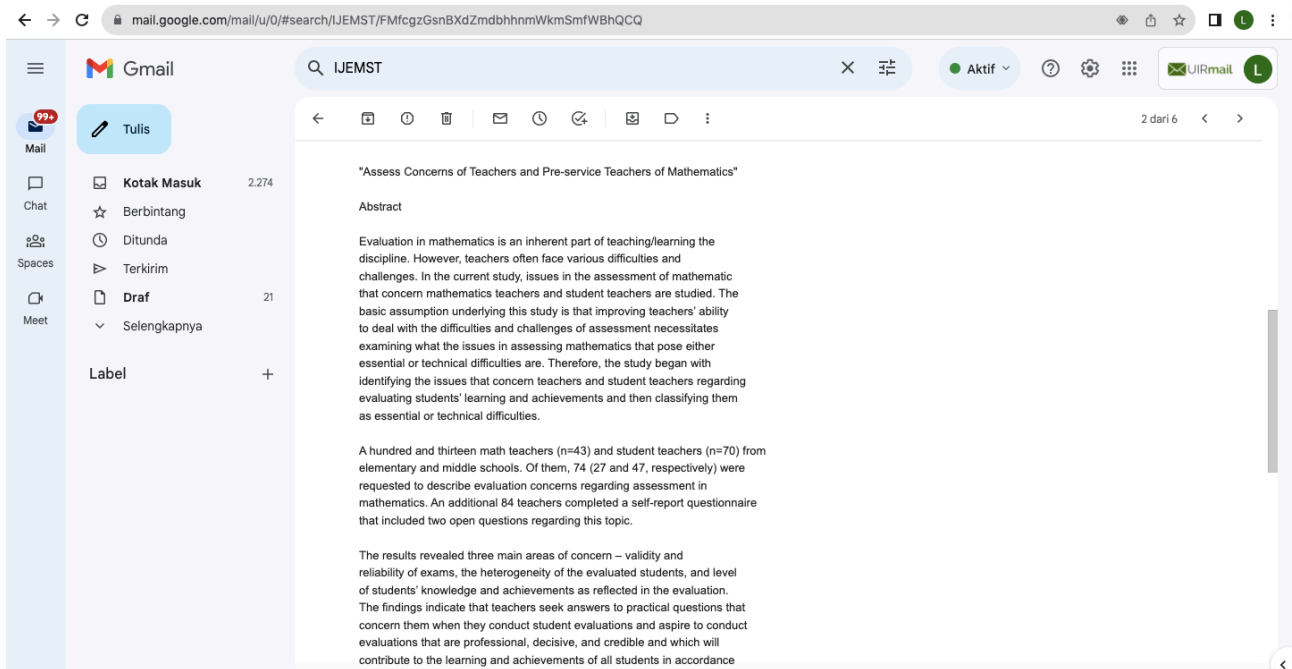
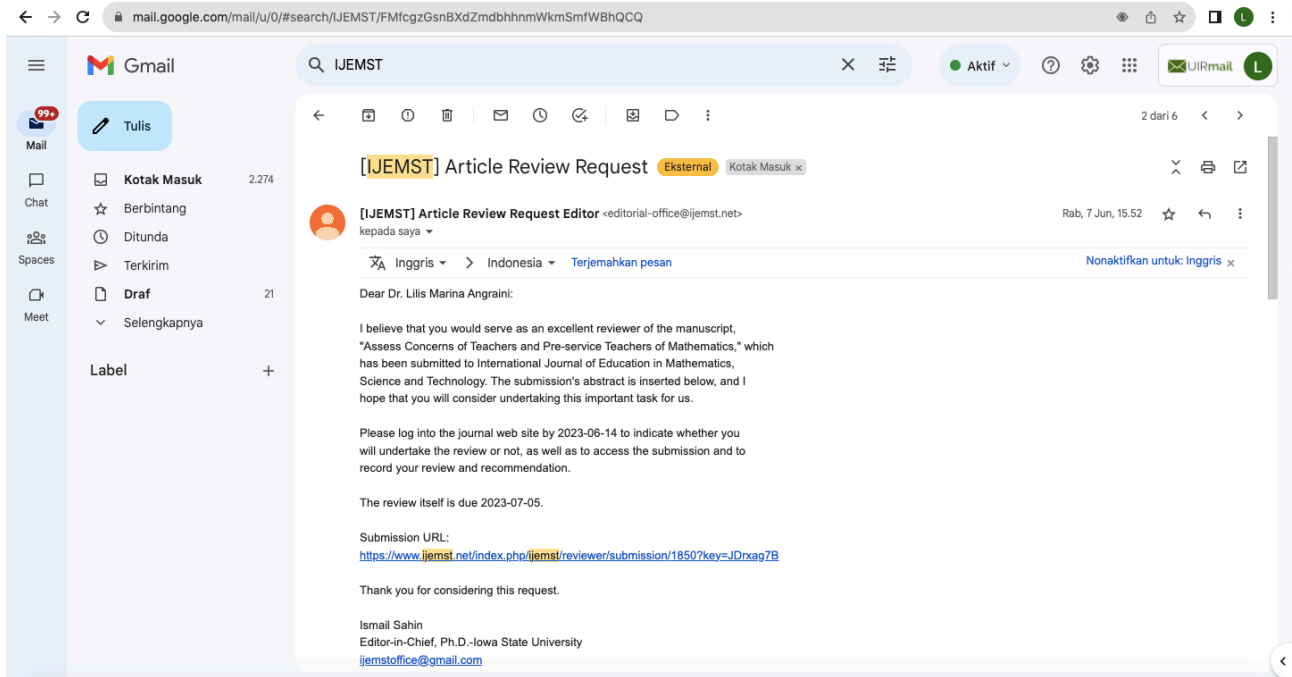
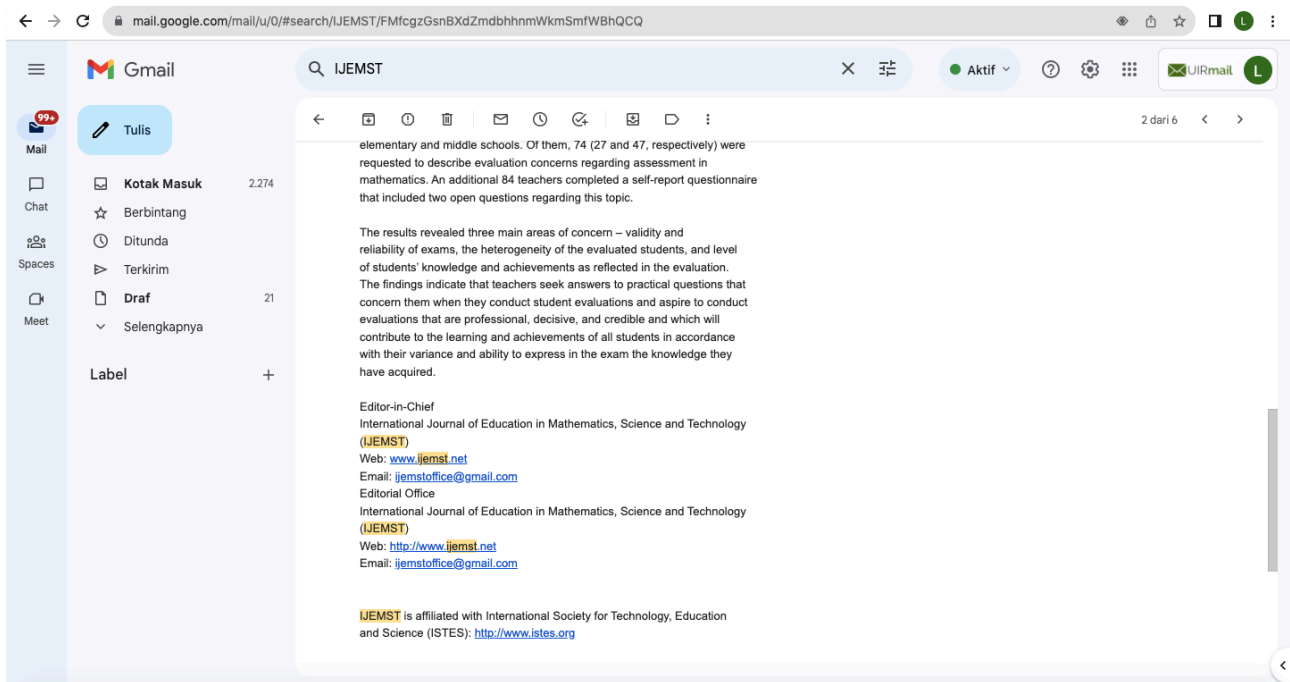
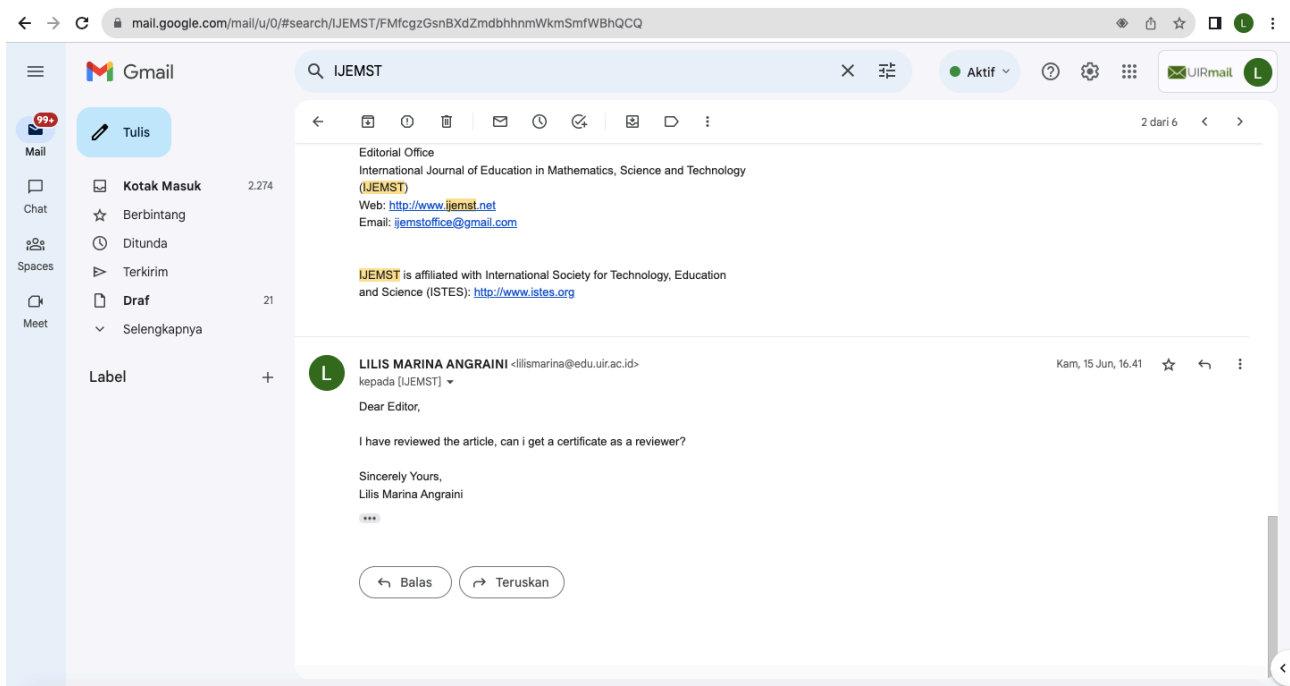


# 1. Permintaan review (07 Juni 2023)

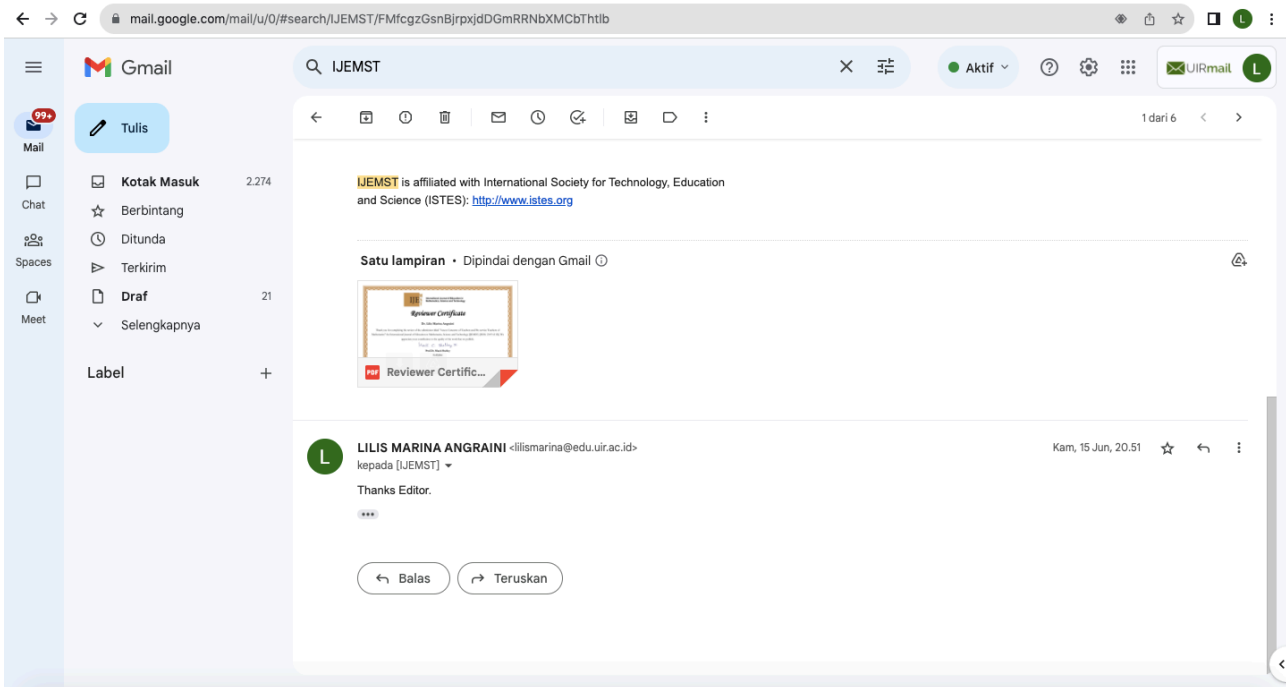
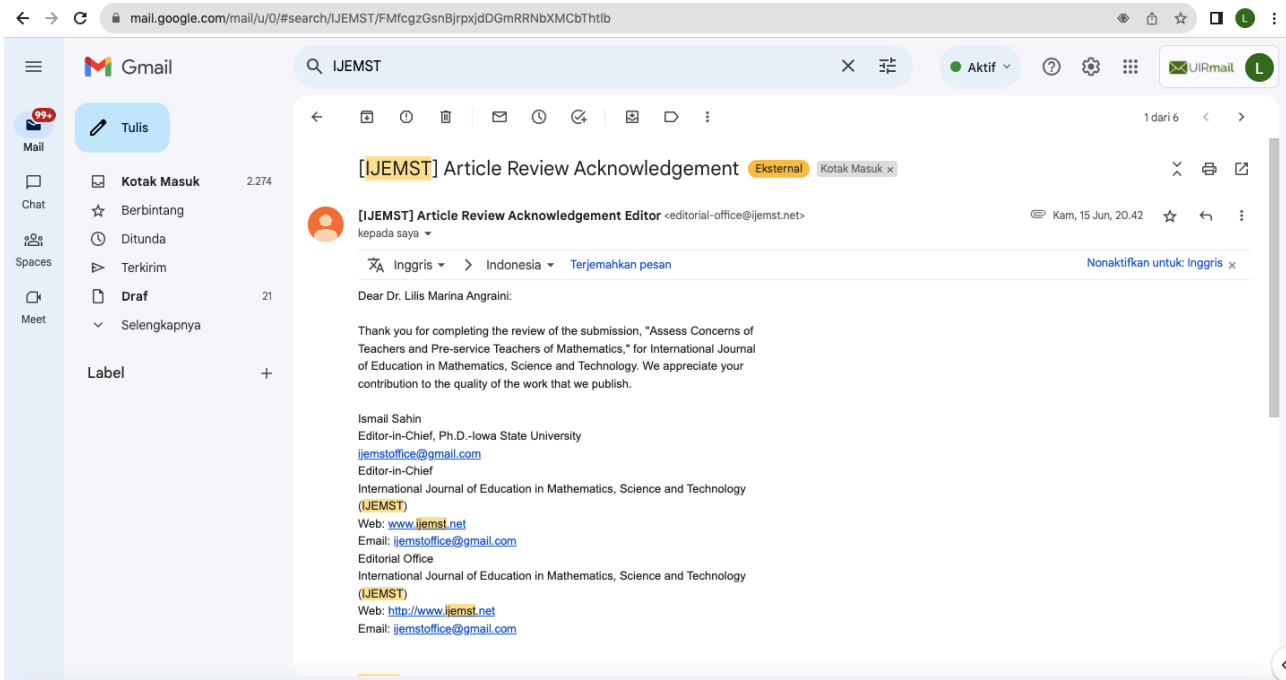




## 2. Penyerahan hasil review (15 Juni 2023)



### 3. Pengiriman sertifikat reviewer (15 Juni 2023)





International Journal of Education in  
Mathematics, Science and Technology

# *Reviewer Certificate*

**Dr. Lilis Marina Angraini**

Thank you for completing the review of the submission titled "Assess Concerns of Teachers and Pre-service Teachers of Mathematics" for International Journal of Education in Mathematics, Science and Technology (IJEMST) [ISSN: 2147-611X]. We appreciate your contribution to the quality of the work that we publish.

*Mack C. Shelley II*

**Prof.Dr. Mack Shelley**

**Co-Editor**

Abstracting/Indexing



## *Assess Concerns of Teachers and Pre-service Teachers of Mathematics*

### *Abstract*

Evaluation in mathematics is an inherent part of teaching/learning the discipline. However, teachers often face various difficulties and challenges. In the current study, issues in the assessment of mathematics that concern mathematics teachers and student teachers are studied. The basic assumption underlying this study is that improving teachers' ability to deal with the difficulties and challenges of assessment necessitates examining what the issues in assessing mathematics that pose either essential or technical difficulties are. Therefore, the study began with identifying the issues that concern teachers and student teachers regarding evaluating students' learning and achievements and then classifying them as essential or technical difficulties.

A hundred and thirteen math teachers ( $n=43$ ) and student teachers ( $n=70$ ) from elementary and middle schools. Of them, 74 (27 and 47, respectively) were requested to describe evaluation concerns regarding assessment in mathematics. An additional 84 teachers completed a self-report questionnaire that included two open questions regarding this topic.

The results revealed three main areas of concern – validity and reliability of exams, the heterogeneity of the evaluated students, and level of students' knowledge and achievements as reflected in the evaluation. The findings indicate that teachers seek answers to practical questions that concern them when they conduct student evaluations and aspire to conduct evaluations that are professional, decisive, and credible and which will contribute to the learning and achievements of all students in accordance with their variance and ability to express in the exam the knowledge they have acquired.

*Key words:* evaluation methods in mathematics, alternative evaluation, teaching and assessment of mathematics, quality mathematic teachers' training

### *Introduction*

All teachers, mathematics teachers (MTs) included, face concerns about assessing student learning and achievement while having to take into consideration content and curricula goals and objectives and suitable teaching methods, especially when dealing with heterogeneous classes and students with diverse needs (Cai et al., 2020; Veldhuis & van den Heuvel-Panhuizen, 2014, 2020). The area of assessment has received much attention in recent decades worldwide, Israel included (Cai et al., 2020; Davison & Leung, 2009; Kulm, 2013; Zhao et al., 2018). In mathematics, assessment is inseparable from teaching and learning, and teachers must evaluate their students using advanced, decisive, and credible methods, which sometimes lead to difficulties (Cai et al., 2020).

The goal of the present study was to pinpoint which issues MTs are concerned about regarding student assessment, whether traditional or alternative, to determine the advantages and disadvantages of each, and to map a course of action to deal with them.

### *Traditional assessment*

Tradition assessment is generally based on quantitative external and internal testing (Nevo, 2002, 2006) designed to determine and quantify acquired student knowledge in the subject studied using a numerical grade to represent the proportion of the desired level (Pellegrino, 2003). Usually, quantitative exams offer summative assessments, meaning that they

“close” the chapter (topic or stage of learning) without allowing a retrospective view of the learning process and an opportunity to correct or improve the grade.

In the traditional approach, teaching and assessment are often viewed as two separate activities, which raises serious doubts regarding the validity of tests and their contribution to learning (Savickiene, 2011). Criticism of the traditional assessment approach began in the late 1980s and has been ongoing since (Kulm, 2013; Research Advisory Committee of the National Council of Teachers of Mathematics, 1988; Watt, 2005).

Several researchers (Kulm, 2013; Watt, 2005) have pointed out a number of negative attributes that stem from traditional assessment: a quantitative test is uniform for all students and does not take into account learner variance; the exam often focuses on superficial knowledge and does not always reflect students’ abilities; and it may lead to partial mastery of the discipline due to the tendency to narrow down the goals of learning to the exam itself. These, in turn, reduce higher-order cognitive processes and does not effectively sample various thinking skills. Moreover, a quantitative exam can reduce studying to nothing more than an exercise in memorization, with the student investing effort only to be able repeat the teacher’s words as exact as possible (even without understanding). All these seriously limit the overall perspective and in-depth comprehension of the discipline.

Additionally, the examination situation itself is often accompanied by negative phenomena –stress and anxiety, desperation and frustration, and discomfort – that affect student performance and not only make test-taking more difficult, they may damage a student’s academic self-image (Ashcraft, 2002; Watt, 2005).

Because achievement tests also reflect teaching skills, teachers (not only MTs) tend to come under pressure to focus their teaching only on the material relevant to the specific

exam, at the expense of ensuring that their students develop proper understanding of the material (Ashcraft, 2002; Geist, 2010; Kulm, 2013; Rameau & Loumine, 2007; Watt, 2005).

Nevertheless, traditional assessment methods do have their advantages, the main one being that a quantitative test can be an objective assessment tool that is constructed in accordance with clear, pre-defined criteria, is uniform for all learners (Savickiene, 2011; Watt, 2005) and has an optimal degree of reliability. Achievements on quantitative tests can be classified by level and defined in broad hierarchical, coherent structures. Based on the results, teachers can define points of strength and weakness in the learning, compare and rate learners' achievements from a neutral starting point, and plan their next steps in teaching – for all the learners or for those with different levels of achievement (Kulm, 2013; Watt, 2005).

#### *Alternative assessment*

As a result of the shortcomings of traditional quantitative assessment, alternative methods of assessment are being developed based on the constructivist approach (Li et al., 2020). These aim to support or replace the traditional assessment that is usually directed toward quantitatively evaluating the final product. The names of these methods are indicative of their uniqueness and contribution to the students: formative assessment, assessment for learning, teacher-based assessment, student self-assessment, informal assessment, and the like (Davison & Leung, 2009; Savickiene, 2011; Veldhuis & van den Heuvel-Panhuizen, 2020). Alternative assessment methods generally encompass a complex of both practical and theoretical skills aimed at providing *qualitative* information about student performance, with the eventual goal of improving learning.

For effective learning to take place, a variety of assessment activities should be available, among them, self-assessment, peer assessment, and reflection as part of the learning



process. Teachers should evaluate students using “why,” “when,” “where,” and “how” questions based on their awareness of “where students should be headed to,” “where are they now,” and “how they are supposed to reach their destination.”

Involving learners in determining assessment criteria can increase their active involvement in and enhance their sense of responsibility for the learning processes (Li et al., 2012). When learners are active partners in the assessment process, they will better understand what goals they are aiming to attain (Veldhuis & van den Heuvel-Panhuizen, 2020). Being involved in defining clear criteria for evaluating learning and achievement helps students better understand what is required of them, while providing practical examples of how to meet the criteria (such as creating a suitable rubric) (Savickiene, 2011).

Given the above, many studies have indicated the need to change traditional assessment methods to multidimensional, formative methods that are curriculum-embedded, credible, and flexible and that allow teachers to provide students with feedback to improve their learning skills and achievements (Bedford & Legg, 2007; Dori, 2003). Formative assessment can enhance the development of learning skills, critical thinking, creative thinking, and more (Abali Öztürk & Şahin, 2014; Black & Wiliam, 2012).

*Student self-assessment.* Student self-assessment is another way to evaluate student achievement. Teachers can improve the correlation between their and their students assessments by teaching students how to conduct accurate self-assessment and understand the objectives of learning and assessment (Savickiene, 2011; Sung, 2005).

*Multiple methods.* Integrating a variety of assessment methods alongside traditional ones allows teachers to include broader curricular content in their assessment, expand their educational focus beyond merely learning for exams (Veldhuis & van den Heuvel-Panhuizen, 2020), and provide a comprehensive, in-depth portrait of learners’ achievements (Nevo, 2002,

2006). Ensuring a good definition of the criteria assessed by formative assessment depends on the teacher's professionalism: the more comprehensive the criteria, the easier it will be to achieve a more objective assessment (Davison & Leung, 2009; Savickiene, 2011). Moreover, alternative assessment allows teachers to establish interpersonal relationships with students and get to know them on a deep, personal level (Birenbaum et al., 2006; Darmody et al., 2020; Zhao et al., 2018).

### *Assessment in mathematics*

Evaluating achievement in mathematics requires more in-depth assessment methods as the discipline includes developing mathematical thinking, being able to properly represent mathematical concepts through precise notation, and developing problem-solving skills. The uniqueness of the discipline calls for integrating alternative ways of assessment (Cai et al., 2020; Kim & Noh, 2010; National Council of Teachers of Mathematics [NCTM], 2000, 2013).

A vast variety of mathematics assessment methods exist and include descriptive assessments (open questions, oral exams, activity reports, and interviews), observational assessment, and analysis of students' responses (Kulm, 2013; Watt, 2005; Zhao et al., 2018). Ideally, they will take into account individual background factors that, while unrelated directly to classroom learning (e.g., behavioral and medical conditions, language ability, special-education definitions), can affect learning (Cai et al., 2020; Mandinach, 2012). Research has shown that consistent use of descriptive assessment leads to more effective teaching and improves learners' attitudes (Kim & Noh, 2010; NCTM, 2000).

Five specific alternative assessment methods used in mathematics have been listed: (i) concept maps (students identify and point out connections between the various concepts they

have learned; (ii) peer assessment (helps develop meta-cognitive thinking and increase learners' self-awareness of their strengths and weaknesses in a subject); (iii) journal writing (encourages learners to identify new knowledge acquired and how it relates to previous knowledge); (iv) a portfolio of assignments and documents (shows investigation and learning, and the development of reflective and creative thinking); and (v) teacher observation (teachers assess the strategies students use, either individually or in interaction with their peers, and thereby better understand students' learning processes). Using any or all of these methods allows teachers to plan their follow-up teaching accordingly (Shahbari & Abu-Alhija, 2018).

Various evaluators can participate in alternative assessment: the MT, the students themselves, student peers, and even parents (Shahbari & Abu-Alhija, 2018; Zhao et al., 2018). In this context, a number of systematic tools have been developed to examine MTs' knowledge and pedagogical knowledge to determine if they are qualified for conducting professional diagnostics in the field (Saderholm et al., 2010).

#### *Advantages of alternative assessment*

The main advantages of alternative assessment methods in mathematics is that they can help improve the learning process, promote the development of students' personal potential, and improve the effectiveness in integrating educational processes (Cai et al., 2020; Davison & Leung, 2009; Galustyan, 2017).

Researchers tend to agree that although alternative assessment assignments are complex and require time and effort, given the advantages, it is important to encourage their use. The first step is to define the content area being evaluated and prepare appropriate indicators to assess it (Davison & Leung, 2009; Galustyan, 2017). Based on the importance of having the content of the assessment material represent the most current knowledge and understand-

ing of the discipline, it is important that experts (both in the discipline and in its assessment) confirm that student assignments are appropriate in terms of context, meaning, and educational value (van den Heuvel-Panhuizen, 2005), including making sure that the assignment reflects a sound learning process and is not merely a result of memorization (Schiefer et al., 2019).

It is agreed that alternative assessment (e.g., self-assessment, peer assessment, group assessment, and investigating sources of knowledge) facilitates teaching in a way that is appropriate for the learners' level and unique needs, which are better revealed by such assessment. This, in turn, enhances achievement and encourages increased involvement in the assessment process (alongside the learning process), leading to improved attitudes regarding self-efficacy, motivation, and perseverance and to the subject itself (Abali Öztürk & Şahin, 2014; Ashcraft, 2002; Ediger, 2013; Kulm, 2013; Popham, 2008; Rameau & Louime, 2007; Zhao et al., 2018).

However, effective alternative assessment requires MTs to have the professional skills to consider the variety of possible alternative assessment methods available, each with its advantages and disadvantages and to plan and execute appropriate assignments while ensuring that they are valid and reliable (Ediger, 2013).

#### *MTs' attitudes to alternative assessment*

Despite the advantages of alternative assessment, the integration and application of alternative assessments methods in mathematics has been slow in recent years, indicating the need for teacher support (Cai et al., 2020), since integration is greatly dependent upon MTs' attitudes toward mathematics' uniqueness (as a subject) on the one hand and toward alternative assessment on the other.

In fact, many MTs find it hard to apply the principles in class and are predominantly using traditional tests (Kulm, 2013; Shahbari & Abu-Alhija, 2018). An Australian study that examined the attitudes of 60 MTs in 11 high schools toward the application of alternative assessment in teaching mathematics concluded that the satisfaction MTs tend to display in traditional examination methods was the result of their conception of such methods as a valid assessment tool for measuring student achievement (Watt, 2005). Furthermore, a number of studies have shown that veteran teachers seem less inclined to adopt alternative assessment methods than novice teachers, presumably because they are accustomed to a culture of summative assessment using standard tests, and do not have the skills to adapt alternative modes of assessment to their teaching objectives (Li et al., 2019; 2020).

One possible reason for this is that teacher-education frameworks are not doing enough to give clear instruction on how to integrate alternative assessment, and schools do not provide support for teachers who face difficulties with alternative assessment (Levy-Vered & Nasser-Abu Alhiha, 2015; Stiggins, 1999). In fact, teachers have reported that the knowledge (regarding alternative assessment) imparted in their preservice training was sparse, practical experience was insufficient or nonexistent, and most importantly, their teacher-educators tended to model preferring traditional assessment methods over alternative ones, yet, on the other hand, the choice and quality of assessment methods may be affected by teachers' beliefs and concepts regarding assessment from their own experiences as students (Levy-Vered & Nasser-Abu Alhiha, 2015).

Furthermore, some studies have reported that there is a wide gap between MTs' self-perceived ability to assess students and their actual professional and practical ability to do so (Kulm, 2013; Levy-Vered & Nasser-Abu Alhiha, 2015; Shahbari, 2018).

In a Dutch study, researchers identified four profiles describing MTs' approach to alternative assessment: *enthusiastic assessors* – teachers who are familiar with a variety of alternative assessment methods, are aware of their potential contribution, and use them frequently; *mainstream assessors* – teachers who occasionally use alternative assessment tools; *non-enthusiastic assessors* – teachers who have a negative view of alternative assessment and use only a few and infrequently; and *alternative assessors* – those who have a vague view of assessment methods: on the one hand, they report extensive use of authentic assessment methods that they themselves develop, but on the other, they claim that they do not recognize the importance and need for this particular form of assessment (Veldhuis & van den Heuvel-Panhuizen, 2014).

The MTs' attitudes will, quite naturally, be greatly affected by any issues that concern teachers when called upon to choose assessment methods. Based on the literature, the predominant issues that teachers give as reasons for their resistance to alternative assessment methods are the following: lack of confidence in their reliability and validity; not enough knowledge regarding the methods; difficulties in implementation, especially when assessing large classes (i.e., the number of students requiring simultaneous assessment); obtaining autonomy in choosing their assessment tool; and the extensive investment in time and energy required to develop valid, reliable assessment tools (Al-Nouh et al., 2014; Cai et al., 2020; Kim & Noh, 2010; Stipek et al., 2001; Watt, 2005). They also cite a shortage of resources (alternative assessment is more expensive than a uniform test) and obstacles in planning, defining, and carrying out suitable assessment assignments.

*Supporting teachers in the crossover*

To encourage exchanging the exclusive use of traditional assessment methods with, at least some, alternative ones, MTs must be made more aware of the inherent potential such tools have for improving students' learning and achievements and then be given guidance and support for developing and using them (Chiang, 2015; Kim & Noh, 2010; Veldhuis & van den Heuvel-Panhuizen, 2020). In fact, based on findings pointing to the positive impact of alternative assessments on learning processes and in an attempt to respond to the concerns and difficulties that teachers have expressed, the NCTM (already in 2013) decided to encourage widespread use of alternative assessment by supplying additional, applicable information to MTs and S-MTs to improve teachers' ability to cope with the difficulties and challenges of alternative assessment methods (NCTM, 2013).

*Two types of challenges: Essential versus technical*

When discussing teachers' concerns regarding alternative assessment methods, it is important to distinguish between two primary types: essential and technical. Essential difficulties emerge from the nature of the assessment method and the professional knowledge and skills teachers need to master them (for example, having a profound understanding of the merits and deficiencies of different assessment methods or having the ability to establish the reasoning and a solid explanation why assessment choices and decisions were made). Technical difficulties are more related to the "toolbox" a teacher needs for such assessment (for example, proficiency in Excel or other spreadsheet software so as to be able to gather the data, or knowledge about how to provide credible, valid, and reliable assessment). "Essential difficulties require the intervention of external parties to help correct a gap in pedagogical or practical knowledge to rectify the difficulty; "technical" difficulties are those that can generally be easily rectified by providing some tool or basic skill.

### *Research goals*

As a result of the growing awareness of the importance of multiple assessment methods in mathematics education, there has been an increase in the volume of research on the topic (Cai et al., 2020; Kim & Noh, 2010; Veldhuis & van den Heuvel-Panhuizen, 2020; Watt, 2005; Zhao et al., 2018). However, these studies tend to focus on *comparing* traditional (which see the discipline as a static corpus of knowledge involving a series of actions and procedures and emphasize evaluating learners' basic knowledge and skills) and alternative methods (particularly those based on the constructivist approach, in which mathematics is seen as a tool for thinking and problem solving, with an emphasis toward investigation) (Stipek et al., 2001).

Based on the aforementioned, the authors of this present study wished to focus on the specific issues that concern MTs and S-MTs: to illuminate which most concern them and which they would like to see addressed in the near future, to map their frequency, and to understand whether there was a difference between the predominance of “essential” or “technical” concerns. The overall goal was to determine ways that these concerns may be alleviated and the assessment skills of teachers be improved.

### *Methodology*

*Research questions.* Based on the above goals, the following research questions were put forth:

1. What concerns do elementary and middle school MTs and S-MTs have with respect to student assessment, particularly with respect to alternative assessment methods?
2. To what degree do these concerns represent “essential” or “technical” difficulties?



3. At what frequency do elementary and middle school MTs and S-MTs encounter these difficulties?

### *Research design*

The study was designed to be qualitative so as to allow the researchers to discover unique issues and concerns expressed by the teachers relevant to their work in their own voice (Creswell & Poth, 2016; Kegler et al., 2019; Saldana & Omasta, 2018). It was divided into three stages as follows:

- Stage 1. Individual meetings held with MTs and S-MTs, during which demographic information was gathered and they were asked to write down their most immediate concern regarding student assessment.
- Stage 2. The participants from Stage 1 were divided into groups of four to six and asked to classify each of the concerns raised (by all the members of the group) as either “essential” or “technical” (see definitions, above) and analysis of differences based on two variables (MT/S-MT; elementary/ middle school).

Stage 3. Distribution of a two-part online questionnaire sent to a new group of MTs followed by analysis of the data collected to determine issues, classification, and frequency based on the variables. Part I asked for general background information (education, number of years teaching and teaching mathematics, background in assessment procedures, grades taught).

Part II comprised two open questions: (i) Is there any issue concerning student assessment (in mathematics) that concerns you as a mathematics teacher? What is it? (ii) Do you believe it warrants immediate help? Please explain.

### *Participants*

All the participants were enrolled in colleges and/or were teaching in Israel. Participants for Stages 1 and 2 (see below) included 43 MTs (16 elementary and 27 middle-school) enrolled in three MT professional development courses and 70 S-MTs (23 elementary and 47 pre-middle-school interns) studying in two academic courses on evaluating achievements in mathematics. Participants for Stage 3 included 84 MTs who answered a call for participants and who responded to the online questionnaire sent to them. (Note: 140 MTs answered the call, but only 84 returned the questionnaire.)

### *Results*

#### *Demographic inf*

The participants' demographic information are summarized in Tables 1 and 2.

*Table 1*

*Demographic Information for Stage 1 Participants (n=113)*

Group	School level:	Elementary school ( <i>f</i> = 39)	Middle school ( <i>f</i> = 74)
Teachers ( <i>f</i> = 43)	<i>f</i>	16	27
	% within group	37%	63%
	% within school level	41%	36%
Student teachers ( <i>f</i> = 70)	<i>f</i>	23	47
	% within group	33%	67%
	% within school level	59%	64%

Table 2

*Demographic Information for Mathematics Teachers Who Responded to the Self-report Questionnaire (n=84)*

		n	%
Teaching seniority (years)	1-3	9	10.7
	4-9	29	34.5
	+10	46	54.8
School level	Elementary school	79	94.0
	Middle school	5	6.0
Specialization in teaching mathematics	Yes	65	77.4
	No	19	22.6
Background in evaluating and assessing student achievement	None	18	21.4
	Academic or advanced study course	66	78.6

As seen in Table 2, over half of the teachers had ten years or more of teaching experience, and most worked in elementary schools, a relatively lower proportion than the statistics of the participants in Stages 1 and 2.

With respect to Research Question 2, essential difficulties, the dealing with which require outside intervention, were more frequent in all groups. In comparison to technical difficulties, for which dealing with do not require outside intervention, of the 113 problems raised by the participants, two-thirds were defined as essential, and only one third, technical. To better understand the attributes of the issues that concern teachers and S-MTs, in Stage 2 of the study, the 113 issues regarding the assessment of mathematics students, both essential and technical, were sorted into three main categories based on validity, reliability, and application.

*Stage 1: List of concerns*

The concerns given by the 113 participants regarding assessment of mathematics students could be sorted into three main categories: validity, reliability, and application (see Table 3 for distributions and Table 4 for examples).

*Table 3.*

Distribution of Categories of Assessment Concerns According to S-MT or MT

	Validity	Reliability	Application	Total
S-MTs	17	17	36	70
MTs	14	5	24	43
Total	31	22	60	113

*Table 4.*

*Examples of Issues Divided into Main and Subcategories*

Main Category (n = 113)	Subcategory	Examples
Assessment validity (n =31)	What is being evaluated	No possibility of calculations in ways that differ from those learned in class.
	Meaning of grade	Difficult to evaluate knowledge using numerical grades. Absence of verbal assessment and tools for ongoing improvement.
	Impact of teacher's instruction on the grade	The teacher who gives the test to evaluate achievements is not fully acquainted with the entire scope of the subject.
	Student variance	Evaluating diagnosed students in relation to the class. On the one hand – accommodations; on the other – in relation to the class

	Student's partial understanding	Student understands but the solution is wrong. Student gives correct solution but does not understand / cannot explain
	Weight of the student's emotional difficulties	Impact of stress (confusion)
Assessment reliability (n = 22)	Scoring is not consistent/uniform	Difficult to quantify the answer to each question according to its importance.
	Giving a grade that is not consistent/uniform	Recurring mistakes.
	Subjective assessment	Teacher's assessment is subjective (affected by previous acquaintance with the student).
Assessment applicability (n = 60)	Lack of comprehensive assessment tools	Difficult to measure and encompass all the material learned in an exam with a limited number of questions
	Numerical grades as norm	Children expect a grade expressed in numbers
	Providing effective feedback	Giving feedback in a way that will motivate the student to improve.

## Stage 2

Participants categorized the 113 issues as either “essential” (e.g., 97) or “technical” (e.g., 26). The results showed that approximately two-thirds were defined as “essential.” The distribution of essential/technical issues based on MT/S-MT and elementary/middle school is shown in Figure 1.

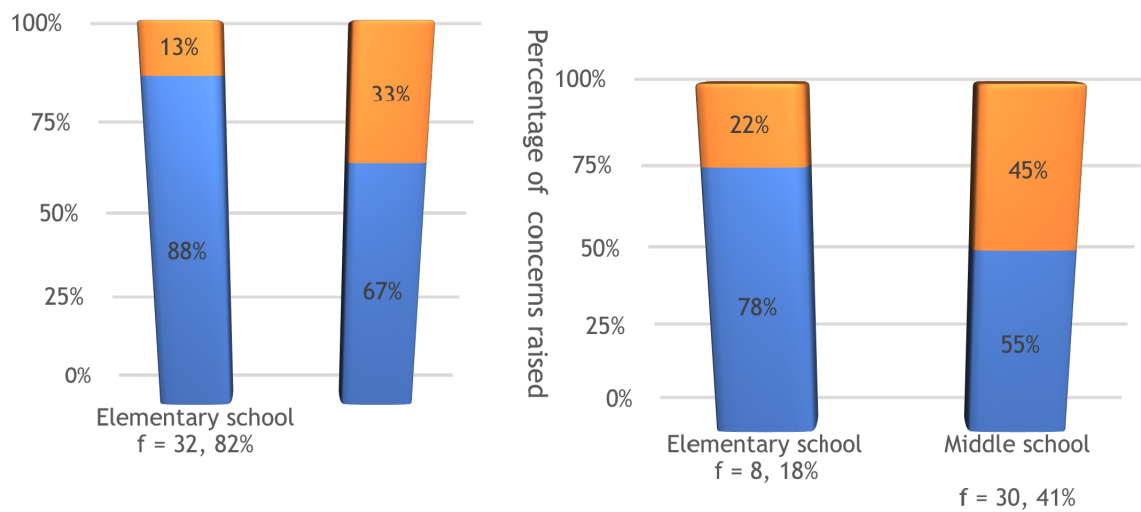
### Figure 1.

*Comparison of Number and Percentage of Essential and Technical Concerns Regarding Student Assessment by Type of Teacher (In-service/pre-service) and Type of School (Elementary/Middle)*

■ Essential concerns ■ Technical concerns

Math Teachers f = 43

Student teachers f = 70



Note that the issues that concern MTs and S-MTs are mostly essential (total 86%).

MTs and S-MTs seemed most concerned with difficulties regarding the validity and reliability of assessment, the challenge in making assessments in a valid way to overcome expecta-

tion of a numerical grade to reflect the learners' knowledge, and in providing effective feedback.

### Stage 3

*Open question 1.* The concerns expressed by the MTs could be divided into four main topics:

(i) class variance, (ii) exams (material covered, test anxiety, timing of exams), (iii) the Meitzav test (a nation-wide measure of school effectiveness and progress), and (iv) the gap between class assignments and test results. Examples for each are listed in Table 5.

Table 5

*Examples of Issues in Response to Question 1 in the Questionnaire, Sorted by Topic*

Topic	Examples
Class variance	<ul style="list-style-type: none"> <li>• How do I evaluate very weak students as compared to the very strong ones in the same class?</li> <li>• Difficulty in choosing assessment questions so that they will be appropriate for all levels of knowledge in class, including the mainstream students.</li> </ul>
Exams (material covered, test anxiety, timing of exams)	<ul style="list-style-type: none"> <li>• Should I give a quiz at the end of each topic?</li> <li>• Should I give an exam that covers several topics?</li> <li>• Does the learning require quantitative assessment?</li> <li>• With respect to graduated/accommodated exams – why do all students usually get the same version of the exam and if an accommodated one is given, how is it evaluated for the report card?</li> </ul>
Meitzav test	<ul style="list-style-type: none"> <li>• Is there an alternative assessment for the Meitzav? It puts pressure on the children, parents, and teachers....</li> <li>• The difficulty with the Meitzav tests, which gets greater every year, is how to cope with it. How do you advance the weak ones?</li> </ul>

---

Gap between class assignments and test results

- How do I incorporate the assessment of the classes given to individual students into the general assessment? At present, it is done using a planning-vs-performance file. But there is a gap between students' advancement in homework and ongoing assignments and their performance on exams.
  - How do you evaluate a student who works and does the assignments really well but does not succeed on the exam?
- 

*Open question 2.* The second question on the online questionnaire concerned which issues the MTs would like to get immediate help for. The answers revealed five main issues: (i) relationship between class size and difficulties in assessment; (ii) difficulties stemming from gaps between students in the class, especially the variance within all students in general and dealing with evaluating weak students in particular; (iii) possible gaps between exam scores and students' true abilities; (iv) issues involving need for external assessment; and (v) difficulties stemming from dealing with ensuring assessment reliability. (Note: Some of these concerns were also raised by the participants for Stages 1/2.)

*Class size.* A major concern was the number of students that must be taught simultaneously. Besides the issue of class management, it is also a concern from the perspective of having to taking into account students' emotional and cognitive needs (e.g., dyscalculia) when teaching the content. Note the following comments:

How do you give your full attention to "everyone" when teaching mathematics in a large group? I feel I want to "touch every student," be aware of each one's needs and difficulties, and adapt assignments as required. But the large number of children in class just won't allow it. Even a supposedly "regular" class is full of complexity, both from the emotional angle and from the cognitive, social, and academic angles (elementary school MT, 10+ years experience).

I'd choose the issue of a class populated by so many students. Splitting the class into learning levels in mathematics could contribute a lot. This way, stu-



dents will move ahead with the material [at their appropriate rate]. Furthermore, the love of mathematics will grow and barriers will be lowered. A discourse between groups of similar levels will be created (elementary school MT).

On the one hand, quantitative assessment doesn't assess students' personal progress; on the other hand, it's very difficult to apply qualitative assessment in a class of 30 students because you have to be involved in each and every student's process, and that's far from simple (elementary school S-MT).

*Differences between student abilities.* Teachers mentioned the gaps in student abilities and the resulting accommodations for testing that weaker students require that affect their test results.

*"I teach math to weak students, and their forgetfulness is very problematic. It's hard to find the way to evaluate what really remains after time"* (novice middle MT); *"I need more tips on how to advance intermediate students and those with difficulties to a higher level"* (elementary and middle MT, 10+ years experience).

*Gaps between test scores and actual knowledge.* An "essential" concern addressed the possible gap between grades received on exams, which are expected to reflect the student's knowledge, and student's true capabilities: *"There are situations when the student knows and understands the material learned, but their exam grade will be low and not reflect the level of knowledge. I chose this issue because it really concerns me"* (novice elementary MT).

A student may understand the material in class but when it comes to the exam, they fail or get a low grade, which represents a gap in relation to their knowledge. I'm really bothered by the situation of a student who demonstrates good comprehension of the material when I sit with him, but then fails the exam. When the exam is over, that same student, having had no further explanations, can sit and successfully solve all the questions they got wrong (novice elementary MT).

*The need for external, periodic assessment* (as opposed to internal assessment, which is performed regularly). For example:

I would like someone to explain to me why we need the *Meitzav* tests. The children are stressed by the test, and the test doesn't reflect their level of general knowledge. In my opinion this test is unnecessary, and its results just create competition among schools without ever really examining the background of the students who took the test with respect to socioeconomic level, learning disabilities, etc. (elementary MT, 10+ years of experience).

*Reliability of assessment.* This was an important issue. “*How can we know if students did their work on their own, without help. Many students have private tutors who do their homework for them at a very high level that is not manifested in the student's performance in class*”; “*How do we evaluate students when they work in pairs? Did the student work with a buddy?*” (novice middle MT).

### *Discussion*

The goal of the present study was to examine what assessment issues concern MTs and S-MTs and whether they were *essential* or *technical* issues. The findings suggest that there is a need to provide MTs and S-MTs with training and professional knowledge on alternative assessment methods in addition to the traditional ones so as to address assessing the unique attributes of mathematical thinking and solving mathematical problems and exercises.

The findings revealed that MTs and S-MTs are mostly concerned with *essential* assessment issues, that is, those requiring intervention on the part of additional professional bodies, and are less occupied by *technical* issues, which they feel they can handle on their own. Professional bodies include out-of-school experts (e.g., from the academic world or the Ministry of Education) or in-school assessment experts (e.g., mathematics coordinator, assessment coordinator, pedagogy coordinator).

Most of the concerns of elementary schools MTs and S-MTs were essential (80%) ones. In middle school, however, the proportion of participants concerned with essential issues was lower: about two-thirds of MTs and about half of S-MTs. This tendency may indicate that middle-school teachers and S-MTs have accrued more assessment skills than their peers in elementary school.

An example of an essential difficulty that was expressed by an elementary school teacher concerned how to analyze students' thinking on their way to the solution of the problem: *"I would be happy to find a solution to the problem of students' line of thought, and not only the final answer. Whether the answer is correct or incorrect... what is the thought process that went through the student's head?"* For comparison, an example of a technical problem that was raised by a middle school teacher is, *"Can we have an oral exam for students who know the material but have difficulties expressing themselves?"*

The findings suggest that there are more assessment challenges in elementary schools, which may be explained by the fact that in (Israeli) elementary schools, students are not divided into learning levels. The classroom is heterogeneous, as opposed to middle school, where mathematics classes are divided according to students' level (Harkabi & Mendel-Levy, 2014). Another possible explanation is that Israeli regulations require middle-school teachers to have a master's degree in teaching mathematics (not the case for elementary school teachers), meaning that most middle-school teachers will have had better training and hence better professional ability for evaluating student achievement. In this context, it seems important that the curriculum for training mathematics teachers should include knowledge of the sequence of learning mathematics so that elementary school teachers will know and understand the importance of the topics they are teaching for future learning, and that middle school teachers will be well acquainted with their students' previous knowledge.

Another interesting finding was that S-MTs, both in elementary and middle schools, reported a lower proportion of essential difficulties than did MTs. This may be explained by the fact that during their practice lessons, students are accompanied by teacher educators. If a difficulty arises, they can turn to the teacher educator (or other support system in the college), and immediately receive help to solve any issue when evaluating their students' achievements.

Comparing the results from Stages 1 and 3 showed similar responses with respect to in which areas of assessment they encounter difficulty with greatest frequency. This finding contributes to the validation of the questionnaire distributed in Stage 3.

The issues that the teachers indicated they were most interested in receiving a true, immediate response to concerned class size, student heterogeneity, evaluating thought processes when solving mathematical problem, ways to measure achievement other than exams, assessment reliability and validity. Other issues included possible gaps between the grade awarded through assessment (e.g., tests) and the students' true abilities and performance on class assignments. These corroborate findings in previous studies that found that most of the issues of concern to MTs relate to assessment credibility and validity, lack of knowledge about how to use alternatives to traditional assessment, and class size. However, these studies also revealed areas not raised in the present study, such as teachers' autonomy to choose an assessment tool and the immense professional investment and time required to develop and validate new assessment tools (Al-Nouh et al., 2014; Kim & Noh, 2010; Stipek et al., 2001; Watt, 2005).

The issues regarding class size and heterogeneity dealt with, among other things, how to deal with the vast gaps in the mathematical abilities of the students in a class; how to evaluate the achievements of student with special needs; how to evaluate the personal progress of

each individual student; the content and pedagogical knowledge required to accommodate assignments and exercises for different students; what tools and knowledge are needed to work with students who are having difficulties without slowing down the rest of the class; and more. It should be emphasized that the finding that issues of student variance and gaps in learning levels are of concern to teachers indicates that they understand the importance of providing differential responses to each student and their desire to undergo professional development in that area.

Teachers also indicated the importance of being able to respond to students' emotional needs, especially special-needs students in integrated classes and those with math anxiety (Geist, 2010). (Note that Israeli law mandates the inclusion of students with special needs in regular classes, which leads to high level of heterogeneity in the classroom [Knesset, 2002]). This also explains why MTs expressed concern with external assessments to which, they claim, students often exhibit test anxiety, especially in mathematics, meaning that they are aware that the results of external assessment do not necessarily indicate the students' true level of mathematical knowledge, given that such exams do not consider the examinees' backgrounds (socioeconomic status) or whether they have learning disabilities or special needs which will effect their grades, as noted by studies that have dealt with the effect of students' personal data. (These are not necessarily related to the class itself, and can include issues regarding behavior, medical status, language abilities, and the inclusion of special-education students in class, Cai et al., 2020; Mandinach, 2012.)

Teachers also demonstrated concern regarding the validity and reliability of mathematics assessments, including issues related to the essence of the content requiring assessment; appropriate modes of assessment; ways of ascertaining and confirming the assessment's validity and objectives; which topics should be included in the assessment; ascertain-

ing whether the assessment assesses only the students' work and not the contribution of people who help them (parents, private tutors, etc.); how to evaluate partial (incomplete) work, unclear assignments, or mistakes that are carried through to the solution; how to evaluate abilities and achievements that cannot be measured or are not unequivocal, such as motivation, amount of investment, and so on. These issues regarding validity and reliability of assessment in mathematics are consistent with previous studies who also noted that one reason that teachers tend to resist the integration of alternative assessment assignments in their work is the difficulty of establishing their validity and reliability (Davison & Leung, 2009; Kulm, 2013; Levy-Vered & Nasser-Abu Alhija, 2015; Li et al., 2019; Savickiene, 2011; Veldhuis & van den Heuvel-Panhuizen, 2014, 2020, Watt, 2005).

#### *Designing appropriate assessment methods*

Issues regarding knowledge of assessment methods in mathematics were raised when teachers pondered the question of the gap between the knowledge students exhibit in class and the knowledge expressed in the exam. This concern indirectly represents a lack of assessment knowledge and the fact that teachers feel they lack professional assessment tools that can help them deal with these gaps. Similarly, the numerical grade produced by mathematics assessment does not give room for students to improve and correct themselves, because, without detailed verbal feedback about every part of the solution and for each stage, students cannot discern their errors and mistakes, and consequently, they do not know on what they should focus to improve and progress. This means that in addition to the numerical grade, there is a need for qualitative measuring tools, and this need is shared by teachers and students alike.

In fact, MTs *are* asking for a rational picture of the situation that will reflect the students' level of understanding of the material learned and the degree of their investment in learning processes. The teachers' expectations that assessment be valid, reliable, unequivocal, and precise is affected by the unique nature of mathematics and the need it imposes for high precision in any exercise or problem. Teachers expect that the assessment will produce a numerical grade that reflects learners' knowledge and provide effective feedback. Therefore, the subject of alternative assessment in mathematics requires specific in-service training for MTs and special training for S-MTs as part of their studies.

In fact, the findings reveal that MTs and S-MTs are eager to get answers to a variety of practical questions regarding assessment. Similarly, other researchers have emphasized the need to use and develop various assessment methods that can support mathematics education (Chiang, 2015; Veldhuis & van den Heuvel-Panhuizen, 2020). This requires guiding and supporting MTs and S-MTs and focusing on developing, validating, and applying varied ways of assessment (Kim & Hon, 2010; Veldhuis & van den Heuvel-Panhuizen, 2020).

Another factor to be considered is involving students in designing and performing alternative assessments so that they understand why and how they are being evaluated, thus giving them responsibility in setting their own course of learning (Chiang, 2015). Peer assessment contributes to the development of students' meta-cognitive thinking, which is very valuable for learning mathematics and for increasing their self-awareness of their own strengths and weakness in their studies of mathematics (Shahbari & Abu-Alhija, 2018). It has thus been recommended that peer assessment should be incorporated (in addition to various other alternative assessments methods) into the assessment protocol.

We may conclude that alternative assessment for mathematics must be constructed professionally so as to complement the lacunae of quantitative assessment and be based on

clear, pre-established criteria. Clear criteria for assessment of learning – at all stages – must be defined, as must be students’ achievements using various methods of examination (periodic quizzes for formative assessment, exams for summative assessment, external assessments, etc.). This will ensure clarity about what is demanded of teachers and the broad criteria for professional, objective assessment.

### *Study contribution*

The findings of this study should be made known to decision-makers in the education system, MT supervisors in the various educational levels, and teacher educators in universities and colleges who train S-MTs or provide professional development to MTs. The findings can form a foundation on which specific, focused, and practical alternative-assessment training programs (in mathematics) can be built.

This is particularly necessary in the case of MTs because despite the clear advantages of alternative assessment methods, many MTs have difficulty applying their principles in the classroom (Cai et al., 2020; Kulm, 2013). In fact, studies have shown that, in most cases, MTs still tend to use traditional quantitative exams that require uniform solutions, although these solutions are not necessarily related to the students’ learning experience and specific context (Shahbari & Abu-Alhija, 2018; van den Heuvel-Panhuizen, 2005). Therefore, it is recommended that programs for training MTs, in-service professional development, guidance, etc., all be based on the findings from specific surveys that examine MTs’ needs (Shriki, 2013), similar to the questions in the present study.

It is important to emphasize that although alternative assessment provides teachers and students with reliable information about knowledge, abilities, and learning and thinking processes, at times, it places more emphasis on the way assignments were performed and the



time and means invested in doing so, and does not necessarily “grade” the outcome (i.e., what knowledge and problem-solving abilities were gained).

Providing MTs with assessment training and practice could improve their teaching-learning processes, their assessment abilities, and support more efficient use of both quantitative and qualitative assessment findings (Levy-Vered & Nasser-Abu Alhija, 2015).

The present study can make a practical contribution in developing curricula that includes not only teaching and assessment methods but also relate to the assessment difficulties teachers might encounter (such as, student variance with respect to mathematical ability, level of cognitive thinking, language and reading comprehension, parents’ education and ability to help their children with their mathematics studies, and even the help that the children get [or do not get] at home), some of which MTs and S-MTs are not necessarily aware.

A practical way to advance these abilities is to guide and instruct MTs and S-MTs on ways to systematically gather data (quantitative and qualitative) over time and use them to evaluate and improve their own mathematics teaching, while also advancing their students’ level of comprehension and improving their thinking skills (Cai et al., 2020). For this process, it is possible – even desirable – to involve the students too. Systematic feedback that teachers themselves produce on their teaching alongside their students’ learning could improve teachers’ abilities to better adapt their level of teaching to the learners’ levels and their real needs, thus contributing to improving achievement. This conclusion is strongly supported by findings in previous studies that found that learners’ active participation in assessment processes, alongside their participation in learning processes, can improve their achievements in mathematics and their positive attitude toward the discipline (Abali Öztürk & Sahin, 2014; Ediger, 2013; Kulm, 2013; Zhao et al., 2018).

### *Ideas for further research*

Further research that examines the unique assessment needs for mathematics would expand the knowledge gleaned in the present study and contribute to a relevant, authentic response for professional, high-quality, valid, and reliable assessment methods.

The questionnaire used in the present study can be distributed to teachers prior to and following a focused in-service professional development course about evaluating achievements and better modes of assessment in mathematics. This would facilitate examining the effectiveness of the in-service course as well as the overall effectiveness of MT training in general.

The school assessment coordinator could be called on to help with assessment for mathematics studies by providing precise answers for needs as they arise in situ.

Expanding in-service courses for teachers, and providing courses aimed at bettering assessment methods in mathematics, as well as providing individual and group guidance, can provide a foundation for improving and advancing teachers' ability to use varied means of assessment, adapt them to their needs, and strengthen personal and interpersonal relationship with students.

### *References*

- Abali Öztürk, Y., & Şahin, Ç. (2014). The effects of alternative assessment and evaluation methods on academic achievement, persistence of learning, self-efficacy perception and attitudes. *Journal of Theory and Practice in Education (JTPE)*, 10(4), 1022-1046.
- Al-Nouh, N. A., Taqi, H. A., & Abdul-Kareem, M. M. (2014). EFL primary school teachers' attitudes, knowledge and skills in alternative assessment. *International Education Studies*, 7(5), 68-84. <http://doi.org/10.5539/ies.v7n5p68>

- Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science*, 11(5), 181-185. <https://doi.org/10.1111/1467-8721.00196>
- Bedford, S., & Legg, S. (2007). Formative peer and self-feedback as a catalyst for change within science teaching. *Chemistry Education Research and Practice*, 8(1), 80-92. <https://doi.org/10.1039/B6RP90022D>
- Birenbaum, M., Breuer, K., Cascallar, E., Dochy, F., Dori, Y., Ridgway, J., Wiesemes, R., & Nickmans, G. (2006). A learning integrated assessment system. *Educational Research Review*, 1(1), 61-67. <https://doi.org/10.1016/j.edurev.2006.01.001>
- Black, P., & Wiliam, D. (2012). Assessment for learning in the classroom. In J. Gardner (Ed.), *Assessment and learning* (2nd ed., pp. 11-32). Sage. <http://dx.doi.org/10.4135/9781446250808.n2>
- Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., Cirillo, M., Kramer, S. L., & Hiebert, J. (2020). Timely and useful data to improve classroom instruction. *Journal for Research in Mathematics Education*, 51(4), 387-398. <https://doi.org/10.5951/jresematheduc-2020-0056>
- Chiang, W. W. (2015). Ninth grade student' self-assessment in science: A Rasch analysis approach. *Procedia - Social and Behavioral Sciences*, 176, 200-210. <https://doi.org/10.1016/j.sbspro.2015.01.462>
- Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). Sage.
- Darmody, M., Lysaght, Z., & O'Leary, M. (2020). Irish post-primary teachers' conceptions of assessment at a time of curriculum and assessment reform. *Assessment in Education: Principles, Policy & Practice*, 27 (5), 501-521. <https://doi.org/10.1080/0969594X.2020.1761290>
- Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, 43(3), 393-415. <https://doi.org/10.1002/j.1545-7249.2009.tb00242.x>

- Dori, Y. J. (2003). From nationwide standardized testing to school-based alternative embedded assessment in Israel: Students' performance in the "Matriculation 2000" Project. *Journal of Research in Science Teaching*, 40(1), 34-52. <https://doi.org/10.1002/tea.10059>
- Ediger, M. (2013). Teacher observation to evaluate mathematics achievement. *Delta-k*, 51(1), 4-6.
- Galustyan, O. V. (2017). Some methodological aspects of the evaluation of students' educational achievements at university. *International Journal of Cognitive Research in Science, Engineering and Education*, 5(1), 43-48. <https://doi.org/10.5937/IJCRSEE1701043G>
- Geist, E. (2010). The anti-anxiety curriculum: Combating math anxiety in the classroom. *Journal of Instructional Psychology*, 37(1), 24-31.
- Harkabi, A., & Mendel-Levy, N. (Eds.) (2014). *Education for all – and for each one in the Israeli education system*. Israel Academy of Sciences and Humanities. <http://education.academy.ac.il/SystemFiles/23011.pdf> (Hebrew).
- Kegler, M. C., Raskind, I. G., Comeau, D. L., Griffith, D. M., Cooper, H. L., & Shelton, R. C. (2019). Study design and use of inquiry frameworks in qualitative research published in health education & behavior. *Health Education & Behavior*, 46(1), 24-31. <https://doi.org/10.1177/1090198118795018>
- Kim, M. K., & Noh, S. (2010). Alternative mathematics assessment: A case study of the development of descriptive problems for elementary school in Korea. *Eurasia Journal of Mathematics, Science and Technology Education*, 6(3), 173-186. <https://doi.org/10.12973/ejmste/75238>
- Knesset (2002). Special Education Law (Amendment 7), 5763-2002. *The law book*, vol. 1876, pp. 90-92. [http://fs.knesset.gov.il/15/law/15\\_lsr\\_300594.pdf](http://fs.knesset.gov.il/15/law/15_lsr_300594.pdf)
- Kulm, G. (2013). Back to the future: Reclaiming effective mathematics assessment strategies. *Middle Grades Research Journal*, 8(2), 1-10
- Levy-Vered, A., & Nasser-Abu Alhija, F. (2015). Modelling beginning teachers' assessment literacy: The contribution of training, self-efficacy, and conceptions

- of assessment. *Educational Research and Evaluation*, 21(5-6), 378-406.  
<https://doi.org/10.1080/13803611.2015.1117980>
- Li, H., Xiong, Y., Hunter, C. V., Guo, X., and Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education* 45 (2), 193–211. [http://doi: 10.1080/02602938.2019.1620679](http://doi.org/10.1080/02602938.2019.1620679)
- Li, L., Fan, J., & Jin, Z. (2019). Comparing multimethod assessment of approaches to learning among preschool children: Direct measure, teacher report, and parent report. *Psychology in the Schools*, 56(8), 1271-1286. [https://doi.org/ 10.1002/pits.22274](https://doi.org/10.1002/pits.22274)
- Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist*, 47(2), 71-85. [https:// doi.org/10.1080/00461520.2012.667064](https://doi.org/10.1080/00461520.2012.667064)
- National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. MISSING INFORMATION
- National Council of Teachers of Mathematics (NCTM). (2013). *Formative assessment: A position of the National Council of Teachers of Mathematics*. [https:// www.nctm.org/uploadedFiles/Standards\\_and\\_Positions/Position\\_Statements/Formative%20Assessment1.pdf](https://www.nctm.org/uploadedFiles/Standards_and_Positions/Position_Statements/Formative%20Assessment1.pdf)
- Nevo, D. (2002). Dialogue evaluation: Combining internal and external evaluation. In D. Nevo (Ed.), *School-based evaluation: An international perspective* (pp. 3-16). Emerald. [https://doi.org/10.1016/S1474-7863\(02\)80004-7](https://doi.org/10.1016/S1474-7863(02)80004-7)
- Nevo, D. (2006). Evaluation in education. In I. F. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The SAGE handbook of evaluation* (pp. 442-460). Sage. [https:// doi.org/10.4135/9781848608078.n20](https://doi.org/10.4135/9781848608078.n20)
- Pellegrino, J. W. (2003). Knowing what students know. *Issues in Science and Technology*, 19(2), 48-52.
- Popham, W. J. (2008). Timed tests for tykes? *Educational Leadership*, 65(8), 86-87.

- Rameau, P., & Louime, C. (2007). Mathematics phobia: Are the mathematical sciences a pothole in the road of life? *Current Science*, *93*(11), 1481-1482 .
- Research Advisory Committee of the National Council of Teachers of Mathematics. (1988). NCTM curriculum and evaluation standards for school mathematics: Responses from the research community. *Journal for Research in Mathematics Education*, *19*(4), 338-344. <https://doi.org/10.2307/749544>
- Saderholm, J., Ronau, R., Brown, E. T., & Collins, G. (2010). Validation of the diagnostic teacher assessment of mathematics and science (DTAMS) instrument. *School Science and Mathematics*, *110*(4), 180-192. <https://doi.org/10.1111/j.1949-8594.2010.00021.x>
- Saldana, J., & Omasta, M. (2018). *Qualitative research: Analyzing life*. Sage.
- Savickiene, I. (2011). Designing of student learning achievement evaluation. *Quality of Higher Education*, *8*, 74-93.
- Schiefer, J., Golle, J., Tibus, M., & Oschatz, K. (2019). Scientific reasoning in elementary school children: Assessment of the inquiry cycle. *Journal of Advanced Academics*, *30*(2), 144-177. <https://doi.org/10.1177/1932202X18825152>
- Shahbari, J. A., & Abu-Alhija, F. N. (2018). Does training in alternative assessment matter? The case of prospective and practicing mathematics teachers' attitudes toward alternative assessment and their beliefs about the nature of mathematics. *International Journal of Science and Mathematics Education*, *16*(7), 1315-1335. <https://doi.org/10.1007/s10763-017-9830-6>
- Shriki, A. (2013). A model for assessing the development of students' creativity in the context of problem posing. *Creative Education*, *4* (7), 430-439. <https://doi.org/10.4236/ce.2013.47062>
- Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational measurement: Issues and Practices*, *18*(1), 23-27. <https://doi.org/10.1111/j.1745-3992.1999.tb00004.x>

- Stipek, D. J., Givvin, K. B., Salmon, J. M., & MacGyvers, V. L. (2001). Teachers' beliefs and practices related to mathematics instruction. *Teaching and Teacher Education, 17*(2), 213-226. [https://doi.org/10.1016/S0742-051X\(00\)00052-4](https://doi.org/10.1016/S0742-051X(00)00052-4)
- Sung, Y. T., Chang, K. E., Chiou, S. K., & Hou, H. T. (2005). The design and application of a web-based self- and peer-assessment system. *Computers & Education, 45*(2), 187-202. <https://doi.org/10.1016/j.compedu.2004.07.002>
- van den Heuvel-Panhuizen, M. (2005). The role of contexts in assessment problems in mathematics. *For the Learning of Mathematics, 25*(2), 2-23.
- Veldhuis, M., & van den Heuvel-Panhuizen, M. (2014). Primary school teachers' assessment profiles in mathematics education. *PLoS ONE, 9*(1), Article e86817. <https://doi.org/10.1371/journal.pone.0086817>
- Veldhuis, M., & van den Heuvel-Panhuizen, M. (2020). Supporting primary school teachers' classroom assessment in mathematics education: Effects on student achievement. *Mathematics Education Research Journal, 32*, 449-471. <https://doi.org/10.1007/s13394-019-00270-5>
- Watt, H. M. G. (2005). Attitudes to the use of alternative assessment methods in mathematics: A study with secondary mathematics teachers in Sydney, Australia. *Educational Studies in Mathematics, 58*, 21-44. <https://doi.org/10.1007/s10649-005-3228-z>
- Zhao, X., van den Heuvel-Panhuizen, M., & Veldhuis, M. (2018). Chinese primary school mathematics teachers' assessment profiles: Findings from a large-scale questionnaire survey. *International Journal of Science & Mathematics Education, 16*(7), 1387-1407. <https://doi.org/10.1007/s10763-017-9841-3>