Research Article

Adapting the teacher formative assessment literacy scale into Turkish: Validation and reliability study

Kemal İzci [1*], **İlhan İlter** [2], **Gökhan Izgar** [1]

[1]Necmettin Erbakan University, Faculty of Education, Department of Educational Sciences, Konya, Türkiye
[2]Kahramanmaraş Sütcü Imam University, Faculty of Education, Department of Social Studies, Kahramanmaraş, Türkiye

Abstract: In the present study, the Teacher Formative Assessment Literacy Scale (TFALS), developed by Yan and Pastore (2022a), was translated into Turkish to examine the factor structure and psychometric characteristics of the scale in a Turkish sample. Data was collected from 318 teachers, of whom 168 were males and 150 were females, working in different state schools in a city in the Central Anatolia Region of Türkiye. As the first step, the scale items were translated and back-translated by experts in both English and Turkish. Afterward, experts were presented with the scale to check for consistency and accuracy based on the feedback received. A pilot study was carried out to establish the linguistic equivalence of the scale, followed by an examination of its structural validity in a Turkish sample. To demonstrate evidence of the confirmed factor structure of TFALS with 22 items, the confirmatory factor analysis (CFA) was conducted. The CFA results demonstrated that the three-factor model of the TFALS-Turkish-version had adequate fit indices. The Cronbach's alpha coefficient of the total scale was found to be .93. The Cronbach's alpha for each of the dimensions of TFALS was .90 for the conceptual dimension (7 items), .86 for the practical dimension (8 items), and .88 for the socio-emotional dimension (7 items). Composite reliability coefficients of the dimensions ranged from .78 to .84. The results of this study provide important evidence for the validity and reliability of the Turkish version of the TFALS, confirming that it has good psychometric properties in a Turkish sample

1. INTRODUCTION

Assessment and evaluation are the essential elements of an effective instructional process. Assessment and evaluation provide data both on the impact of education policies at large and on the quality and level of learning that takes place at the classroom level. In particular, studies conducted in recent years have shown that teachers' data-based decisions in the teaching process support student success (Bennet, 2011; Brookhart, 2018; Lee et al., 2020; Li, 2016). In this context, it is important that teachers both collect data showing their students' learning as a process and result and use these data to support teaching and learning (Yan & Pastore, 2022a). Assessment and evaluation can be used for diverse purposes in the educational process.

*CONTACT: Kemal İzci ✉ kemalizci@gmail.com ▤ Necmettin Erbakan University, Faculty of Education, Department of Educational Sciences, Konya, Türkiye

Researchers classify the purposes of assessment and evaluation in education into three main categories: accountability, certification, and instructional purposes (Archer, 2017; Black & Wiliam, 2009; Brookhart & McMillan, 2020). Firstly, each country needs to monitor the success of its education policies at a broad level. Thus, educational institutions use national and international comparative assessments to illustrate the results of educational efforts for accountability purposes. Secondly, educational institutions need to provide certifications or diplomas to illustrate that an individual student has the knowledge and skills required to progress to the next level of education, to transfer to new types of schools, and to apply for a job requiring the certification. Thus, institutions use assessment and evaluation to provide certification.

Finally, at the classroom level, assessment and evaluation can be used for instructional purposes, including diagnosing students' prior knowledge and misconceptions, monitoring learning, providing feedback, and demonstrating the success of teaching and learning. When we look at the instructional purposes of assessment and evaluation, we see that they are generally classified as summative and formative (Black & Wiliam, 1998). Summative assessment focuses on determining students' level of learning in order to award them grades that certify their learning. Therefore, it is believed that the influence of summative assessment on shaping instruction and improving learning is limited. Formative assessment, on the other hand, is thought to support learning by enabling the teacher to monitor the learning process and the students to organize their own learning. For this reason, it is important that teachers understand formative assessment and use it in their teaching to facilitate effective and high-quality learning (Bennet, 2011; Black & Wiliam, 1998; Brookhart & McMillan, 2020; Izci, 2016; Shepard et al., 2017).

1.1. Theoretical Framework

1.1.1. *Formative assessment*

Researchers define formative assessment as the use of assessment processes to gather data about students' learning in order to aid their learning (Black & Wiliam, 2009; Haritage, 2010; Wylie, 2020). Teachers can use a variety of methods, such as observations, quizzes, and class discussions, in order to gain knowledge about their students' learning progress. However, the critical point in formative assessment is to use the collected information in a way to support learning. Providing effective feedback and adapting teaching to meet students' needs are some ways to use formative assessment results to facilitate students' learning. Formative assessment stands out as an important instructional approach because of its role in regulating instruction and enhancing learning (Black & Wiliam, 2009; Wylie, 2020).

When looking at the relevant literature, two approaches to formative assessment come to the fore. The first of these approaches is data-based decision-making, and the second is assessment for learning. The data-based decision-making approach involves collecting, interpreting, and using data from different sources, formally or informally, in order for students to achieve a specific learning goal (Schildkamp & Kuiper, 2010; Schildkamp et al., 2020). The students' learning levels are determined by analyzing the collected data, and tailored teaching activities are then provided to help them achieve the desired level. In this approach, the goal is the realization of learning. On the other hand, the assessment for learning approach aims to improve the quality of the learning process. Klenowski (2009) states that the purpose of assessment for learning is to aid and encourage the ongoing learning process.

What is important here is the interpretation and use of data collected through teaching activities such as dialogues, demonstrations, and observations made by students, teachers, and others. Based on this definition, the information collected about the learning process generally includes the information collected informally in the process. The effective use of this approach enhances the effectiveness of teaching and the success of students, as shown in previous studies (Pinger

et al., 2018; Shavelson, 2013). According to Bennet (2011), the success of assessment for learning depends on teachers' obtaining useful information about their students, interpreting this information, and using this interpretation for instructional decisions and feedback to students.

Although formative assessment offers significant potential, the desired effect of formative assessment depends on its comprehensive understanding and application by teachers. Previous research studies have yielded mixed results on the effect of formative assessment (Furtak et al., 2016; Kepek & Izci, 2021). Some researchers argue that the reason why formative assessment does not produce the desired effect is that teachers do not use assessment information to make data-supported instructional decisions (Bennet, 2011; Schildkamp et al., 2020). However, it has been concluded in some studies in the literature that how teachers use formative assessment is determinative of increasing learning (Black & Wiliam, 1998; Torrence, 2012). Hence, it can be posited that teachers hold a pivotal role in the achievement of formative assessment. In an investigation conducted by Cañadas and colleagues (2021), the potential gains and benefits of formative assessment were explored. This study's findings showed that implementing formative assessment techniques in educational environments could enhance students' learning outcomes, cultivate self-directed learning abilities, facilitate significant and contextual learning experiences, and encourage metacognitive processes, especially through collaborative and self-evaluation. If the competencies that teachers must have for a successful formative assessment are known, then teachers can be trained within these competencies. Upon examination of the pertinent literature, it becomes apparent that various concepts, including assessment literacy, assessment identity, and assessment expertise, are utilized to describe teachers' assessment competencies (Abell & Siegel, 2011).

These models also cover cognitive, affective, and practical competencies related to teacher assessment competencies. For example, in a literature review conducted by Gotch and French (2014), it was determined that 36 scales existed in the literature on assessment literacy, and most of these scales had low validity and reliability evidence. Assessment literacy pertains to the knowledge, skills, and techniques required by teachers to appraise the comprehension and competencies of their students, analyze the outcomes of these assessments, and apply the findings to provide constructive feedback or adapt their pedagogical strategies (Abell & Siegel, 2011; Xu & Brown, 2016). Yan and Pastore (2022a) defined the concept of formative assessment literacy based on the concept of assessment literacy. Based on the definition provided, the concept of formative assessment literacy includes the fundamental knowledge, skills, and attitudes that educators need to have in order to proficiently use the assessment process. Such a process must improve the learning experience and refine pedagogical practice.

However, it has been observed that limited studies have been conducted on teacher competencies in formative assessment literacy. In one of these studies, Schildkapm et al. (2020) analyzed 54 studies that were conducted within the scope of formative assessment. The examination led to the recommendation of a comprehensive set of skills that encompass three dimensions. The first element comprises the knowledge and skills necessary for teachers to possess in relation to formative assessment. The second element consists of social factors that relate to teachers. The last element comprises psychological factors that relate to formative assessment. According to the framework, the dimension of knowledge and skills includes various competencies, such as but not limited to data and assessment literacy mastery, proficiency in pedagogical content knowledge, effectiveness in goal-setting, provision of valuable feedback, leading meaningful in-class discussions, and expertise in utilizing information and communication technologies. The social factors dimension, on the other hand, involves the competencies of teachers to cooperate with their colleagues and their ability to involve students in the process. The psychological factors dimension consists of the competencies of attitude or belief, ownership, social pressure, and perception of control.

Although a model was proposed by Schildkapm et al. (2020) within the scope of formative assessment, only one scale that could be used to determine and improve teachers' formative assessment literacy was reached (Yan & Pastore, 2022a).

1.1.2. *Formative assessment in Türkiye*

In recent years, there has been an increased emphasis on formative assessment in Türkiye's learning and teaching process, which aligns with international developments in this area (Cañadas, 2023; Menéndez et al., 2019; Schildkamp et al., 2020). This is to ensure that student needs are met effectively. It is possible to state that formative assessment is deemed important for evaluating the successful implementation of the curriculum. In this sense, teachers can guide their students to develop meaningful learning by choosing formative assessment tasks and methods (Biggs & Tang, 2011). In addition, within the scope of teacher competencies in Türkiye, the formative assessment is also highlighted by the Ministry of National Education (MoNE). While the MoNE does not provide a distinct definition for formative assessment, the MoNE (2013) explained the assessment vision of the 2013 curriculum as "it has adopted an assessment approach aimed at identifying learning difficulties and providing continuous feedback in order to monitor and guide instruction to support meaningful learning." (MoNE, 2013, p. IV).

Similarly, the "General Competencies for the Teaching Profession" document, which was issued by the MoNE in 2017, constitutes the Turkish teacher competencies and places great emphasis on formative assessment by requiring teachers to meet the following two competencies: "1) Giving correct and constructive feedback to students and other stakeholders by taking into account the results of assessment; and 2) Undertaking a reorganization of the procedures utilized for both teaching and learning, taking into account the outcomes obtained from assessments." (MoNE, 2017, p. 15). This shows that the MoNE of Türkiye values the formative function of assessment as it emphasizes the use of assessment to elicit student learning, provide feedback, and adjust instruction within the two documents. While the beneficial impact of formative assessment on the learning and teaching process is recognized, its potential practical value depends on teachers' application of the approach (Bennet, 2011; Black & Wiliam, 2009; Schildkamp et al., 2020). Thus, it is important to focus on teachers' formative assessment literacies.

A limited number of studies that focus directly on formative assessment literacy are noticeable in Türkiye (Aras, 2019; Bayrak et al., 2019; Buldur & Hasbek, 2020; Karaman, 2017; Karaman & Karaman, 2017; Kaya et al., 2021; Yasar, 2017). One of the reasons for this is that alternative assessment and evaluation are more emphasized in MoNE documents, and, therefore, researchers focus more on alternative assessment and evaluation in their studies. However, alternative assessment and evaluation refers to the use of different and flexible techniques that are different from the traditional techniques that can be used in the assessment of students' learning as a process and a product (Şahin & Kaya, 2020). Therefore, the studies conducted in this area in Turkey mostly focused on teachers' opinions, knowledge, and frequency of using alternative assessment and evaluation techniques (Ayan & Erdemir, 2023; Şahin & Kaya, 2020). When we look at the limited studies that focus on teachers' formative assessment literacies, it is seen that these studies mostly focus on any of the social-emotional, conceptual, and practical parts of formative assessment rather than focusing on teachers' formative assessment literacies as a whole. In addition, it is seen that the related studies mostly examined the formative assessment of the teachers by collecting qualitative data.

Aras (2019) conducted a study to support the development of formative assessment practices among three preschool teachers and obtained research data through semi-structured interviews and observations. The results of the study indicate that teachers made improvements, especially in the areas of collecting data about students' learning in a planned manner, developing lesson plans according to the collected data, and using portfolios in a way that supports student

participation. In another study conducted with preschool teachers, Karaman and Karaman (2017) examined the formative assessment practices of 12 service preschool teachers. In the study, in which the data was obtained through the notes kept by the participants and open-ended questions posed by the researcher, the findings showed that the participants used observation and follow-up as an assessment tool, integrated assessment and learning processes, and used process-oriented assessments.

In their study, Bayrak et al. (2019) focused on biology teachers' ability to identify students' learning deficiencies and their plans for addressing the deficiencies. According to the data obtained through semi-structured interviews, the authors concluded that teachers mostly identified students' learning deficiencies by using verbal questions and discussions during the lesson and tried to overcome learning deficiencies by giving short repetitions of the subject and different examples. Buldur and Hasbek (2020) used the metaphor test as a data collection tool in their study, in which they examined pre-service teachers' perceptions about formative assessment through metaphors. The findings of the study showed that there were 42 metaphors produced by 127 pre-service teachers, and these metaphors were classified in the categories of "improving student learning; identifying the learning gap; re-planning the teaching process; and assessment throughout the process." It was determined that the most metaphors were in the "Improving student learning" category. Kaya et al. (2021) aimed to examine science teachers' formative assessment awareness and the effect of this awareness on formative assessment practices. The authors collected data through a form consisting of open-ended questions and classroom observations. The findings showed that the participants exhibited three different levels of formative assessment awareness and practices: naive, eclectic, and conscious. It was found that the majority of the participants had an eclectic level, which meant that they had a high level of awareness but could not reflect this awareness into practice. Yaşar (2017) examined the perceptions of pre-service science teachers about formative assessment and obtained data from 17 participants through semi-structured interviews. The findings showed that pre-service science teachers mostly knew the definition and function of formative assessment, but they were very inadequate in terms of the purposes and uses of formative assessment and adapting instruction according to the results of formative assessment. Lastly, Karaman (2017) adapted a scale for determining pre-service teachers' attitudes and intentions towards formative assessment into Turkish and collected data from 301 pre-service teachers. The results showed that pre-service teachers' attitudes towards formative assessment and behavioral control levels affect their intentions to implement formative assessment.

As mentioned above, Turkish teachers' formative assessment literacy areas were mostly addressed by using qualitative research methods and data. Only Karaman (2017) translated a quantitative measurement tool that addressed teachers' perceptions and goals of formative assessment into Turkish. Thus, there is no quantitative tool to measure Turkish teachers' formative assessment literacy. Teachers' formative assessment abilities can be measured accurately and reliably through a valid tool, which in turn can help focus on learning-related elements. This can contribute to the development of teachers' teaching competencies and aid in the creation of effective strategies for instructional planning. In this context, the present study aims to adapt the TFALS developed by Yan and Pastore (2022a) into Turkish in order to measure the formative assessment literacy of teachers.

1.2. Aim of the Present Study

Yan and Pastore (2022a) have created a measuring tool called "The Teacher Formative Assessment Literacy Scale [TFALS]," which measures the formative assessment literacy of teachers. In the process of scaling relevant factors, the studies on formative assessment and the models and scales proposed concerning assessment literacy have been employed. The TFALS comprises three dimensions: conceptual, practical, and socio-emotional. It measures teachers' formative assessment literacy based on their statements. A study carried out by Yan and Pastore

(2022a) on primary and secondary school teachers showed that the TFALS had acceptable psychometric properties. One of the reasons for conducting the present research is that formative assessment has not been adequately addressed nationally, as mentioned above. Furthermore, just one study (Karaman, 2017) employed a quantitative approach by adapting a scale to investigate preservice teachers' aims and attitudes toward formative assessment. Given the constraints of time and cost, a decision was made to forego the development of a new scale in favor of translating an established, reliable scale that has been introduced to the international literature into Turkish. Besides, adapting a scale to a new group helps researchers investigate the validity and reliability of the scale for different groups of participants, which allows them to use the scale to compare groups. Furthermore, the adaptation of a pre-existing scale to a new cultural context has the potential to enrich the literature on the cultural validity of said scale.

Formative assessment undeniably impacts teaching and learning processes in the classroom (Schildkamp et al., 2020). Formative assessment has become a policy pillar with educational significance (Van der Kleij et al., 2018) due to its potential to improve student learning (Black & Wiliam, 1998). The fact that the scale has been used in different cultures, such as Hong Kong and Italy (Yan & Pastore, 2022b), and has high psychometric properties, created a motivation for its adaptation to Turkish culture. Adaptation of the TFALS into Turkish will be beneficial in terms of determining the levels of formative assessment literacy of our teachers, meeting their formative assessment needs by identifying them, and allowing for different comparisons. The lack of any quantitative studies on teachers' formative assessment literacy, especially at the national level, is attributed to the absence of a data collection instrument that can measure teachers' formative assessment literacy levels. Therefore, adapting the scale is expected to enhance teaching competencies and facilitate the implementation of formative assessment processes in education. In addition, the use of an instrument that measures teachers' formative assessment literacy can give future teachers an idea about making decisions concerning their instructional steps and how to improve their assessment practices. This research is expected to benefit teachers, students, and researchers.

2. METHOD

2.1. Research Method

This study aimed to examine the psychometric properties of the Teacher Formative Assessment Literacy Scale (TFALS) in a sample of Turkish teachers. The study employs a survey model, which is one of the quantitative research approaches, in order to achieve its aim. For the purpose of this study, the investigation was executed through the implementation of the stages involved in the process of scale adaptation.

2.2. Study Group

The sample consists of elementary and secondary school teachers who registered at the non-thesis master's degree in educational sciences at a higher education institution in Türkiye. The sample was composed of 318 teachers, of whom 150 were females (47.16%), and 168 were males (52.83%). Participants' ages ranged from 24 to 46, with a mean age of 34 years. Convenience sampling was used to recruit participant respondents for data collection. When we examined the teaching experience of the sample, 10 (3.1%) participants reported less than five years of experience (0- 5 years), 66 (20.8%) were between 6 and 10 years, 60 (18.9%) between 11 and 15 years, 70 (22%) between 16 and 20 years, and 112 (35.2%) between 21 years or more. Participants differed according to their teaching profession fields and worked at different school levels, such as elementary and secondary schools. The distribution of participants according to their fields of expertise is shown in Table 1.

Table 1. *Distribution of participant teachers by their teaching profession fields*.

Teaching profession fields	N	%	Teaching profession fields	N	%
Elementary school teachers	38	11.95	English teachers	24	7.55
Social studies teachers	35	11.01	Physical education & sports teachers	15	4.72
Turkish teachers	34	10.69	Special education teachers	13	4.09
Science teachers	34	10.69	Religious culture & moral teachers	12	3.77
Math teachers	38	11.95	Information technology teachers	15	7.72
History teachers	15	4.72	Biology teachers	9	2.83
Geography teachers	18	5.66	Chemistry teachers	8	2.52
Physics teachers	10	3.14			

Additionally, in this study, to ensure the language validity of the scale, data were collected from an independent sample group of 25 pre-service teachers enrolled in the final year of the English language-teaching department at a university. Most of these participants were female ($N = 18$).

2.3. Data Collection Tool

2.3.1. *Teacher formative assessment literacy scale*

The TFALS was developed by Yan and Pastore (2022a) to measure the formative assessment literacy profile of teachers. The scale was developed with two randomly selected sample groups from Hong Kong and Italy. A total of 449 teachers, 295 females and 151 males, working in 12 schools in Hong Kong, participated in the administration of the scale. Of these teachers, 263 worked in elementary schools and 186 in high schools. In the Italian sample, data were collected from 309 teachers, most of whom ($N = 278$) were female, in ten selected schools in the Apulian region. Out of 309 teachers, 134 work in primary schools and 175 in high schools. It was designed based on the theoretical model of assessment literacy introduced by Pastore and Andrade (2019). Recognizing the critical role of teachers' assessment literacy, this model was developed to assess the conceptual, practical, and socio-emotional aspects of teachers' formative assessment literacy. Therefore, the original scale has three factors (dimensions) and twenty-two items based on a deductive approach. The items are designed to provide a comprehensive picture of the formative assessment literacy of teachers.

The main factor in the development of this TFALS was to develop a self-report tool that would comprehensively assess teachers' formative assessment literacy. The conceptual dimension has to do with formative assessment principles and content knowledge. The practical dimension aims to assess a teacher's formative assessment practices in order to promote learning and teaching. The socio-emotional dimension assesses a teacher's understanding of the social and emotional aspects of formative assessment. The conceptual dimension consists of seven items (items 1, 2, 3, 4, 5, 6, and 7); the practical dimension consists of eight items (items 8, 9, 10, 11, 12, 13, 14, and 15); and the socio-emotional dimension includes seven items (items 16, 17, 18, 19, 20, 21, and 22). For all items, a six-point Likert-type response scale was used (1 = strongly disagree, 6 = strongly agree). The factor structure of the original TFALS was tested with exploratory factor analysis (EFA) by Yan and Pastore (2022a). The EFA results indicated that three factors had eigenvalues greater than one that accounted. In addition, confirmatory factor analyses (CFA) were performed by the authors to determine the construct validity of the TFALS with 22 items. The CFA results showed that the three-factor structure of the original scale had good model fit [(χ^2/df = 2.370; TLI = .910, CFI =.921; RMSEA = .069). Cronbach's alpha (α) reliability coefficients was found to be for the dimensions of conceptual, practical, and socio-emotional .88, .88, and .89, respectively (Yan & Pastore, 2022a).

2.4. Adaptation Processes

In this study, language equivalence (language and cultural adaptation), construct validity, item discrimination, and Cronbach's α coefficients were examined within the framework of the psychometric properties of TFALS. To translate the TFALS, a six-step process was followed consistent with recommendations in the literature (Eremenco et al., 2005; World Health Organization [WHO], 2017).

2.4.1. *Language and culture adaptation*

In the literature, it is apparent that the steps employed for cultural adaptation are intricately described. In this study, the adaptation process was refined by the WHO through numerous studies and then reduced to the steps shown in Figure 1 (WHO, 2017).

Figure 1. *Steps to be followed in the adaptation process*.
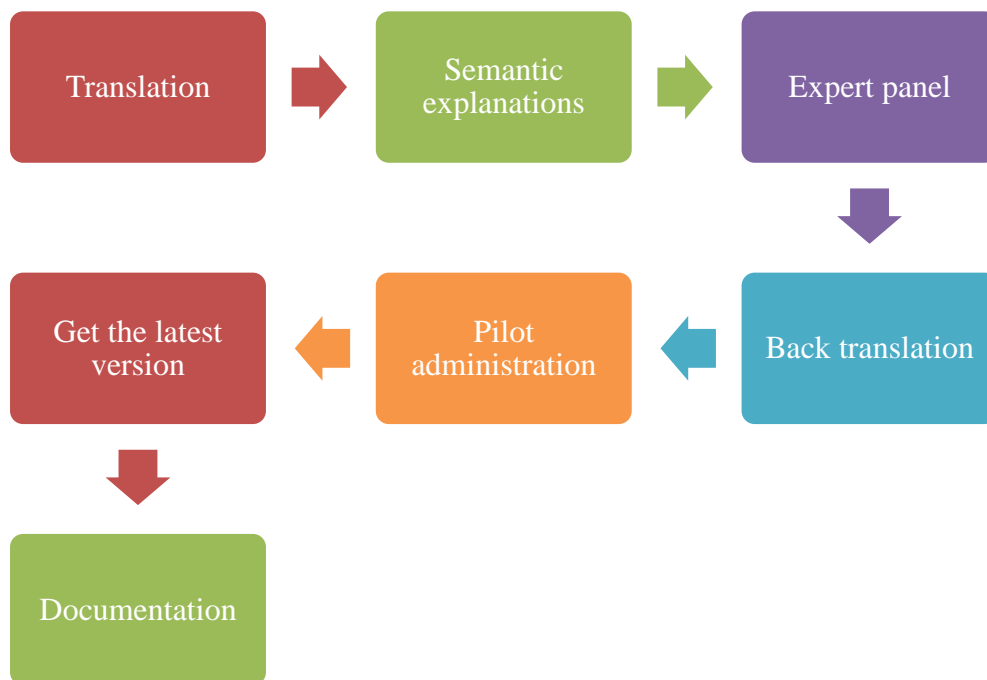
In order to adapt the original TFALS into Turkish, the authors of the scale were contacted via email to seek permission. Following the permission process, at the first stage, four faculty members who have mastery of both Turkish and English translated all of the items in the TFALS into Turkish. The experts are familiar with the culture of the original scale. During translation, they aimed for conceptual equivalence by independently translating the words or phrases in the items in the scale while staying very faithful to the original text. These experts avoided the use of any jargon when translating articles, such as technical terms, phrases, or colloquial terms that are not clearly understood. They determined whether items had changed meaning culturally (idiomatic equivalence). In particular, they examined whether it was culturally appropriate to apply each item to the target group (experimental equivalence). Thus, they evaluated whether the scale items questioned the same concepts in the new culture (conceptual equivalence) (Borsa et al., 2012).

The quality of the translation and the cultural relevance and intelligibility of the items on the scale were examined. By comparing the scale items obtained from the experts, a common Turkish form was created. In the second step, the TFALS-Turkish form was given to two experts who obtained their master's and doctoral degrees in an English-speaking country, and the experts were asked to translate the items in the Turkish form of the scale back into English. According to WHO (2017), this method, the back-translation of the scale, which passed the expert panel, should be done by an independent translator whose native language is English but

who also knows the translated language and has no knowledge of the scale. In the third step, a one-to-one comparison was made between each item's original expression and the expression that resulted from this translation. In the fourth step, after the back translation of the scale, the authors compare the two versions to identify differences between the back-translated form of the scale and the original scale. All experts reviewed all items in Turkish and English and investigated any semantic differences. In the fifth step, the TFALS-Turkish form obtained as a result of the comparison and the original scale were found to be generally equivalent, and thus the translation process was completed. In the sixth phase, the items in the scale translated into Turkish were individually evaluated by two experts who have studied Turkish language education in terms of language, intelligibility, significance, and clarity, and the scale was given its final form. According to WHO (2017), a pilot study should be conducted before claiming that a new scale is ready for data collection. Pilot implementation actually constitutes the last stage of the adaptation process. At this stage, a pilot application was made to 30–40 students for the clarity of the questions. In the selection of the sample, the aim was to reach the group that best reflected the target group. To ensure its linguistic equivalence, the TFALS-English version and the TFALS-Turkish version were administered at two-week intervals to a group of students (N = 25) from a higher education institution's English language education program. After passing all the steps described above, the authors achieved the final version of the scale. The results of the correlational analysis are presented in Table 2.

Table 2. *Correlation coefficient between the TFALS-Turkish and English versions.*

Factor	r
Factor 1	.91**
Factor 2	.85**
Factor 3	.83**
Total	.87**

**p < .01, Factor 1 = Conceptual, Factor 2 = Practical, Factor 3 = Socio-emotional

As seen in Table 2, the results revealed that all dimensions in both the TFALS-Turkish and TFALS-English versions were significantly correlated with each other (conceptual dimension r= .91, p<. 01, practical dimension r= .85, p<.01, and socio-emotional dimension r= .83, p<. 01). According to Büyüköztürk (2014), correlation coefficients between .30 and .70 indicate a moderate correlation, whereas those greater than .70 indicate a high correlation. The result indicated that linguistic equivalence was achieved between the TFALS-English and TFALS-Turkish versions.

2.5. Normal Distribution Analysis

Before conducting factor analyses, this study examined the assumption of normality using the AMOS tests for normality and outliers. To verify the existence of outliers, the Mahalanobis distance was used in the data set. Mahalanobis distance is a value used to detect the presence of extreme values that make it difficult to meet linearity and normality assumptions in regression analysis (Çokluk et al., 2012; Kline, 2011). There was no outlier that had a Mahalanobis distance score greater than the critical value, $\chi^2(10) = 29.558$, $p < .05$. In order to determine whether the univariate normal distribution assumption is met, the skewness and kurtosis coefficients for each item were examined. The normality results for the study data are shown in Table 3.

Table 3. *Normality results of the study data.*

Item/Factor	N	Skewness	Kurtosis
Factor 1: Conceptual	318	-1.548	-1.777
Item 1	318	-1.219	1.986
Item 2	318	-1.548	-1.778
Item 3	318	-1.496	-.571
Item 4	318	-.948	-1.501
Item 5	318	-1.793	-889
Item 6	318	-1.417	-1.369
Item 7	318	-1.239	-.413
Factor 2: Practical	318	-1.150	2.159
Item 8	318	-.696	-.485
Item 9	318	-.778	-1.552
Item 10	318	-1.359	1.023
Item 11	318	-1.589	-.548
Item 12	318	-1.423	.901
Item 13	318	-.974	-1.842
Item 14	318	-1.716	.844
Item 15	318	-1.158	-.548
Factor 3: Socio-emotional	318	-1.539	-1.457
Item 16	318	-1.896	.548
Item 17	318	-1.853	-1.471
Item 18	318	-.959	.963
Item 19	318	-1.112	-1.785
Item 20	318	-1.264	.875
Item 21	318	-1.489	-1.916
Item 22	318	-.895	1.750

Furthermore, normality was tested by using skewness and kurtosis coefficients. Results indicated that the skewness coefficients of the data ranged from -1.896 to -.696, and the kurtosis coefficients ranged from 1.916 to 2.159 from -2.0 to +2.0, suggesting that the data had a normal distribution (Tabachnick & Fidell, 2013; Trochim & Donnelly, 2006). In addition, the distribution of the data on a 45-degree line in the Q-Q graphs is another indicator that indicates a univariate normal distribution (Tabachnick & Fidell, 2013). The correlation coefficients calculated to determine the status of multicollinearity between items were below .90 coefficient specified by Kline (2011), indicating no multicollinearity problems.

2.6. Data Analysis

A confirmatory factor analysis (CFA) was conducted for the TFALS-Turkish version in the study sample using the AMOS 23 version. A three-factor model was created based on Yan and Pastore (2022a) with the scales used in the study. Model fit was assessed using Chi-square (χ^2), Tucker-Lewis Index (TLI), Incremental Fit Index (IFI), Comparative Fit Index (CFI), and Root-Mean-Square Error of Approximation (RMSEA (Brown, 2006; Kline, 2011). In order for the model constructed in CFA to have an acceptable fit, it should have values of $2 \leq \chi^2/df \leq 3$, $.05 \leq RMSEA \leq .08$, $.90 \leq CFI \leq .95$, $.90 \leq TLI \leq .95$ and $.90 \leq IFI \leq 95$ (Kline, 2011). The reliability of the scale was tested using Cronbach's alpha and composite reliability (CR), where a value

of α above 0.7 is considered acceptable (Büyüköztürk, 2014). Furthermore, the discriminatory power of the items was examined by calculating adjusted item-total correlations and comparing the upper 27% of the participants with the lower 27% (Erkuş, 2014; Tekindal, 2015). Therefore, item analyses were performed to estimate item-total correlation values. SPSS 22.0 program was used for composite reliability, Cronbach's alpha, construct validity and item analysis of TFALS.

3. FINDINGS

3.1. Factor structure of the TFALS

Researchers have suggested that when adapting an instrument developed in a culture, whose factor structure has been determined, to another culture, the adaptation should be done using CFA. It is also suggested that the convergent validity of the scale should be tested with CFA (Çokluk et al., 212; Seçer, 2015). Construct validity is a crucial aspect of data analysis that examines the extent to which a measure purports to measure or whether the factorial structure of the measurement tool is valid (Byrne, 2001). To examine the construct, the three-factor, 22-item structure of the TFALS-Turkish version, a CFA was performed by using the maximum likelihood estimation. The initial model included three latent variables with 25 items. Results of CFA showed that the three-factor model was not an acceptable level for close model fit χ^2 value (2.771, $p < .05$), but other fit indices (TLI = .88, IFI = .89, CFI = .89, RMSEA = .078) were within the acceptable range (Kline, 2011).

In order to improve the model, modification indices were tested among the items on the scale thought to contribute significantly to model fit. According to the experts, when conducting modifications, attention should be paid to the theoretical rationale and the items with associated error terms under the same factor (Çokluk et al., 2016; Karagöz, 2016). In this respect, modifications were conducted between the error terms (item1 & item2) of two items in the conceptual subscale of the scale, between which a latent relationship could be accepted and which was thought to contribute significantly to the model fit. Both modifications were made on items of the same size and assumed to measure similar phenomena. As a result of the renewed CFA after the modification suggestions that were provided to contribute to the model were processed, it was found that there were significant improvements in the goodness of fit indices (see Table 4) of the three-factor model of the Turkish version of the TFALS that met the adequate data fit values and were confirmed (χ^2/df = 2.379; IFI= .923; TLI= .912, CFI= .922; RMSEA= .066; SRMR= .036). The standardized factor loadings for the constructs of conceptual dimension were between .61 and .77, for practical dimension between .75 and .76 and for socio-emotional dimension between .77, and .82, which were similar to TFALS–English version (.78, .79, and .88, respectively).

Table 4. *The goodness-of fit indices and CFA results.*

Fit Indices	Perfect Fit	Acceptable Fit	Model fit indices (Yan & Pastore (2022a)	Model fit indices (Turkish version)
χ^2/df	$0 \leq \chi^2/df \leq 2$	$2 \leq \chi^2/df \leq 3$	2.370	2.379
RMSEA	$0 \leq RMSEA \leq .05$	$.05 \leq RMSEA \leq .08$	.069	.066
CFI	$.95 \leq CFI \leq 1$	$.90 \leq CFI \leq .95$	.92	.92
TLI	$.95 \leq TLI \leq 1$	$.90 \leq TLI \leq .95$	.91	.91
IFI	$.95 \leq IFI \leq 1$	$.90 \leq IFI \leq .95$	--	.92

3.2. Correlations between dimensions of the TFALS

Pearson correlation analysis was used to determine the relationship between the dimensions of the TFALS. The results of the descriptive and correlational analyses are presented in Table 5.

Table 5. *Correlations between dimensions of TFALS-Turkish version.*

	M	*SD*	Factor Correlation		
			1	2	3
Conceptual	4.96	.795	1.00		
Practical	5.02	.778	.780**	1.00	
Socio-Emotional	5.11	.819	.762**	.840**	1.00

**p< .01

As seen in Table 5, all correlations between variables were positively significant (*p*<.01). The correlations between the three dimensions of the TFALS ranged from .762 to .840. Büyüköztürk (2014) stated that the correlation between .00-.30 is low, between .30-.70 is moderate, and between .70-1.0 is high. Based on these parameters, it can be said that there were high correlations between the dimensions of TFALS. Furthermore, according to the correlation matrix for the items TFALS, all 22 items showed positive, significant, and above .30 correlations with the total score of the scale, and the level of correlation varies between medium and high.

3.3. Reliability Analysis

Cronbach's alpha (α) and composite reliability (CR) coefficients for each dimension of TFALS were used to assess the reliability of the TFALS-Turkish version. The results are shown in Table 6.

Table 6. *Reliability coefficients of the scale for the 22-item TFALS-Turkish version.*

	Number of items	Cronbach's α	Composite Reliability
Conceptual	7	.90	.84
Practical	8	.86	.81
Socio-emotional	7	.88	.78
Total	22	.93	.85

The Cronbach's alpha coefficients found for each of the dimensions of the TFALS were .90 for the conceptual (7 items), .86 for the practical (8 items), .88 for the socio-emotional (7 items), and .93 for the total scale. The CR coefficients of conceptual, practical, and socio-emotional dimensions were .84, .81, and .78, respectively, and .85 for the total scale. Fraenkel et al. (2012) and Hair et al. (2010) noted that measurements of reliability coefficients of .70 and above are generally agreed upon as an acceptable value. Additionally, George and Mallery (2003) suggested a tiered approach for the range of Cronbach's alpha consisting of the following: ≥0.9-excellent, ≥0.8-good, ≥0.7-acceptable, ≥0.6. All reliability coefficients are above.72 provides additional evidence for the reliability of the TFALS-Turkish version. Figure 2 shows the DFA results for the Turkish version of TFALS.

Figure 2. *The confirmatory factor analysis model tested with the TFALS-Turkish version.*
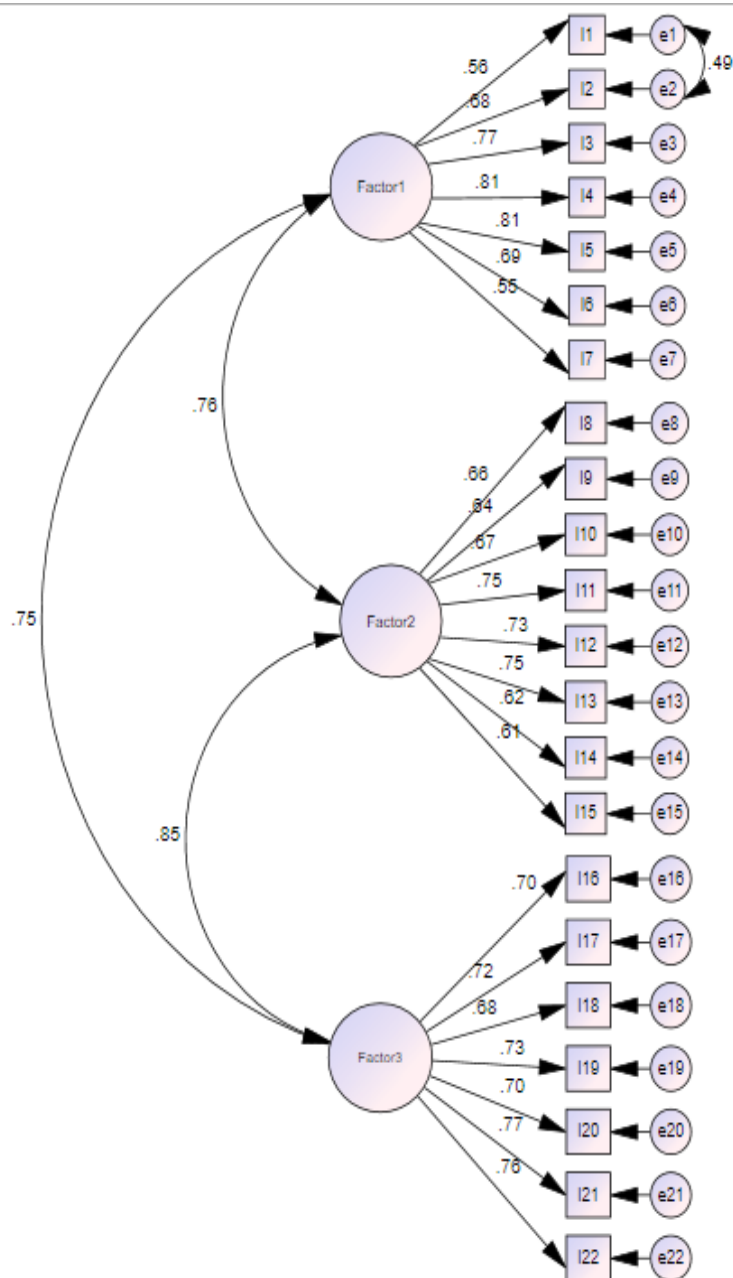
Note. TFALS= Teacher Formative Assessment Literacy Scale (TFALS) Factor 1: Conceptual, Factor 2: Practical, Factor 3: Socio-Emotional

3.4. Item Analysis

In this study, the discriminatory power of the items was examined by comparing the participants' lower 27% and upper 27% and calculating adjusted item-total correlations. The lower and upper groups for each subscale of the TFALS must be determined by the scales formed by the multidimensional structure (Büyüköztürk, 2014). A t-test was used to determine whether there is a significant difference between the upper 27% and lower 27% of the group in terms of the subscale of the TFALS-Turkish version. Table 7 displays the item-total correlations and the lower-upper 27% t-test results.

Table 7. *Results of analysis for the 22-item TFALS.*

Items	Corrected Item Total Correlations (*r*)	*M*	*SD*	Lower -Upper 27% *t*-Test
Factor 1: Conceptual				
1	.78	4.62	1.054	15.48[*]
2	.72	4.81	.978	13.95[*]
3	.65	5.18	.962	11.78[*]
4	.52	5.13	.993	118.41[*]
5	.49	5.11	.979	17.74[*]
6	.79	5.02	.981	10.12[*]
7	.68	4.89	.908	13.89[*]
Factor 2: Practical				
8	.67	4.82	.977	14.75[*]
9	.77	4.85	1.021	5.69[*]
10	.58	5.03	.972	17.26[*]
11	.70	5.04	.977	18.47[*]
12	.78	5.18	.969	21.34[*]
13	.65	4.99	.884	19.47[*]
14	.74	5.09	.961	15.03[*]
15	.70	5.16	.978	17.85[*]
Factor 3: Socio-emotional				
16	.63	5.04	.933	20.24[*]
17	.57	5.04	1.068	14.96[*]
18	.75	4.96	1.037	22.23[*]
19	.71	5.13	.937	16.45[*]
20	.78	5.24	1.007	10.99[*]
21	.74	5.19	.933	20.45[*]
22	.62	5.16	.906	17.36[*]

[*]*p< .01*

Taking 27% as a cutoff value (lower and upper groups), the results showed t values for the difference between the upper 27% and lower 27% of the participants ranged from 10.54 to 14.55 for the conceptual dimension, between 14.85 and 14.25 for the practical dimension, and between 13.25 and 12.45 for the socio-emotional dimension. T-test values were significant for all items according to the comparison result between the participants' lower 27% and the upper 27%. Significant t-values in comparisons between the lower and upper groups of the participants were accepted as evidence of the items' discriminatory power (Erkuş, 2014). Table 7 also presents that item-total correlations ranged from .47 to .57 for the conceptual dimension, .54 to .62 practical dimension, and .37 to .50 socio-emotional dimension. Items with .30 and above coefficients are considered to have sufficient discriminatory power when interpreting item-total correlations (Büyüköztürk, 2014; Erkuş, 2014). The results indicate that all items met this requirement. Therefore, all items in the TFALS had discriminatory power, according to these findings.

3.5. Scoring of the Scale

The TFALS-Turkish version consists of 22 items total with three dimensions. It included 7 items in the conceptual dimension, 8 items in the practical dimension, and 7 items in the socio-emotional dimension. There are no reverse items among the items that constitute the scale. The

scores that can be obtained from the scale range from 22 to 132 points. Higher scores indicate higher levels of teachers' formative assessment literacy.

4. DISCUSSION and CONCLUSION

Assessment is an important element of the teaching process in terms of the quality of education and training services. Accurate and effective feedback can be given to students during the assessment and evaluation process, and more effective guidance can be provided. By doing this, variables that result from teaching methods, learning environments, course contents, and other factors that negatively affect learning can be identified and corrected (Yılmaz, 1998). In this context, the effectiveness of assessment activities today is related to whether they aim to improve the learning of students rather than reach a conclusion about the success or failure of students by using letter or numerical grades (National Council of Teachers of Mathematics [NCTM], 2023). At this point, the issue of formative assessment, which aims to support the learning of the students, gains importance in increasing the quality of educational services. Although early studies aimed to determine the features of formative assessments compared to other assessment activities, recent research has focused on how frequently teachers utilize formative assessment and its impact on students' learning outcomes (Furtak et al., 2016; Pinger et al., 2018; Yan & Pastore, 2022a). Furthermore, there is a requirement for research into scale development aimed at evaluating teachers' formative assessment literacy. Therefore, Yan and Pastore (2022a) devised a self-reported scale for evaluating teachers' formative assessment literacy in response to this demand. The purpose of the study was to translate and evaluate the psychometric properties of the TFALS developed by Yan and Pastore (2022a). When an instrument is translated from one language or dialect into another, reliability and validity studies should be conducted for the intended use across linguistic groups (Geisinger, 1994). Accordingly, in this study, initially, a language equivalence study was carried out while adapting the original scale into Turkish. Experts in educational sciences and the English language translated the scale into Turkish.

The resulting Turkish form was translated back into English using the back-translation method, and expert advice was provided. Adaptation studies have shown that the items in the English and Turkish versions of the TFALS are highly correlated with each other. A common source of validation information on scale structure is the factor analysis technique applied to test data. Factor analysis techniques are the most frequently used procedures for evaluating tests adapted to linguistically diverse populations (Geisinger, 1994). In this study, after the linguistic equivalence study of the scale, CFA was used to evaluate the three-factor model of the TFALS-Turkish version with 22 items for a Turkish teacher sample. The CFA results indicated that fit indices showed that the model was not within acceptable ranges. In order to improve the model, modification indices were examined among the items that were thought to contribute significantly to the model fit, and modifications were made between the error terms of two items (item 1 and item 2) in the conceptual sub-dimension of the scale. The results of the CFA repeated after the modifications showed that the model fit indices were within the acceptable limits of goodness of fit (Kline, 2011). The CFA results support the three-factor structure of the TFALS-Turkish version. The factor loadings of the items in the scale ranged from .61 to .82. According to the experts, factor loadings of .45 or more are a good measure of selection. (Büyüköztürk, 2014; Tabachnick & Fidell, 2013). Accordingly, it can be said that the three-factor structure and item factor loadings of the TFALS-Turkish version were found to be sufficient. Therefore, the results suggest that the Turkish version of the scale is structurally reliable and can be used in a Turkish sample.

The reliability of the Turkish-adapted scale was determined by calculating Cronbach's alpha and the composite reliability coefficient. For the Turkish version of the scale, the Cronbach's alpha for each of the dimensions of TFALS were .90 for the conceptual dimension (7 items), .86 for the practical dimension (8 items), and .88 for the socio-emotional dimension (7 items). In

addition, the composite reliability values of the dimensions in the adapted scale were found to be greater than .70. In the original form of the scale, Cronbach's alpha was .88 for the conceptual dimension, .88 for the practical dimension, .89 for the socio-emotional dimension, and .86 for the whole scale (Yan & Pastore, 2022a). The internal consistency coefficient obtained is similar to the results of previous research on the reliability of TFALS (Yan & Pastore, 2022a; Yan & King, 2023; Yan et al., 2022). At this point, it is seen that the reliability coefficients of the original scale and the Turkish-adapted scale are close to each other. According to Geisinger (1994), there may be differences in the understanding of the scale due to cultural differences in scale adaptation studies. Therefore, there may be differences in reliability values. Previous research has recommended that for a measure to be reliable in scale development or adaptation studies, the internal consistency and reliability coefficients should exceed .70 (Büyüköztürk, 2014; Fraenkel & Wallend, 2006; Robinson et al., 1999; Tezbaşaran, 1997).

The findings suggest that the internal consistency and composite reliability coefficients of the Turkish version of the scale are satisfactory. Currently, based on the evidence of validity and reliability, it is appropriate to claim that the Turkish-adapted version of the scale can be employed to assess teachers' formative assessment literacy. To evaluate whether the items on the scale measure the targeted attributes, item-total score correlations and mean scores collected from the upper and lower 27% groups were computed and compared according to Büyüköztürk (2014). On the scale, the differences observed in the item averages of the lower and upper 27% groups were found to be significant at the $p<.001$ level. These results provided evidence for the discrimination of the items in the scale (Erkuş, 2014). Item-total score correlations were found to vary, ranging from .55 to .75. The correlation values between the dimensions of the scale ranged from .32 to .60 and there were significant positive relationships ($p<.05$) between the dimensions of the scale (Erkuş, 2014). On the scale, the differences observed in the item averages of the lower and upper 27% groups were found to be significant at the $p<.001$ level. This result is evidence for the discrimination of the items in the scale (Erkuş, 2014).

4.1. Recommendations and Contributions to Education

All results showed that the multidimensional structure of the original TFALS had appropriate psychometric properties. The validity and reliability analyses of this scale adapted into Turkish were carried out with a systematic approach. However, there is a limitation in terms of research results that must be taken into consideration for future research. This limitation is that the participants are teachers who have a non-thesis master's degree in the field of education. In terms of the generalizability of the results of the study, validity and reliability analyses can be conducted on the data to be collected from a sample group consisting of teachers who have not undertaken postgraduate education in Türkiye. Although there is such a limitation in the study, the findings provided evidence that the TFALS adapted into Turkish has good reliability and validity. Consequently, this study translated the TFALS into Turkish and tested its reliability and validity among Turkish teachers. It can be said that the adapted scale is suitable for measuring the literacy levels of teachers from different teaching fields in Türkiye for the practice of formative assessment. Examining teachers' assessment literacy is one of the most important issues in education and teaching practice today. There is convincing evidence in the literature that assessment, especially formative assessment, is an important leverage point for improving students' academic outcomes (Andrade & Heritage, 2018; De Simone, 2009). In this direction, the adaptation of TFALS into Turkish may be informative in terms of teacher education and teacher practice. In addition, as far as we know, this study presents the first attempt to translate the TFALS into Turkish and test its reliability and validity among a group of teachers working in Türkiye. In other words, to date, there is no inductively developed and psychometrically tested instrument that measures the multidimensional construct of teachers' formative assessment literacy in Türkiye. The Turkish version of the scale is presented in the Appendix.

The scale adapted into Turkish can actually provide researchers, teacher educators, school administrators, and policymakers with important ideas about teachers' profiles (strengths and weaknesses) in formative assessment. In this sense, it may be possible to support teachers in developing their formative assessment literacies through the data to be collected through the scale. The adapted scale can also be used to evaluate the effectiveness of teacher education programs or interventions to increase teachers' formative assessment literacy (Yan & Pastore, 2022a). Furthermore, the three-dimensional formative assessment literacy model underpinning TFALS serves critical dimensions for the effectiveness of teaching practice and the improvement of student development. In future studies, teachers' formative assessment literacy can be measured using TFALS. Again, by using TFALS, researchers can get an idea about teachers' competencies related to formative assessment in teaching processes and how they use formative assessment. This is because the scale provides a multidimensional perspective and enables the complex structure of formative assessment literacy to be seen. In this context, the TFALS, adapted into Turkish, is expected to fill an important gap in the existing literature. Both the original form of the scale and the adaptation study, which were applied to teachers, can give researchers an idea about the efficiency and excellence of teaching services. The use of this scale in different studies and with different samples will also help to better understand its psychometric properties.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors. **Ethics Committee Number**: Bartın University, E-23688910-050.01.04-2300023164

Contribution of Authors

Kemal İzci: Literature review**,** Investigation, Data collection, Resources, and Writing-original draft. **İlhan İlter**: Methodology, Supervision, Statistical analysis, Validation, and Writing-original draft. **Gökhan Izgar**: Data collection, Data Interpretation, Resources, Writing-original draf.

Orcid

Kemal İzci https://orcid.org/0000-0002-4228-8845
İlhan İlter https://orcid.org/0000-0002-1473-7172
Gökhan Izgar https://orcid.org/0000-0002-6835-9701

REFERENCES

Andrade, H.L., & Heritage, M. (2018). *Using formative assessment to enhance learning. achievement and academic self-regulation*. Routledge.

Aras, S. (2019). Improving early childhood teachers' formative assessment practices: Transformative role of collaborative action research. *International Journal of Curriculum and Instructional Studies, 9*(2), 221-240.

Archer, E. (2017). The assessment purpose triangle: Balancing the purposes of educational assessment. *Frontiers in Education, 2,* 1-17.

Ayan, A.D., & Erdemir, N. (2023). EFL teachers' perceptions of automated written corrective feedback and Grammarly. *Ahmet Keleşoğlu Eğitim Fakültesi Dergisi (AKEF) Dergisi, 5*(3), 1183-1198.

Bayrak, N., Çalık, M., & Doğan, S. (2019). Biology teachers' practice of formative assessment: A case of the identifying learning gap element. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi, 46*, 132-149.

Bennett, R.E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5-25.

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university.* Open University Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5(*1), 7–74.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31.

Borsa, J.C., Damásio, B.F., & Bandeira, D.R. (2012). Cross-cultural adaptation and validation of psychological instruments: Some considerations. *Paidéia, 22*(53), 423–432.

Brookhart, S.M., & McMillan, J.H. (2020). *Classroom assessment and educational assessment. (p. 1-9).* Routledge.

Buldur, S., & Hasbek, G. (2020). Pre-service teachers' perceptions about formative assessment: A metaphorical study. *Inonu University Journal of the Faculty of Education, 21*(2), 595-615.

Büyüköztürk, Ş. (2014). *Sosyal bilimler için veri analizi el kitabı [Data analysis notebook for social sciences].* Pegem Akademik Yayın.

Byrne, B.M. (2001). Structural equation modeling with AMOS, EQS, and LISREL: Comparative approaches to testing for the factorial validity of a measuring instrument. *International Journal of Testing, 1*(1), 55–86.

Brookhart, S.M. (2018). Summative and formative feedback. In A. Lipnevich & J. Smith (Eds.), *The Cambridge handbook of instructional feedback (pp. 52–78).* Cambridge University Press. https://doi.org/10.1017/9781316832134.005

Cañadas, L. (2023). Contribution of formative assessment for developing teaching competences in teacher education. *European Journal of Teacher Education, 46*(3), 516-532.

Cañadas, L., Santos-Pastor, M.L., & Ruíz, P. (2021). Percepción del impacto de la evaluación formativa en las competencias profesionales durante la formación inicial del profesorado [Perception of impact of formative assessment on professional competences during initial teacher training]. *Revista Electrónica De Investigación Educativa 23*: 1–14. e03.

Çokluk, Ö, Şekercioğlu G, Büyüköztürk Ş. (2012). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları [Multivariate statistics for social sciences: SPSS and LISREL applications].* Pegem Akademi Yay.

Desimone, L. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181–199.

Eremenco, S.L., Cella, D., & Arnold, B.J. (2005). A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Evaluation & the Health Professions, 28*(2), 212–232.

Erkuş, A. (2014), *Psikolojide ölçme ve ölçek geliştirme-I: Temel kavramlar ve işlemler [Measurement and scale development in psychology-I: Basic concepts and processes*]. Pegem Akademi Yay.

Fraenkel, J.R., Wallen, N.E., & Hyun, H.H. (2012). *How to design and evaluate research in education* (Vol. 7, p. 429). McGraw-hill.

Fraenkel, J.R., & Wallen, N. (2006). *How to design and evaluate research in education.* McGraw-Hill.

Furtak, E.M., Kiemer, K., Circi, R.K., Swanson, R., de León, V., Morrison, D., & Heredia, S. C. (2016). Teachers' formative assessment abilities and their relationship to student learning: Findings from a four-year intervention study. *Instructional Science, 44*, 267-291.

Geisinger, K.F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment, 6*(4), 304-312. https://doi.org/10.1037/1040-3590.6.4.304

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Allyn & Bacon

Gotch, C.M., & French, B.F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues and Practice, 33*(2), 14–18.

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2010). *Multivariate data analysis* (7th Ed). Pearson.

Heritage, M. (2010). *Formative assessment: Making it happen in the classroom.* Corwin Press.

Izci, K. (2016). Internal and external factors affecting teachers' adoption of formative assessment to support learning. *International Journal of Educational and Pedagogical Sciences, 10*(8), 2800–2807.

Karagöz, Y. (2016). *SPSS 23 ve AMOS 23 uygulamalı istatistiksel analizler* [SPSS 23 and AMOS 23 applied statistical analysis]. Nobel Akademik Yay.

Karaman, O. (2017). Adapting of formative assessment attitude and intention scale for pre-service teachers and a structural equation modelling. *Inonu University Journal of the Faculty of Education, 18*(3), 181-131.

Karaman, P., & Karaman, A. (2017). Investigating teacher candidates' use of formative assessment: Action research model. *Kastamonu Eğitim Dergisi, 25*(6), 2385-2400.

Kaya, G., Atasoy, V., Candan-Helvacı, S., & Pektaş, M. (2021). The role of science teachers' awareness in their classroom practice of formative assessment. *Education and Science, 46*(205), 335-357.

Kepek, T., & Izci, K. (2021). Examining the effect of alternative assessment methods on academic achievement and student opinions about these methods. *Necmettin Erbakan Üniversitesi Ereğli Eğitim Fakültesi Dergisi, 3*(1), 34-50.

Klenowski, V. (2009). Assessment for learning revisited: An Asian-Pacific perspective. *Assessment in Education: Principles, Policy and Practice, 16*(3), 263-268.

Kline, R.B. (2011). *Principles and practice of structural equation modeling.* The Guilford Press.

Li, H. (2016). How is formative assessment related to students' reading achievement? Findings from PISA 2009. *Assessment in Education: Principles, Policy & Practice, 23*(4), 473–494. https://doi.org/10.1080/0969594X.2016.1139543

Lee, H., Chung, H. Q., Zhang, Y., Abedi, J., & Warschauer, M. (2020). The effectiveness and features of formative assessment in US K-12 education: A systematic review. *Applied Measurement in Education, 33*(2), 124-140.

MoNE (2013). İlköğretim fen ve teknoloji dersi öğretim programi [Primary school science and technology curricula). Milli Eğitim Bakanlığı. Ankara.

MoNE (2017). Öğretmenlik mesleği genel yeterlikleri. [General competencies of the teaching profession] Öğretmen Yetiştirme ve Geliştirme Genel Müdürlüğü. Ankara. https://oygm.meb.gov.tr/dosyalar/StPrg/Ogretmenlik_Meslegi_Genel Yeterlikleri.pdf

Menéndez, I.Y.C., Napa, M.A.C., Moreira, M.L.M., & Zambrano, G.G.V. (2019). The importance of formative assessment in the learning teaching process. *International Journal of Social Sciences and Humanities, 3*(2), 238-249.

NCTM (2023). Using classroom assessment to improve student learning: Math problems aligned with nctm and common core state standards. https://www.nctm.org/store/Products/Using-Classroom-Assessment-to-Improve-Student-learning--Math-Problems-Aligned-with-NCTM-and-Common-Core-State-Standards/

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

Pinger, P., Rakoczy, K., Besser, M., & Klieme, E. (2018). Implementation of formative assessment–effects of quality of programme delivery on students' mathematics achievement and interest. *Assessment in Education: Principles, Policy & Practice, 25*(2), 160-182.

Robinson, J.P., Shaver, P.R., & Wrightsman, L.S. (1999). *Measures of political attitudes*. Academic Press.

Seçer İ. (2015). *Psikolojik test geliştirme ve uyarlama süreci: SPSS ve LISREL uygulamaları [Psychological test development and adaptation process: SPSS and LISREL applications].* Anı Yayıncılık.

Schildkamp, K., & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education, 26*(3), 482-496.

Schildkamp, K., van der Kleij, F.M., Heitink, M.C., Kippers, W.B., & Veldkamp, B.P. (2020). Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International Journal of Educational Research, 103*, 101602. https://doi.org/10.1016/j.ijer.2020.101602

Shavelson, R.J. (2013) On an approach to testing and modeling competence. *Educational Psychologist, 48*(2), 73-86. https://doi.org/10.1080/00461520.2013.779483

Shepard, L.A., Penuel, W.R., & Davidson, K.L. (2017). Design principles for new systems of assessment. *Phi Delta Kappan, 98(6), 47-52.*

Şahin, Ç., & Kaya, G. (2020). Alternatif ölçme değerlendirme ile ilgili yapılan araştırmaların incelenmesi: Bir içerik analizi [A review of the research on alternative assessment evaluation: A content analysis]. *Nevşehir Hacı Bektaş Veli University Journal of ISS, 10*(2), 798-812.

Tabachnick, B.G., & Fidell, L.S. (2013). *Using multivariate statistics.* (6ᵗʰ Ed). Pearson.

Tekindal, S. (2015). *Duyuşsal özelliklerin ölçülmesi için araç oluşturma [Creating a tool for measuring affective traits].* Pegem Akademi Yayın.

Tezbaşaran, A. (1997). *Likert tipi ölçek geliştirme kılavuzu*. Türk Psikologlar Derneği Yayın.

Torrance, H. (2012). Formative assessment at the crossroads: Conformative, deformative and transformative assessment. *Oxford Review of Education, 38*(3), 323-342.

Trochim, W.M., & Donnelly, J.P. (2006). *The research methods knowledge base* (3ʳᵈ ed). Atomic Dog.

Van der Kleij, F.M., Cumming, J.J., & Looney, A. (2018). Policy expectations and support for teacher formative assessment in Australian education reform. *Assessment in Education: Principles, Policy & Practice, 25*, 620–637.

World Health Organization (WHO) (2017). Process of translation and adaptation of instruments. Retrieved from http://www.who.int/substance_abuse/research_tools/translation/en/

Wylie, E.C. (2020). Observing formative assessment practice: Learning lessons through alidation. *Educational Assessment, 25*(4), 251–258.

Xu, Y., & Brown, G.T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education, 58*(1), 149-162.

Yan, Z., Chiu, M.M., & Cheng, E.C.K. (2022). Predicting teachers' formative assessment practices: Teacher personal and contextual factors. *Teaching and Teacher Education, 114,* 103718. https://doi.org/10.1016/j.tate.2022.103718

Yan, Z., & King, R.B. (2023). Assessment is contagious: The social contagion of formative assessment practices and self-efficacy among teachers. *Assessment in Education: Principles, Policy & Practice, 30*(2), 130–150.

Yan, Z., & Pastore, S. (2022a). Are teachers literate in formative assessment? The development and validation of the Teacher Formative Assessment Literacy Scale. *Studies in Educational Evaluation, 74*, 101183. https://doi.org/10.1016/j.stueduc.2022.101183.

Yan, Z., & Pastore, S. (2022b). Assessing teachers' strategies in formative assessment: The teacher formative assessment practice scale. *Journal of Psychoeducational Assessment, 40*(5), 592-604. https://doi.org/10.1177/0734282922107.

Yaşar, M.D. (2017). Prospective science teachers' perception related to formative assessment approaches in Turkey. *Journal of Education and Training Studies, 5*(4), 29-43.

Yılmaz, H. (1998). *Eğitimde ölçme ve değerlendirme [Assessment in education].* Mikro Yayın.

APPENDIX: The Teacher Formative Assessment Literacy Scale-Turkish Version

Öğretmen İzlemeye Dayalı Değerlendirme (Formative Assessment) Okuryazarlığı Ölçeği Türkçe Versiyonu

Aşağıdaki ifadelere ne derece katılıp/katılmadığınızı belirtiniz.	Kesinlikle katılmıyorum	Katılmıyorum	Kısmen katılmıyorum	Kısmen katılıyorum	Katılıyorum	Kesinlikle katılıyorum
1. İzlemeye dayalı değerlendirmenin mantığını/ gerekçelerini açıklayabilirim.						
2. Öğrencilerin öğrenme ihtiyaçlarının izlemeye dayalı değerlendirme yoluyla belirlenebileceğini biliyorum.						
3. Değerlendirme etkinliklerinin öğrenme hedefleriyle örtüşmesi gerektiğini düşünüyorum.						
4. İzlemeye dayalı değerlendirme etkinliklerinin öğrencilerin öğrenme düzeylerine yönelik geri bildirim sunması gerektiğini düşünüyorum.						
5. Öğrencilerin öğrenme ihtiyaçlarını belirlemek adına izlemeye dayalı değerlendirme sonuçlarının öğretmenler için yararlı olduğunu biliyorum.						
6. Öğrenmenin arttırılması için öğrencilerin izlemeye dayalı değerlendirmeye katılmaları gerektiğini düşünüyorum.						
7. Öğrencilerin öğrendiklerini göstermelerine olanak sağlayacak çeşitli değerlendirme yöntemlerini biliyorum.						
8. Öğrencilerin öğrendiklerini göstermelerine olanak tanıyan çeşitli değerlendirme yöntemlerini kullanırım.						
9. Öğrencilere akran değerlendirme süreçlerine katılmayı öğretiyorum.						
10. Öğrencilerin öz değerlendirme becerilerini geliştirmelerine yardımcı olurum.						
11. Öğrencilere sunmuş olduğum geri bildirim bilgilerini kullanmalarını sağlarım.						
12. Değerlendirme sonuçlarına göre, öğrencilere o konuyu nasıl daha iyi öğrenebileceklerini gösteririm.						
13. Öğrencilere, öğrenmelerini geliştirmek amacıyla değerlendirme sonuçlarını kullanmalarını öğretirim.						
14. Kafa karışıklığı oluştuğunda, öğrenciler için değerlendirme amaçlarını netleştiririm.						
15. Değerlendirme kriterlerini /ölçütlerini öğrencilerle paylaşırım.						
16. Öğretmenler ve öğrenciler arasında izlemeye dayalı değerlendirmeye ilişkin ortak bir anlayış oluşturma ihtiyacının farkındayım.						
17. Öğrencilerin değerlendirmelere verdikleri duygusal tepkileri dikkate alırım.						
18. Öğrencilerin değerlerinin, inançlarının ve tutumlarının izlemeye dayalı değerlendirmeyi kullanmalarını etkilediğinin farkındayım.						
19. Değerlendirme sürecindeki geri bildirim bilgilerinin öğrencilerin öğrenme motivasyonunu nasıl etkilediğinin farkındayım.						
20. İzlemeye dayalı değerlendirmenin adalet ve öğrenci mahremiyeti gibi etik yönlerine duyarlıyım.						
21. İzlemeye dayalı değerlendirme sürecinde öğrencilerin iyi oluşlarını sağlamak için sorumluluklarımın farkındayım.						
22. Öğrencilerin izlemeye dayalı değerlendirme uygulamalarından yararlanma hakkına sahip olduğunun bilincindeyim.						