Yaşadıkça Eğitim, Cilt 37, Sayı 1, Yıl 2023, s.253-268
Journal of Education for Life, Volume 37, Issue 1, Year 2023, pp.253-268
DOI: 10.33308/26674874.2023371507

A Critical Point About Early Childhood Assessments: Validity and Reliability Issues in Teachers' Formative Assessment

Elif BULDU[1], Çağla ÖNEREN ŞENDİL[2]

Abstract: In the current study, Turkish early childhood teachers' self-reported beliefs and practices on ensuring reliability and validity on their formative assessment implementations were investigated. A total of 17 female early childhood teachers participated in this study. The data for this phenomenological study was collected through semi-structured interviews. The researchers developed some themes and categories based on participant teachers' responses about validity and reliability through constant-comparative data analysis method. The finding of the study revealed that the teachers who took part believe that obtaining valid and reliable information is critical for their assessment. Their self-reported practices, on the other hand, revealed that they have some difficulty using multiple assessment procedures with one another due to crowded classrooms and other excessive paper shuffling. The findings once again underline the importance of practical teacher training in understanding that assessment is an evidence-based process.

Keywords: Early Childhood Education, Formative Assessment, Validity, Reliability, Early Childhood Teachers

Erken Çocukluk Değerlendirmesinde Kritik Bir Nokta: Öğretmenlerin Biçimlendirici Değerlendirmelerindeki Geçerlik ve Güvenirlik Konuları

Öz: Bu çalışmada, Türk okul öncesi öğretmenlerinin biçimlendirici değerlendirme uygulamalarında güvenirlik ve geçerlilik sağlamaya yönelik kendi bildirdikleri inanç ve uygulamaları incelenmiştir. Bu çalışmaya toplam 17 kadın okul öncesi öğretmeni katılmıştır. Gerçekleştirilen fenomenoloji çalışmanın verileri yarı-yapılandırılmış görüşmeler yoluyla toplanmıştır. Araştırmacılar, katılımcı öğretmenlerin geçerlik ve güvenirlik konusundaki yanıtlarına dayalı olarak sürekli karşılaştırmalı veri analizi yöntemiyle bazı temalar ve kategoriler geliştirmiştir. Araştırmanın bulgusu, katılan öğretmenlerin değerlendirmeleri için geçerli ve güvenilir bilgi edinmenin kritik olduğuna inandıklarını ortaya koymuştur. Öğretmenlerin kendi deneyimleri, kalabalık sınıflar ve diğer kâğıt işleri nedeniyle çoklu değerlendirme prosedürlerini bir arada kullanmakta zorlandıklarını göstermiştir. Bulgular, değerlendirmenin kanıta dayalı bir süreç olduğunu kavramada, uygulamaya yönelik öğretmen eğitiminin ne derece önemi olduğunun bir kez daha altını çizmektedir.

Anahtar Sözcükler: Erken Çocukluk Eğitimi, Biçimlendirici Değerlendirme, Geçerlik, Güvenirlik, Erken Çocukluk Öğretmenleri

Received: 02.08.2022 Accepted: 10.01.2023 Article Type: Research Article

[1] TED University, Faculty of Education, Department of Elementary and Early Childhood Education, Ankara, Turkey, e-mail: elif.buldu@tedu.edu.tr, ORCID: https://orcid.org/0000-0003-0585-0138

[2] TED University, Faculty of Education, Department of Elementary and Early Childhood Education, Ankara, Turkey, e-mail: cagla.sendil@tedu.edu.tr, ORCID: https://orcid.org/0000-0003-0622-4315

Atıf için/ To cite:
Buldu, E., & Öneren Şendil, Çağla. (2023). A Critical point about early childhood assessments: validity and reliability ıssues in teachers' formative assessment. *Journal of Education for Life*, *37*(1), 253-268. https://doi.org/10.33308/26674874.2023371507

In recent years, educational reforms have mostly focused on how to ensure the quality of teaching and learning, as well as how to assess learners' learning on an ongoing basis (Peters et al., 2019). This viewpoint has aroused interest in formative assessment strategies for supporting children's learning and making decisions about the learning process (Black & William, 2009; Boyle & Charles, 2010; Cauley & McMillan, 2010; Clark, 2011; Riley-Ayers, 2014; Wortham, 2008). Although people outside the early childhood educational setting associate the word "assessment" with a paper-and-pencil test, this idea is inappropriate for early childhood education, which conducts assessment every day in the classroom and beyond in an authentic manner (National Research Council [NRC], 2008). Assessment can be used to learn about the needs of children and how to best support their learning and development. This form of assessment is known as assessment for learning or formative evaluation. According to the National Research Council (NRC, 2008, p. 425), "an assessment designed to evaluate progress toward an objective and used to guide curricular and instructional decisions".

Teachers and early childhood programs can be more sensitive and attentive to children thanks to diverse and ongoing observations (McAfee & Leong, 2016). Teachers focus on the rationale and arguments they attach on their decisions to answer in formative assessment rather than the children's answers to questions (Leung, 2004). Teachers modify experiences and instruction as they acquire knowledge to help children learn more effectively. Teachers can scaffold children's learning and development by providing appropriate guidance and assistance because their reasoning skills become available to them during the assessment process. Using assessment as a formative manner allows teachers to continuously collect information about children' learning and then use that information to improve learning opportunities (NRC, 2001). As a result, formative assessment is a cyclical process of planning, design, data collection, interpretation, and decision-making (Earl, 2003; McAfee & Leong, 2016; Stobart, 2012). Some assessment decisions must be taken before preparation for systematic assessment in a cyclical manner. The purpose of conducting assessment, what will be assessed, when will be assessed, methods of documenting-recording assessment information, summarizing all assessment pieces, assigning meaning to that information, and deciding how to use all of that information for the reasons or intended purposes are all major decisions in this cycle. In other words, the basic decisions in the assessment cycle are why, what, and when to assess, documenting (gathering & recording), summarizing, interpreting, and using assessment information (McAfee & Leong, 2016).

Apart from the numerous disagreements over the purpose, strategy, and method of formative assessment, the advantages for learners and ways in which it enhances learning are now clearly obvious. (Boyle & Charles, 2010; Pastore et al., 2019; Pellegrino et al., 2001; Popham, 2011), because any type of assessment practices can provide some information about teaching and learning, and have an observable impact on the speed of children's learning (Boyle & Charles, 2010; Popham, 2011). To make the assessment process powerful, regardless of the assessment strategies employed in the classroom—formative, summative, or large-scale—the collected information should have a reasonable level of validity and reliability (NRC, 2008). Rather than debating the effectiveness of assessment in early childhood education, it would be more useful to focus on the validity and reliability issues that lie at the heart of appropriate assessment. Early childhood teachers, according to Wortham and Hardin (2020), prefer to use formative assessment procedures that they establish; therefore, they must know how to develop valid and reliable assessment strategies as well as make bias-free interpretations in order to effectively organize learning process. Formative assessment implementations that are valid, reliable, and accurate have a broader scope, including observation of children in and out of the classroom, data collection from both children and their primary caregivers, and direct communication with the child without the use of formal test items or materials (NRC, 2008). Due to this, there has been a growing concern regarding the validity and reliability of formative assessment implementation, especially in early childhood education settings. (Wortham, 2008).

Validity and Reliability in Formative Assessment

Validity and reliability are two crucial and interconnected terms in the realm of practice that have numerous definitions and explanations. Because a large portion of the information collected through

formative assessment is qualitative, a positivist investigation of validity and reliability in assessment can serve the purpose (Frey, 2018).

Validity is primarily concerned with how the acquired information is interpreted and applied during the assessment process (Stobart, 2012; Shepard, 2009). Content, construct, and instructional validity are the three sub-dimensions of validity (Fraenkel et. al., 2012). The assessment procedure is required to obtain valid information in order to make appropriate inferences and ensure that all parts of the children's intended learning are covered (NRC, 2008; Organization for Economic Co-operation and Development [OECD], 2013). Capturing and assessing behaviors appropriately, assuring trustworthiness and accuracy, balancing behavior sampling and assessment, and assessing children's specific behavior using multiple strategies to support inferences are some of the key indicators of a valid assessment procedure in formative assessment (McAfee & Leong, 2016; Newton, 2012; Wortham, 2008). Furthermore, reliability means that the assessment is based on the consistency of assessing what it sets out to measure and that it is repeated later to ensure that the assessment process is not impacted by other occurrences (OECD, 2013). It ensures consistency across multiple assessments or measurements (NRC, 2008). The validity and reliability of formative assessment procedures, on the other hand, differs depending on how teachers conceptualize them. As a result, formative assessment reliability is determined by the consistency of many components created by teachers to provide information on a child's learning and development. Validity is related to the accuracy and efficiency of instructional practices developed by teachers in response to assessment results, much as reliability is (Way et al., 2010).

According to Katz (1997), teachers' awareness of making inappropriate inferences while assessing children can help them make less mistakes in their assessment interpretation. Nevertheless, teachers often have insufficient levels of assessment literacy, according to investigations of their assessment practices (De Luca et al., 2016; Işıkoğlu-Erdoğan et al., 2021; Matters, 2006). External sources, personal knowledge, beliefs and attitudes, and professional experimentation, according to Clarke and Hollingsworth (2002), are some of the factors that contribute to teachers' professional development. Any change to one of the constructs can assist in the change of the others. According to this understanding, working on teachers' knowledge and practice can extend assessment literacy and the spectrum of assessment frameworks that teachers use. Many researchers have also claimed that formative assessment has some drawbacks in terms of validity and reliability, which might lead to its misuse (Black et al., 2010). Some predetermined evaluation criteria that are important for consistency and uniformity cannot be determined since formative assessment is a component of everyday teaching and learning procedures (Leung, 2004). Formative assessment procedures can be influenced by teachers' opinions by relying significantly on the values ascribed by teachers to children' performances. As a result, both teachers' knowledge and beliefs about formative assessment are important (Leung, 2004).

In different educational settings and circumstances, empirical investigations on assessment for learning have been conducted. Despite the fact that summative assessment is the most-studied strategy (Harlen, 2007; Taras, 2009; Trotter, 2006), researchers generally agree on the effectiveness of formative assessment, which is a less-studied topic (Cauley & McMillan, 2010; Katz, 1997; Moss et al., 2006; Pastore et al., 2019). Researchers recognize that formative assessment strategies are difficult to implement and blend with the learning process (Yılmaz et al., 2020) and that most teachers do not understand them properly (Bennett, 2011), which could be one reason for the lack of research into formative assessment studies. Işıkoğlu-Erdoğan et al. (2021) conducted a study to investigate preschool teachers' level of competency in child assessment. They found that teachers have mostly intermediate and low levels of competency in assessing young children. Teachers with low levels of competency largely use assessment practices to inform parents, understand children's skills/abilities, and determine their needs. Furthermore, while much of the research on validity and reliability issues has focused on the adoption of summative assessment strategies (Black et al., 2010; Crawford et al., 2001), there have been a limited number of studies on the validity and reliability of formative assessment (Brookhart & Moss, 2014; Shepard, 2009). Studies on formative assessment in early childhood education, on the other hand, rarely investigated the notion of the validity and reliability issues.

Assessment in Turkish Early Childhood Education

Since it is highly important to determine to what extent performed education and teaching activities have been effective, assessing children objectively by taking their personal differences into consideration is crucial in the early childhood curriculum of Turkey (Ministry of National Education [MoNE], 2013). Because educational activities are planned in light of the curriculum's gains and indicators which represent behaviors associated with desired outcomes on children's whole developmental areas, determining whether or not the learning outcomes predicted during the process have been achieved at the end of the learning process is vital. The assessment in the national curriculum is multidimensional which includes various aspects such as observing the child in all developmental areas (cognitive, linguistic, motor, social and emotional development, and self-care skills), reporting the results of the observations, evaluating the activity plans, and the teacher's self-evaluation expected to be considered.

The first dimension of assessment in the national curriculum refers to the development of children (MoNe, 2013). Teachers can keep records of a child's development by employing observational tools. The observation forms can be filled out twice a year by relying on the recordings. Teachers can reflect on their analyses and evaluations based on the information in the development observation form in this report, and they can make educational and developmental recommendations to families. The evaluation of the everyday educational process and activities is the second dimension. Activities should also be evaluated in the curriculum to identify any emerging program requirements. This can be accomplished by asking questions (descriptive, affective, gain-related, and life-related), using worksheets/memory cards, sketching drawings, creating banners/posters/graphics, and the photos taken regarding the activity. Teachers' self-evaluation is the final dimension of national curriculum assessment. Teachers also assess themselves by examining personal characteristics connected to their teaching profession. Teachers can use the results of this self-assessment to make efforts to improve themselves in accordance with their level of competence, look for professional development facilities and resources, and, to that purpose, request assistance and help from the school administration (MoNE, 2013).

Present Study

The theoretical value of implementing formative assessment in early childhood education has long been recognized (Clark, 2011; Earl, 2003), and its implementation in classrooms has been studied extensively (Boyle & Charles, 2010; De Luca et al., 2016; Pastore et al., 2019; Tariq, 2013). However, research on the validity and reliability of formative assessment is quite limited (Nitko & Brookhart, 2011; Black & William, 2009), and more research into the reliability and validity of formative assessment in early childhood education is needed. As a result, the current study aims to identify Turkish early childhood teachers' self-reported beliefs and practices in terms of conducting a valid and reliable formative assessment. In light of these assumptions, the research question was framed as follows: "What are the self-reported beliefs and practices of Turkish early childhood teachers about reliability and validity in formative assessment?"

Method

The purpose of this study is to investigate preschool teachers' self-reported beliefs and practices about their formative assessment process, as well as how they incorporated validity and reliability into their assessment implementations, using the phenomenological research paradigm. Phenomenological research is a powerful qualitative research method to uncover the commonality of participants' lived experiences for a specific phenomenon (Moustakas, 1994). In phenomenological research, the researchers acquire in-depth information using qualitative methods such as interviews, observations, and documents to understand participants' personal perspectives and interpretations (Creswell, 2013; Paley, 2017). Parallel with this, the current study used phenomenological research to learn about teachers' self-reported beliefs and practices regarding validity and reliability through interviews.

Participants and Settings

Purposive sampling was employed to select participants in order to reach out to information-rich individuals who were related to the topic of interest (Patton, 2015). Phenomenology utilizes a purposive sample, which is a group of individuals chosen on the basis of their experience with the topic of interest in qualitative research (Mapp, 2008; Paley, 2017). In phenomenological research, similar to other qualitative research methods, the participant selection process emphasizes saturation and getting rich information rather than representing a population (Miles & Huberman, 1994). To ensure maximal diversity in the participants, the researchers included 17 female early childhood teachers from ten different cities in Turkey. The age of the participants ranged from 23 to 39, and their professional experience ranged from one to 14 years. As presented in the Table 1, 13 of the participants had undergraduate degrees, four had open education undergraduate degrees, and one had a high school degree in early childhood education.

Table 1. *Participant ECE Teachers' Profiles*

Years of Experience	N
1-4	8
5-9	6
10-15	3
Graduation Degree	
Bachelor's Degree	14
Bachelor's Degree (Open)	3
Number of student	
5-9	1
10-14	3
15-19	7
20-24	5
25-30	1
Special Training on Assessment	
Yes	2
No	15
Institution	
Private	5
Public	12

All of the participant teachers had ongoing teaching practices from private and public early childhood institutions where they implemented a variety of early childhood curriculum, as shown in Table 1. MoNE's early childhood curriculum is used in both private and public institutions. Only two teachers received special assessment training through in-service training from MoNE, whereas the remaining teachers did not receive special assessment training.

Data Collection Tool and Procedure

Semi-structured interview form was the main data collection tool in the current study. The participant teachers reflected their self-reported beliefs and practices on how they embedded validity and reliability in their assessment cycle through the interview questions. The researchers developed a semi-structured interview form after conducting in-deep investigation of the relevant literature. Two different experts in the relevant field reviewed the interview questions before the final version. The experts suggested a few phrasing changes and one additional question. What participating teachers routinely do as a part of the assessment process was the proposed question. The interview questions were finalized following the reviews. There were 19 questions on the interview form, including (a) How do you start, continue, and end an assessment process? So, could you tell us about your evaluation procedure in a few words?, (b) How do you decide what to assess before collecting data?, and (c) Are children's developmental characteristics and individual differences a criterion for selecting assessment techniques? Why? The researchers conducted a pilot study by interviewing three early childhood teachers before beginning the main data collection process. After the pilot research, nothing changed in the interview form.

Individual interviews were conducted via Zoom meetings, which determined the best times for the participants' teachers to meet. The interviews lasted an average of 50 minutes. The researchers recorded all of the online meetings and then transcribed them. The investigator triangulation was used in this study to minimize sampling, procedural, and measurement bias (Denzin, 1973). The validation of gathered data was attempted by involving the researchers.

Data Analysis

The data was analyzed using the constant-comparative data analysis method. This inductive data coding approach sorts the data by assessing and coding the information obtained from participants on a continuous basis (Glaser & Strauss, 1967).

All of the interview responses were arranged chronologically. As shown in Figure 1, the researchers conducted extensive study into the relevant literature to determine what constitutes valid and reliable indicators and compiled a list of these indicators.

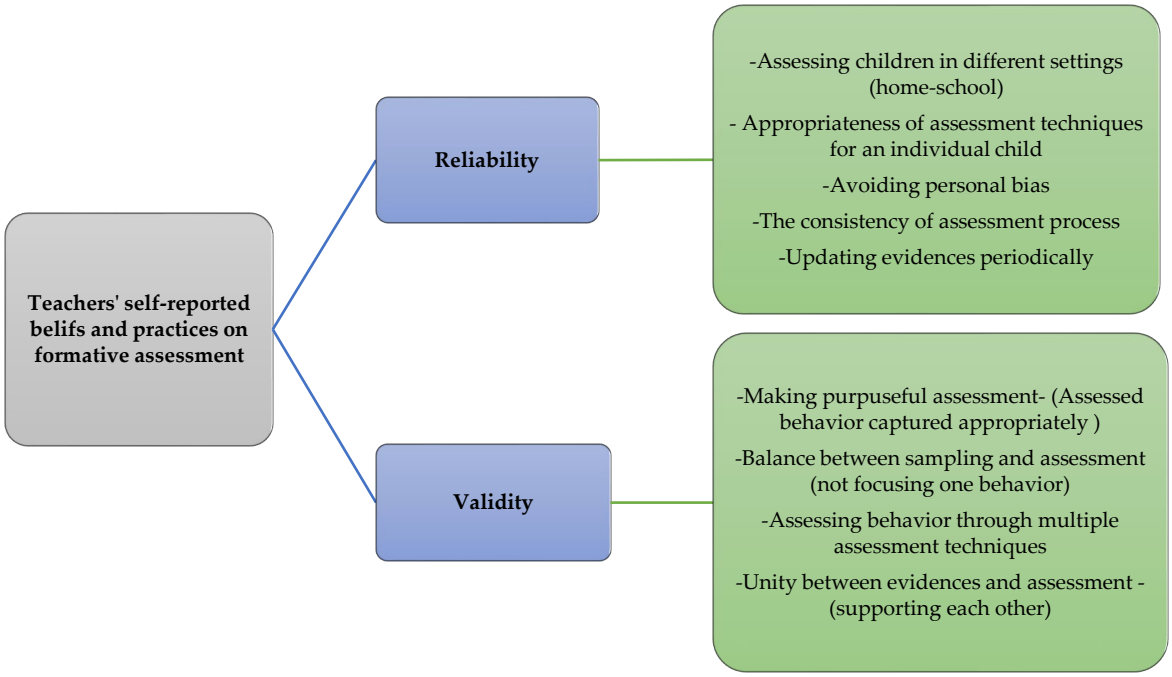

Figure 1. *Main Themes, and Categories*

It is critical to work closely with the indicators of validity and reliability while conducting data analysis. The researchers' codes and categories were compared with each other after numerous individual close readings. The second coder who experts in early childhood education examined 30% of the interview transcripts in order to increase the study's accountability by removing any potential researcher biases. The inter-coder reliability proposed by Miles and Huberman (1994) was applied in this situation, and the researchers' consensus was calculated as (108/108+7x100). 94). After everyone agreed on all of the codes and categories, the final codes and categories were created to define the study's main themes.

The current study may have some possible limitations in its nature. The findings were presented based on the participant teachers' self-reported responses, and included just 17 participant, therefore it is subject to biases that may have influences to the study. To overcome this limitation, participant variation was made as much as possible.

The following are some reliability indictors;

- Assessing children in different settings (home-school) whether to obtain similar performance at another time-place

- Appropriateness of assessment techniques for an individual child,
- Avoiding personal bias,
- The consistency of assessment process,
- Updating evidences periodically.

Furthermore, the indicators for validity were determined as follows:

- Making purposeful assessment (Assessed behavior captured appropriately),
- Balance between sampling and assessment (not focusing one behavior),
- Assessing behavior through multiple assessment techniques,
- Unity between evidences and assessment (supporting each other).

Based on the participating teachers' responses, the researchers made some connection between participant teachers' assessment practice during their assessment cycle, and reliability validity indicators. As presented in the Figure 2, the developed framework reflects the assessment process that participant teachers follow as well as the indicators of validity and reliability they consider during assessment.

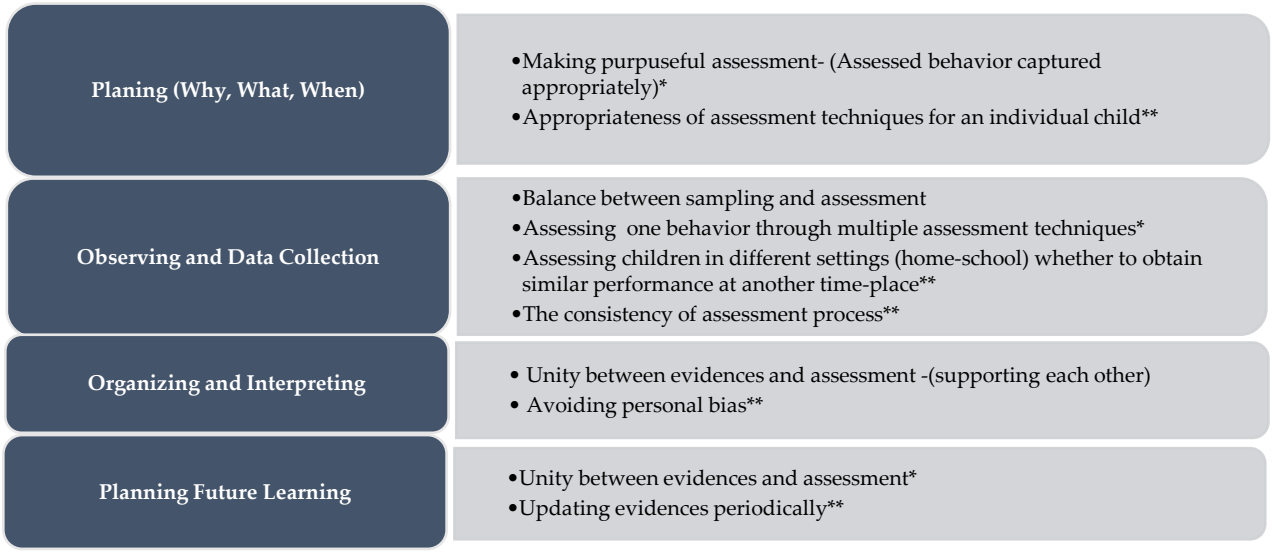

Figure 2. *The Developed Framework for Validity and Reliability Based on the Cycle of Assessment*

Note: *Represents validity indicators, and **Represents reliability indicators

The assessment cycle is represented by the left column, which includes (a) planning, (b) observing and collecting data, (c) organizing and interpreting, and (d) planning for future learning. Validity and reliability indicators are provided in the right column.

Results

The interview revealed that Turkish early childhood teachers' beliefs and practices on reliability and validity in formative assessment. The major categories under the validity theme are (1) making purposeful assessment, (2) balancing sampling and assessment, (3) assessing behavior using multiple assessment techniques, and (4) unity between evidences and assessment. Furthermore, under the reliability theme, major categories include (1) assessing children in different settings (home-school) to determine whether to obtain similar performance at another time-place, (2) appropriateness of assessment techniques for an individual child, (3) avoiding personal bias, (4) the consistency of the assessment process, and (5) updating evidences periodically.

Reliability

Assessing Children in Different Settings (Home-School) Whether to Obtain Similar Performance at Another Time-Place

Some teachers (n=7) stated that they should evaluate children in various situations, such as at home and at school, while monitoring and collecting data in order to acquire similar results at a later time and location. T3 underlined the importance of not relying on a single assessment technique to determine a child's performance. Another teacher said,

> *I take notes to see if the child is doing this behavior right now or at every opportunity, and I try to compare these notes with their future behaviors to determine if it is changeable or sustained. I speak with the family about how he or she behaves not only at school but also at home* (T2).

Furthermore, one of the teachers noted that maintaining balance and integrity in assessment findings is a challenge that she tries to solve by assessing children multiple times (T14). The rest of teachers did not mention about this reliability indicator while explaining their assessment process. One of the teachers, on the other hand, stated what it meant for her to reflect children's true performance as follows: "*If I observe the behavior I want the child to gain, it means I have made a proper and true assessment*" (T11).

Appropriateness of Assessment Techniques for an Individual Child

Almost all of the teachers (n=13) agreed that they were aware of appropriate assessment techniques and that they planned the proper assessment process for each child to ensure the formative assessment process' reliability. Many of the teachers emphasized the significance of considering the children's age and developmental differences, as well as the use of different assessment tools. One of the teachers mentioned that she began the assessment process by evaluating the children's individual characteristics and monitoring them individually in various activities (T14). Another teacher stated that due of the individual differences and developmental characteristics of children, utilizing a single assessment technique would result in erroneous assessment results (T12).

On the other hand, some of them stated that selecting assessment procedures and instruments was challenging. For one of the teachers, selecting the appropriate assessment tool (T3) can be difficult. Similarly, one of the teachers stated that because she had a crowded classroom and a lot of paper work, she couldn't choose an appropriate assessment technique and tool based on the children's individual differences (T5). Furthermore, a large number of teachers (n=14) indicated that they only use observation as an assessment tool and did not identify any other techniques.

Avoiding Personal Bias

Seven of the participating teachers' statements (n=7) could be considered under the category of "avoiding personal bias". Teachers regularly acknowledged the objectivity of the assessment process while analyzing the gathered information as they presented their interpretations. "*I consider assessment as a type of self-criticism for teachers*," one teacher said, "*and I use it as evidence retrospectively while filling out questionnaires and other form*s" (T9). They also stated that assessment allows them to examine children holistically and properly reflect the truth about their learning and development, as long as it is done objectively. One teacher also stated that their observations are reliable because they are already familiar with each child's development in their classroom. On the other hand, one of the participating teachers stated that children's development is apparent, hence the assessment result cannot possibly reflect reality (T5).

The Consistency of Assessment Process

The consistency of the assessment process is another category considered for reliability. Six of the teachers (n=6) mentioned the challenge of collaborating with parents and achieving similar results from different assessment tools while observing and collecting data from children. For a reliable and consistent assessment result, the teachers who cited the consistency of the assessment process mainly agreed that the behavior should be repeated multiple times in different environments and situations. Two of the teachers

mentioned that conducting assessments on a regular basis allows them to make the best decisions possible when comparing different assessment results (T9). T11 mentioned that consistency of long-term sustained observations is important for collected information. She also emphasized that consistency requires making decisions based on observations of the child in a variety of settings. Furthermore, many teachers (n=16) indicated that they prefer to collect information from children through child observation as a main formative assessment method.

Updating Evidences Periodically

One of the most commonly stated terms while explaining their practices for planning future learning under the reliability theme is the category of updating evidences on a regular basis. While some of the participants' teachers (n=8) reported that they conducted assessments on a regular basis at specific times, the majority stated that they conducted assessments based on the needs of the children and the program. Teachers who responded that they used assessment at specific periods stated that they did it once a month on average. Teachers, on the other hand, who indicated that they assessed according to the needs, stated that "*I update and review my assessment results at the end of the time designated beforehand for the acquisition of the behavior*" (T3). Another teachers reported that "*there is no fixed and definite time period. Depending on the conduct, I'm assessing, it varies*" (T5). One of teachers also mentioned following excerpt;

> *I use the assessment at the end of every day's circle time. In addition, I communicate with the children about the activities and their learning process, and I include all of the activities in their portfolios once a month* (T15).

Validity

Making Purposeful Assessment

The majority of teachers (n=16) stressed the importance of conducting purposeful assessments while planning their assessment process, as well as measuring what is intended to be measured. Some teachers (n=10) stated that gains and indicators are used as a foundation for purposeful assessment. They noted that through engaging in appropriate assessment, they were able to identify whether they had made progress and whether the child had met a specific goal. Aside from improvements and indicators, some of the teachers (n=6) noted starting the assessment with a specific curriculum or developmental area. Teachers reported using appropriate tools or techniques in the direction of the assessment's purpose. T3 summed it up this way:

> *It is impossible to use each assessment tool for measuring every kind of behavior. In order to appropriately measure the behavior I want to assess, I need to select the appropriate tool suitable with my aim. For example, tools used to measure motor development will be differ from the tools measuring language skills.*

Teachers also stressed that (n=7) they employ age-related similarities, different needs, abilities, and developmental milestones in specific developmental areas to make more purposeful assessments and to accurately capture the assessed behavior.

Balance Between Sampling and Assessment

Convergence of the behavior evidenced from a various sources and context are mentioned by some of the teachers (n=6) for achieving valid and representative sample. Balance was emphasized by some of the teachers (n=2) during the observation and data collection process in order to avoid relying on just one type of developmental area. One of the participant teachers put it as;

> *Each of the developmental areas works in conjunction with the others. It would be a mistake to focus our efforts just on one area of development. All of the activities we did in class, as well as the assessment methods we used, should focus on all of the developmental areas of children.* (T6)

Other teachers (n=2) acknowledged the importance of balancing assessment data from multiple sources. For example, T11 stressed the need of asking family members or primary caregivers about their child's learning and development in order to complement the teacher's assessment information. "*I understand that, if the assessment information received from multiple sources supports each other, the assessment I conducted is appropriate*," T11 said. To put it another way, "*getting information from a variety of sources ensures that the assessment data is representative*" (T3).

Assessing behavior through multiple assessment techniques

Some of the teachers (n=10) emphasized the need of having enough samples to represent the entire behavior or developmental level of the children during the observation and data gathering processes. T14 mentioned about adequately representing the behavior in means of providing evidences from multiple methods by saying that "*gathering evidences by using variety of methods (portfolios, daily evaluations, monthly evaluations, observation forms, anecdotes, etc.) can also be used to measure the behavior or development of children*". T17 also stated that "*in order to remediate the possible deficit of one method, it is critical to collect evidences on the development of the child using a variety of methods such as checklists, rubrics, anecdotal records, observation lists, portfolios, and audio record*." Furthermore, T14 emphasized the necessity of having a sufficient sample of behavior to accurately depict the behavior, stating;

> *If you are not pleased with the information in your hand, you should continue to gather information throughout the assessment process. Because children are ever-evolving beings, their actions change with time. As a result, you should collect data on the changing needs of a developing child throughout the semester and consider the assessment process to be ongoing* (T14).

Unity between evidences and assessment

In terms of interpreting and planning future learning stages of the assessment cycle, several of the teachers (n=10) emphasized the importance of gathering multiple evidences to support the assessment process. T3, for example, emphasized the importance of using continuing assessment techniques to assist teachers see the learning process from the beginning through the end of the semester. In addition, a few of the teachers (n=4) can use parent-teacher conferences to supplement classroom observation or assessments. Observing children's natural play activities and asking them questions are two of the methods indicated for ensuring consistency in the evidence and assessment data.

Discussion and Conclusion

The current study investigated 19 early childhood teachers' views on the validity and reliability of their assessment processes. As a result of the analysis, nine different categories were developed under the themes of reliability (assessing children in different settings whether to obtain similar performance at another time-place appropriateness of assessment techniques for an individual child, avoiding personal bias, the consistency of assessment process, updating evidences periodically), and validity (making purposeful assessment, balance between sampling and assessment, assessing behavior through multiple assessment techniques, unity between evidences and assessment). The findings showed that the participating teachers did not prefer to use different assessment tools to collect information due to reasons such as class size, time and excessive workload.

Formative assessment is an ongoing process in which early childhood teachers use the information to improve their teaching and guide the learning and development of children (Popham, 2011). Formative assessment is quite ideal for early childhood teachers because it can be used to assess everyday activities in a natural and ongoing manner (Riley-Ayers, 2014). Formative assessment implementations, which include planning, observing and data collection, interpretation, and decision making, are daily routines of the assessment cycle at various times of the day (Yılmaz et al, 2020); however, bringing these steps together with reliability and validity issues can be difficult for teachers. Assessment of learning and development, on the other hand, only gives useful information if it creates a representative framework for what the children have learnt (Mushi, 2001). Within the scope of formative assessments, reliable and valid processes for measuring children's overall development and learning should be ensured, and it is necessary to evaluate whether it meets or does not meet the expectations (NRC, 2008). As a result, we can use reliability and validity indicators to assist and improve assessment practices (Wesolowski, 2020).

Some of the teachers who took part in the current study emphasized convergence of behavior evidenced from various sources and contexts. The reported experiences of teachers are consistent with the literature in this regard. McAfee and Leong (2016) emphasize that assessment should focus on evaluating the whole child

in many areas while being sensitive and responsive to children's needs. Rather than focusing on holistic evaluation, teachers tend to evaluate the cognitive, language, social, and emotional areas of development more, according to Işıkoğlu-Erdoğan et al. (2021). It is critical to consider the individual needs of children as well as the appropriateness of assessment practices while drawing on a variety of sources and contexts at this point. Teachers can gain an ecologically valid knowledge of a child's development and learning by using observational methods of assessment, which allow them to collect data from a variety of sources across time (Meisels et al., 2001; NRC, 2008; Riley-Ayers, 2014). Despite the widespread use of formative assessment in early life, discrepancies in children's performance scores or assessments might be attributed to variances in teachers' evaluations (NRC, 2008; Waterman et al., 2012). Because a teacher is the source of information, ensuring reliability and validity may be difficult. When teachers make judgments based on their observations in the classroom, several external circumstances (time of day, relationship with the child, etc.) may influence the teacher's inferences. As a result, assessor variance should be addressed, particularly when formative assessments are involved. In other words, changes in a child's assessment outcomes should be due to the child, not the assessor or other factors that compromise and threaten assessment reliability and validity (Waterman et al., 2012). As the use of performance-based and observational metrics increases, so does the necessity for ensuring the reliability and validity of these constructs (Russo et al., 2019). As it is evident in the findings of the current study, teachers reported the use of observing children's play activities to provide integrity among the evidence and assessment information, using multiple methods by gathering evidence using a variety of methods can be considered among validity indicators. However, teachers also reported some challenges in choosing assessment techniques and tools for an individual child which may threaten the reliability of the assessment process. The findings support the suggestion of Katz (1997). In this sense, minimizing potential errors in formative assessment strategies is possible by using instruments according to their specific purpose in the planning process. However, the rest of the teachers in the current study stated that they had some difficulties while choosing appropriate assessment techniques for children's age and development. Similarly, Nah and Kwak (2011) indicate that teachers often lack plans for formative assessment, and therefore it makes formative assessment unsystematic and inefficient to catch children's learning and development individually. As teachers emphasized, class size can be another barrier to the appropriateness of assessment techniques for an individual child. In a similar vein, Cagasan et al. (2020) noted that class size can make it difficult for teachers to respond to the needs of individual children and can make adopting effective formative assessment procedures difficult.

Teachers can learn more about different learning styles and modes of expression in children by participating in various forms of assessment (Mushi, 2001). In other words, when children are examined using various assessment methods, they are better able to express their learning. Furthermore, assessing children in multiple settings (home-school) to see if they can achieve similar results at a different time is an important indicator of a reliable assessment process. The content of the assessment, according to Cazden (2001), is an effective technique of determining children's performance because working in a different context, such as familiar, comfortable, and informal settings, can influence children's performance. Teachers can gather a range of data over time in many contexts to ensure a more valid and reliable understanding of the children's learning and development (NRC, 2008). As a result, curriculum-based performance assessments are intrinsically linked to teachers' judgments and evaluations, and are generated from classroom activities and daily curriculum practices (Meisels et al., 2001; Wesolowski, 2020). However, if the assessment results are used for purposes other than informing classroom instruction, there is a risk for bias, threatening the information's reliability. Observing classes or implementing various forms of formative assessments requires teachers received the necessary training to ensure validity and reliability (NRC, 2008). As a result, teachers' abilities to make accurate inferences about children's learning and development are critical for making proper educational decisions (Wesolowski, 2020). In the current study, only a few teachers emphasized the necessity of avoiding personal bias while interpreting information from children. Personal biases can influence assessments in qualitative data, such as formative assessment (Morse et al., 2002). As a result, it is critical to ensure making accurate and trustworthy judgments. Besides it, different approaches for assuring trustworthiness should be used when compared to familiar techniques used in standardized assessment (Baker et al., 1993). Getting training to agree

on performance standards, validating and double-checking the assessment content to ensure appropriate assessment standards, specifying the criteria, reaching a shared understanding of performance standards, and so on are some of the suggested techniques. (Baker et al., 1993). Although the participating teachers stated that their observations are quite reliable because they are already familiar with each child's development and that the assessment result possibly reflects reality, they may need to use some of the suggested ways to ensure trustworthiness. Teachers' professional development might be improved by offering education in some basic assessment contents such as portfolios, screening tools, systematic observation, and so on (Işıkoğlu-Erdoğan et al., 2021). The participating teachers mentioned that they were able to appropriately depict the behavior by using numerous approaches such as portfolios, daily assessments, monthly evaluations, observation forms, anecdotes, and so on. In a study conducted by Işıkoğlu-Erdoğan et al. (2021), it was revealed that preschool teachers mostly use observation and portfolio techniques to share the results with relevant people. However, due to the nature of portfolios' richness, uniqueness of contents, and openness to interpretation, they also have certain reliability difficulties in terms of subjectivity, objectivity, and consistency. As a result, Tigelaar et al., (2005) advocated for at least two assessors who are familiar with the students being evaluated. Assessors should also engage in continuous assessment, revisiting and discussing their interpretation. Portfolio evaluation also requires the preparation and implementation of assessment rules, as well as the maintaining of interpretation records, to ensure trustworthiness (Tigelaar et al., 2005).

The majority of the teachers in the current study underlined the need of conducting purposeful assessments, particularly throughout the assessment planning process. Apart from that, they emphasized the significance of measuring what is intended to be measured. Due to the cyclic nature of the assessment process, it may be checked consistently throughout the assessment process from beginning to end. If teachers regard assessment as a practice that occurs at the end of the teaching and learning process, the harmony between what is taught and what is assessed may be threatened (Mushi, 2001). Participating teachers report that by conducting an appropriate evaluation, they are able to identify whether or not they have made progress and whether or not the child has met a certain goal. The majority of teachers appear to utilize evaluation as a last activity to ensure that advances have been realized. By providing adequate education for teachers, this assumption about the use of assessment practice should be reconsidered. Similarly, validity can be an issue in classroom-based assessment for the quality of inferences made by the teacher in order to measure student-learning outcomes. To achieve validity in the classroom, the teacher should find a balance between sampling and assessment, the level of thinking patterns, and congruency (Wesolowski, 2020). For example, relevance is related with the association between the learning objectives, national standards and issue being taught, as well the content of the assessment. Therefore, the representativeness of collected information is an important factor for a valid assessment in the classroom assessment perspective.

In light of these findings, the current study has some important implications for early childhood teachers who want to use valid and reliable formative assessment in their classrooms. According to the findings of the current study, it is vital to provide teachers with adequate assistance and training in order to emphasize the importance of obtaining valid and reliable findings from in-class assessment implementations. To do this, state-run organizations should give a series of practical workshops to promote teachers' professional development in performing purposeful assessments and assessing children using a variety of assessment techniques. The current study also makes an important recommendation for teachers who deal with overcrowding in the classroom. Technological devices and tools can provide a new avenue and promise advanced features for early childhood assessment such as apps, video, and voice recorders. Additionally, for working teachers with limited access to professional development training, an informative guidebook can be created to introduce formative assessment methods and their implementation with examples.

Declarations

Authors' contributions: Both of the authors worked on this manuscript equally during writing and revising.

Ethics approval and consent to participate: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and national research committee.

In this context, the necessary ethical permission to conduct the study was obtained with the decision numbered "2021" at meeting number 2 and dated 28.02.2021 of TED University Human Research Ethics Committee Regulations.

Competing interests: No potential conflict of interest was reported by the author(s).

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Baker, E. L., O'Neil, H. F., & Linn, R. L. (1993). Policy and validity prospects for performance-based assessments. *American Psychologist, 48*(12), 1210-1218. https://doi.org/10.1037/0003-066X.48.12.1210

Bennett, R. (2011). Formative assessment: A critical review. *Assessment in education: Principles, Policy & Practice, 18*(1), 5-25. https://doi.org/10.1080/0969594X.2010.513678

Black, P., & William, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5-31. https://doi.org/10.1007/s11092-008-9068-5

Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2010). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice, 17*(2), 215-232. https://doi.org/10.1080/09695941003696016

Boyle, B., & Charles, M. (2010). Defining ongoing assessment: The effective method for supporting teaching and learning in early years and primary education. *School Leadership and Management Journal, 30*(2), 285-300. https://doi.org/10.1080/13632434.2010.485184

Brookhart, S. M., & Moss, C. M. (2014). *The validity and reliability of information from a formative walk-through observation instrument.* https://www.duq.edu/assets/Documents/castl/_pdf/CASTL_tech_report_1-14.pdf

Cagasan, L., Care, E., Robertson, P., & Luo, R. (2020). Developing a formative assessment protocol to examine formative assessment practices in the Philippines. *Educational Assessment, 25*(4), 259-275. https://doi.org/10.1080/10627197.2020.1766960

Cauley, K. M., & McMillan, J. H. (2010). Formative assessment techniques to support student motivation and achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 83*(1), 1-6. https://doi.org/10.1080/00098650903267784

Cazden, C. B. (2001). *Classroom discourse: The language of teaching and learning* (2nd ed.). Heinemann.

Clark, I. (2011). Formative assessment: Policy, perspectives and practice. *Florida Journal of Educational Administration & Policy, 4*(2), 158-180.

Clarke, D., & Holllingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education, 18*, 947- 967. https://doi.org/10.1016/S0742-051X(02)00053-7

Crawford, L., Tindal, G., & Steiber, S. (2001). Using oral reading rate to predict student performance on statewide achievement tests. *Educational Assessment, 7*, 303-323. https://doi.org/10.1207/S15326977EA0704_04

Creswell, J. W. (2013). *Qualitative inquiry & research design: Choosing among the five approaches.* SAGE Publications.

De Luca, C., LaPointe, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability, 28*(3), 251–272. https://doi.org/10.1007/s11092-015-9233-6

Denzin, N. K. (1973). *The research act: A theoretical introduction to sociological methods.* Transaction Publishers.

Earl, L. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Corwin Press.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8th ed.). Mc Graw Hill.

Frey, B. (2018). *The SAGE encyclopedia of educational research, measurement, and evaluation* (1-4). SAGE Publications. https://doi.org/10.4135/9781506326139

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Aldine. https://doi.org/10.1097/00006199-196807000-00014

Harlen, W. (2007). Teachers' summative practices and assessment for learning tensions and synergies. *The Curriculum Journal, 16*(2), 207-223. https://doi.org/10.1080/09585170500136093

Işıkoğlu Erdoğan, N., Aydoğan, S., Efe-Kendüzler, S., Dülger-Ceylan, E., Aydın, A., & Dinler, H. (2021). Okul öncesi öğretmenlerinin çocukları değerlendirmedeki yeterlilik düzeyleri ve kullandıkları araçlar [Preschool Teachers' Competence Levels and Tools Used in Child Assessment]. *Yaşadıkça Eğitim, 35*(1), 1-19. https://doi.org/10.33308/26674874.2021351230

Katz, L. (1997). *A developmental approach to assessment of young children*. (ERIC ED 407172). ERIC Digest.

Leung, C. (2004) Developing formative teacher assessment: Knowledge, practice, and change, *Language Assessment Quarterly: An International Journal*, *1*(1), 19-41. https://doi.org/10.1207/s15434311laq0101_3

Mapp T. (2008). Understanding phenomenology: The lived experience. *British Journal of Midwifery, 16*(5), 308–311. https://doi.org/10.12968/bjom.2008.16.5.29192

Matters, G. (2006). *Using data to support student learning.* https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/dddm_pg_092909.pdf

McAfee, O., & Leong, D. (2016). *Assessing and guiding young children's development and learning* (6th ed.). Pearson.

Meisels, S. J, Bickel, D. D., Nicholson, J., Xue, Y., & Atkins-Burnett, S. (2001). Trusting teachers' judgements: A validity study of a curriculum-embedded performance assessment in kindergarten to grade 3. *American Educational Research Journal, 38*(1), 73–95. https://doi.org/10.3102/00028312038001073

Miles M.B. & Huberman A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage Publication.

Ministry of National Education [MONE] (2013). *Okul Öncesi Eğitim Programı [Early Childhood Education Program]*. MEB.

Morse, J. M., Barrett, M., Mayan, M., Olson, K., & Spiers, J. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods, 1*, 1-19. https://doi.org/10.1177/160940690200100202

Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Validity in educational assessment. *Review of Research in Education, 30*, 109– 162. https://doi.org/10.3102/0091732X030001109

Moustakas, C. E. (1994). *Phenomenological research methods*. Sage Publications. https://doi.org/10.4135/9781412995658

Mushi, S. L. P. (2001, April 10-14). *Evaluating validity and reliability of classroom assessments using secondary data.* [Paper presentation] Annual Meeting of the American Educational Research Association, Seattle, WA, USA.

Nah, K.O., & Kwak, J.I. (2011). Child assessment in early childhood education and care settings in South Korea. *Asian Social Science, 7*(6), 66-78. https://doi.org/10.5539/ass.v7n6p66

National Research Council [NRC] (2001). *Classroom assessment and the national science education standards*. The

National Academies Press.

National Research Council [NRC] (2008). *Early childhood assessment: Why, what, and how*. The National Academies Press.

Newton, P. (2012). Validity, purpose and the recycling of results from educational assessments. In J. Gardner (Ed.) *Assessment and learning* (pp.264-276). SAGE Publications. https://doi.org/10.4135/9781446250808.n16

Nitko, A. J., & Brookhart, S. M. (2011). *Educational Assessment of Students* (6th Ed.). Pearson.

Organization for Economic Co-operation and Development (2013). Student assessment: Putting the learner at the center. In *Synergies for Better Learning: An international perspective on evaluation and assessment.* OECD Publishing. https://doi.org/10.1787/9789264190658-7-en

Pastore, S., Manuti, A., & Scardigno, A. F. (2019). Formative assessment and teaching practice: The point of view of Italian teachers. *European Journal of Teacher Education, 42*(3), 359-374. https://doi.org/10.1080/02619768.2019.1604668

Paley J. (2017). *Phenomenology as qualitative research—A critical analysis of meaning attribution*. Routledge. https://doi.org/10.4324/9781315623979

Patton, M. Q. (2015). *Qualitative research & evaluation methods* (4th ed.). SAGE Publishing.

Pellegrino, J., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment.* National Academies Press.

Peters, L. E., Graves, S. B., Liang, E., & Akaba S. (2019, December 6). *Case studies on authentic assessment: Perspectives on fidelity, utility, and applications to practice* [Poster Presentation]. NYC Early Childhood research Network, Hunter College at the City University of New York.

Popham, J. W. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator, 46*(4), 265-273. https://doi.org/10.1080/08878730.2011.605048

Riley-Ayers, S. (2014). *Formative assessment: Guidance for early childhood policy makers.* (Ceelo Policy Report). http://ceelo.org/wp-content/uploads/2014/04/ceelo_policy_report_formative_assessment.pdf

Russo, J. M., Williford, A. P., Markowitz, A. J., Vitiello, V. E., & Bassok, D. (2019). Examining the validity of a widely-used school readiness assessment: Implications for teachers and early childhood programs. *Early Childhood Research Quarterly, 48,* 14-25. https://doi.org/10.1016/j.ecresq.2019.02.003

Shepard, L. (2009). Commentary: Evaluating the validity of formative and interim assessment. *Educational Measurement: Issues and Practice, 23*(3), 32-37. https://doi.org/10.1111/j.1745-3992.2009.00152.x

Stobart, G. (2012). Validity in formative assessment. In J. Gardner (Ed.) *Assessment and learning* (pp.233-242). SAGE Publications. https://doi.org/10.4135/9781446250808.n14

Taras, M. (2009). Summative assessment: the missing link for formative assessment. *Journal of Further and Higher Education, 33*(1), 57-69. https://doi.org/10.1080/03098770802638671

Tariq, M. A. (2013). Engaging professionals: Investigating in service teachers use of formative classroom assessment**.** *Universal Journal of Educational Research 1*(4), 318-322. https://doi.org/10.13189/ujer.2013.010407

Tigelaar, D. E. H., Dolmans, D. H. J. M., Wolfhagen, I. H. A. P. & van der Vleuten, C. P. M. (2005). Quality issues in judging portfolios: Implications for organizing teaching portfolio assessment procedures. *Studies in Higher Education, 30*(5), 595-610. https://doi.org/10.1080/03075070500249302

Trotter, E. (2006). Student perceptions of continuous summative assessment. *Assessment & Evaluation in Higher Education, 31*(5), 505-521. https://doi.org/10.1080/02602930600679506

Waterman, C., McDermott, P. A., Fantuzzo, J. W., & Gadsden, V. L. (2012). The matter of assessor variance in early childhood education or whose score is it anyway. *Early Childhood Research Quarterly, 27*, 46-54. https://doi.org/10.1016/j.ecresq.2011.06.003

Way, W. D., Dolan, R. P., & Nichols, P. (2010). Psychometric challenges and opportunities in implementing formative assessment. In H. L. Andrade, & G. J. Cizek (Eds.) *Handbook of Formative Assessment* (pp.297-315). Routledge.

Wesolowski, B. C. (2020). Classroometrics: The validity, reliability, and fairness of classroom music assessments. *Music Educators Journal, 106*(3), 29-37. https://doi.org/10.1177/0027432119894634

Wortham, C. S. & Hardin, B. J. (2020). *Assessment in early childhood education*. (8th ed.). Pearson.

Yılmaz, A., Şahin, F., Buldu, M., Ülker Erdem, A., Ezmeci, F., Somer Ölmez, B., Aydos, E. H., Buldu, E., Ünal, H. B., Aras, S., Buldu, M., Akgül, E. (2020). An examination of Turkish early childhood teachers' challenges in implementing pedagogical documentation. *Early Childhood Education Journal*. https://doi.org/10.1007/s10643-020-01113-w