	1.00
11	0.92	33	1.00
12	1.00	34	1.00
13	1.00	35	1.00
14	1.00	36	1.00
15	1.00	37	1.00
16	1.00	38	1.00
17	1.00	39	1.00
18	1.00	40	1.00
19	1.00	41	1.00
20	1.00	42	1.00
21	1.00	43	1.00
22	1.00	44	1.00

After establishing the validity of the items, the TPFAS was set to be administered to science teachers for pilot testing using an online survey called Google Form.

### Construct Validity

The construct validity of TPFAS was determined using confirmatory factor analysis such as running a factor analysis on the items in the scale to determine the covariation among the items and to identify whether the patterns fit into the constructs of the Feedback Loop. The dimensionality of the 44-items was analyzed using principal components factor analysis. The Kaiser-Guttman rule was used to identify a number of factors and their components based on the data analysis. Also, no items must cross-load on more than one factor, and indicators with factor loadings less than 0.4 was excluded (Ertz, et al., 2016). As a result, 10 items were deleted from the instrument including 3 items from *Setting Goals*, 4 items from *Designing, Selecting and Adapting Tools*, 2 items from *Collecting Data*, and 1 item from *Making Inference*. These items were excluded in the final instrument because of the following reasons:

- (a) Items that implied negative statements such as: *"I find it challenging to guide my students to focus on the learning goals"* (Setting Goals), *"I find it difficult to ask "intriguing questions" that stimulate students' thinking"* (Designing, Selecting and Adapting Tools), *"I am having trouble finding an appropriate tool that fits the learning goals"* (Designing, Selecting and Adapting Tools), *"I don't see any relevance in collecting data (e.g. worksheets, lab reports, tests) to examine what my students have learned"* (Collecting Data) and *"I feel less confident in giving informative feedback"* (Making Inference) were all loaded in one factor; a misfit to the four constructs needed by the researcher and therefore decided by the researcher to discard the items despite the recoding process of negative items was done before applying the factor analysis.
- (b) Items that did not load during the factor loading which includes statements, *"I believe that students need to have an understanding of what is expected of them"* (Setting Goals), *"I prefer to ask questions that start with "why" and "how" than with "what"* (Designing Tools) and *"I do tasks that provide quick information like having them vote on the responses by raising their hands or writing the answer on the board"* (Designing Tools). Also this statement under the construct of Collecting Data, *"I pose open-ended questions that allow my students time to reflect and respond"* was excluded.
- (c) Item that cross-loaded in another factor like the item, *"I use learning progressions as a*

way of mapping out a sequence of instruction" (Setting Goals) was also removed.

the final results including the items' statements are reported in Table 3.

After eliminating the problematic items, another factor analysis was conducted on the remaining scales and

**Table 3**  
*Factor Matrix of TPFAS*

ITEMS	Factor Loading	Internal Consistency (alpha) (Over-all $\alpha = .93$ )
<i>Setting Goals</i>		<b>.819</b>
1. I clearly communicate to my students the intended learning or goals for the lesson.	.776	
2. I see to it that the learning goals are specific, measurable and easy to understand.	.553	
3. I believe that constructing learning goals that are clear can increase students' learning.	.522	
4. I use well-defined learning goals to help students take more control of their learning.	.479	
5. I inform my students before the start of the lesson on what they need to learn, demonstrate and be able to do.	.746	
6. I discuss the learning goals many times during the lesson.	.623	
7. I explore students' thinking through their common ideas, general prior knowledge and/or misconceptions.	.443	
8. I try to redirect my students' attention on what needs to be focused on.	.682	
<i>Designing, Selecting and Adapting Tools</i>		<b>.825</b>
9. I use activities based on "observable or anchoring events" (such as earthquakes, volcanic eruptions, or the cooling of coffee in a mug) to explain and elicit answers from students.	.763	
10. I can design learning tasks (such as contextualized scenarios, concept or mind maps) to capture my student's understandings.	.734	
11. I can easily select and adapt tools to assess my students' thinking.	.596	
12. I involve my students by listening and asking their ideas when developing assessment tasks.	.592	
13. I usually ask students a question with a series of follow-up questions to probe the student's ideas in depth.	.589	
14. I find it easy to develop discussion tools (such as whole-class or small-group discussions) that elicit ideas.	.575	
<i>Collecting Data</i>		<b>.855</b>
15. I engage students in small group discussions.	.496	
16. I listen to students' responses and conversation in their small group.	.589	
17. I actively engage students in generating data by listening and responding to the ideas.	.573	
18. I pay attention to my students' expressions whenever they are working on the tasks.	.582	
19. I pay attention to students' participation or their tone of voice, facial expression or do anything else to help me track how my lesson is going.	.642	
20. I can interpret students' ideas correctly when I use varied forms of data (e.g. responses during discussions, lab reports, problem sets).	.700	
21. I can utilize the assessment data of my students to adjust my teaching.	.559	
22. I provide support to students without taking over by using thought-	.503	

provoking questions, rather than explanations or direct instruction.	
23. I can identify learning patterns from the data I collected from students.	.533
24. I plan activities that require short and simple responses to yield essential information.	.487
25. I like to develop activities that make students discuss in small groups and let them share their answers in the class.	.414
<b>Making Inferences</b>	
<b>.884</b>	
26. I encourage my students to write feedback in helping other students improve their tasks.	.785
27. I engage my students to do peer assessment.	.757
28. I let students evaluate the quality of what their peers have done, objectively.	.744
29. I ask my students to do self-assessment about them after the activity.	.680
30. I involve my students in the formulation of criteria for the rubrics of peer and self-assessments.	.653
31. I analyze students' recorded ideas and pick up on themes.	.631
32. I let students create a written record of what they know and able to do.	.551
33. I use "rubrics" for both peer and self-assessments.	.527
34. I capture student's thinking by collecting their written work or jotting students' ideas on a piece of paper.	.514

### Reliability

Cronbach's alpha ( $\alpha$ ) showed high internal consistency with an overall scale reliability of 0.93 for the TPFAS instrument, based on the 157 total interpretations. The resulting TPFAS instrument exhibited internal consistency reliability and included 34 items having 50.18% of total variance explaining the variability. Moreover, reliability analysis was conducted within each subscale and exhibited internal consistency reliability (alpha) ranging from .819 to .884 for the four subscales or constructs. According to Hulin et al., (2001) as cited in Ursachi et al., (2015), this range was considered as reliable. Nevertheless, even if an item was deleted, the Cronbach's alpha values found in the tables of total-item statistics using SPSS for the overall and per subscale showed reliable values. Thus, the items in the instrument held together as an entity and as separate factors.

### Face Validity

According to the comments and suggestions of the experts and target users, to make some items more understandable, other items can be revised and improved. Some items that have the same implications to them are: the item "I apply sequences of learning goals to find out what my students have learned so far." was changed into a more explicit statement like "I use learning progressions as a way of mapping out a sequence of instruction." for the term "sequences of learning" could be an incomprehensible phrase to the teachers. Also, the item "I try to engage into their small group's discussions and ask questions." was revised into a simpler statement like "I engage students in small group discussions." and was advised to separate the item on the part of asking questions.

In addition, some commented that there were more positively stated items in the instrument and

suggested to include negatively stated items in order to check the consistency of responses.

Face validity in principle is not considered as validity as far as measurement principles are concerned (Anastasi & Urbina, 1997) however, the researcher used the experts' and prospect respondents' viewpoints in improving the face value of the instrument in terms of the relevance, simplicity and incomprehensibility of the meaning of words used in the items.

### Practicality

In terms of the usability of TPFAS instrument, which was based from the factors mentioned by Asaad and Hailaya (2004), the researcher received some respondents' remarks on the usefulness of the instrument. One responded how TPFAS reminded him to reflect on his practices of formative assessment while another used the messenger to express gratitude for the tips of applying self and peer-assessments into her classes. Another junior teacher pointed out the ease of answering the questionnaire and how practical the approach especially to the public school teachers handling large size classes. Because of technology, the ease of administration was stressed-free and economical for it was evidently accessed by other science teachers across the Philippines. Similarly, the ease of scoring, coding and decoding of results were effortlessly used by the researcher because of the Google form's feature of converting responses into the Excel sheet. Consequently, it was easier for the researcher to convert the data into the SPSS program which provided further ease of interpretation and application of the study.



## CONCLUSION

Based on the evidences provided by the reliability and validity values, it can be suggested that TPFAS or Teachers' Practices on Formative Assessment Scale reveals reliable and valid properties in determining teachers' practices in conducting formative assessment. As shown by its results, an overall reliability coefficient consistency of 0.93 indicates an acceptable standard for an instrument to be used. Likewise, the four constructs or elements that were based on the Feedback Loop Model such as *goal, tool, data and inference* also showed high internal reliability values of 0.819, 0.825, 0.855 and 0.884 respectively. This implies that the items in the instrument are dependable, consistent and stable (Colton & Covert, 2007) and with this result, it shed a new light on the literature for formative assessment practices. Though several instruments have been developed concerning teachers' practices on formative assessment, this study differs from the previous studies because of the utilization of Feedback Loop's data-driven approach on formative assessment.

Moreover, with the advantages of technology, the practicalities of administering this instrument to its target users have shown efficiency in terms of its utility and economy with respect to time, money and effort. While it has shown good psychometric properties, further studies are encouraged by adding more respondents and experts for better data analysis. Perhaps, a new study focused on science teachers from the junior high school alone, or may be senior high school or from the college and universities. The TPFAS instrument anchored on the Feedback Loop Model could be used as an alternative measure in reporting the teachers' practices on formative assessment proposed by Yan and Cheng's (2015) using their TPB model.

## REFERENCES

- Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability, and Validity of Ratings. *Educational and Psychological Measurement*, 45, 131-142.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211.
- Albuquerque, C., & Igor, D. (2014). *Grounding formative assessment in high-school chemistry classrooms: Connections between professional development and teacher practice*. (Unpublished doctoral dissertation) Michigan State University, East Lansing, Michigan.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Prentice Hall/Pearson Education.
- Asaad, A. S., & Hailaya, W. M. (2004). *Measurement and Evaluation, Concepts and Principles*, (1<sup>st</sup> ed.). Rex Book Store, Inc.
- Ash, D., & Levitt, K. (2003). Working within the zone of proximal development: Formative assessment as professional development. *Journal of Science Teacher Education*, 14(1), 1-313.
- Athanases, S. Z., & Achinstein, B. (2003). Focusing new teachers on individual and low performing students: The centrality of formative assessment in the mentor's repertoire of practice. *Teachers College Record*, 105(8), 1486-1520.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Open University Press.
- Buck, G. A., & Trauth-Nare, A. E. (2009). Preparing Teachers to Make the Formative Assessment Process Integral to Science Teaching and Learning. *J Sci Teacher Educ* 20, 475-494.
- Cattell, R. B. (1978). Matched determiners vs. factor invariance: A reply to Korth. *Multivariate Behavioral Research*, 13(4), 431-448.
- Colton, D., & Covert, R. W. (2007). *Designing and Constructing Instruments for Social Research and Evaluation*. John Wiley & Sons.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, 78(1), 98.
- Danielson, C. (2013). *The framework for teaching: Evaluation instrument*. The Danielson Group.
- Ertz, M., Karakas, F., & Sarigöllü, E. (2016). Exploring pro-environmental behaviors of consumers: An analysis of contextual factors, attitude, and behaviors. *Journal of Business Research*, 69, 3971-3980.
- Furtak, E. M., Glasser, H. M. & Wolfe, Z. M. (2016). *The Feedback Loop: Using formative assessment. Data for Science Teaching and Learning*. National Science Teachers Association. David Beacom.
- Langub, M. K. (2019). Development and validation of a scale on technology integration in physics class (STIPC) for junior high school science teachers. *European Journal of Education Studies*, 6 (2), 287-298.
- Laurillard, D. (1994). Teaching and technology: which leads? *Alt-N*, 5-7.
- Marshall, J. C., Smart, J., & Alston, D. (2016). Development and validation of Teacher Intentionality of Practice Scale (TIPS): A measure to evaluate and scaffold teacher effectiveness. *Teaching and Teacher Education*, 59, 159-168.
- National Research Council (NRC) (2001). *Classroom Assessment and the National Science Education Standards*. National Academy Press.
- Retnawati, H. (2016). Proving content validity of self-regulated learning scale (the comparison of Aiken index and expanded Gregory index). *Research and Evaluation in Education*, 2(2), 2016, 155-164
- Ruiz-Primo, M. A., & Furtak, E. M (2007). Exploring Teachers' Informal Formative Assessment Practices and Students' Understanding in the Context of Scientific Inquiry. *Journal of Research in Science Teaching*. 44(1), 57-84.

- Ursachi, G., Horodnic, I.A., & Zait, A. (2015). How reliable are measurement scales? External factors with indirect influence on reliability estimators. *Procedia. Economics and finance*, 20, 679-686.
- Williams, B., Onsmann, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine*, 8(3), 1-13.
- Yan, Z., & Cheng, E. C. (2015). Primary Teachers' attitudes, intentions and practices regarding formative assessment. *Teaching and Teacher Education*, 45, 128-136.
- Yang, W. (2011). *Applying content validity coefficient and homogeneity reliability coefficient to investigate the experiential marketing scale for leisure farms*. Corpus ID: 33921641.