

Assessment literacy: The process of evaluating early-career teachers and its association with assessment self-efficacy

Keren Tal-Amsili, Avidan Milevsky, Aleksandra Gerkerova* and Nitza Davidovitch

Education Studies, Ariel University, POB 3, Kiryat Hamada, Ariel, Israel 40700, Israel.

Accepted 30 July, 2024

ABSTRACT

Assessment is an important skill acquired by graduates of higher education, especially those from teacher education programs. This pilot study examined the effects of formative assessment processes on early-career teachers, focusing on their perspectives on student evaluation, their perceived self-efficacy in assessment, and the assessment methods they chose for their students. Participants, who were 152 teachers interning at elementary schools, completed self-report questionnaires midway through and at the end of their internship year. Out of these, 98 teachers completed the questionnaire at both time points. Key findings showed that assessment self-efficacy increased by 9.6% and perceptions of assessment as a formative process also rose significantly by 10%. The study suggests that formative assessment during teacher internships should be reinforced to enhance teachers' use of assessment in their classrooms. Additionally, it highlights the importance of improving assessment processes during internships, enabling teachers to maximize the benefits of assessments. This improvement not only optimizes their professional skills but also enhances their ability to conduct formative assessments among their students.

Keywords: Assessment, higher education, self-efficacy, early-career teachers, teaching internships.

*Corresponding author. E-mail: gerkerova@ukr.net..

INTRODUCTION

In recent decades, educational approaches have shifted from traditional teacher-led instruction to a constructivist paradigm, where teachers guide students in constructing knowledge (Libman, 2013; Levy-Feldman, 2020). Research by Vygotsky (1983) and Bruner (1987) emphasizes the importance of social interaction and discovery learning, which are central to constructivist theories. However, the application of constructivist approaches to student evaluation has received less attention. Summative assessment, aligned with the traditional-positivist approach, remains widely used to grade and report student performance (Birenbaum, 2018). In contrast, formative assessment focuses on providing feedback to enhance learning, promoting student motivation and self-efficacy (Yan and Brown, 2021).

Despite its benefits, formative assessment is not widely utilized.

Teachers' use of formative assessment is influenced by their beliefs about student evaluation. Research identifies four main beliefs: 1) formative assessment improves learning efficiency; 2) summative assessment fosters accountability (Opre, 2015); 3) summative assessment allows for comparison of achievements; 4) assessment is irrelevant to teaching and learning, potentially causing anxiety and detracting from self-esteem (Harris et al., 2008).

Teachers' assessment conceptions are shaped during their early training. In Israel, after completing academic studies, teachers undergo a one-year internship with mentorship and peer workshops. Midway through the

their licensure (Ministry of Education, 2020). Interns also self-evaluate their achievements.

While there is considerable research on teachers' beliefs about assessment, empirical studies specifically examining how these beliefs evolve during the early stages of teachers' careers are scarce. Furthermore, little is known about the impact of formative feedback received during the induction year on teachers' assessment self-efficacy. This pilot study examines how formative assessment during the induction year affects interns' perceptions of student assessment, their assessment self-efficacy, and their chosen assessment methods. Assessment self-efficacy, enhanced by mentor feedback, reflects interns' confidence in conducting assessments (Bandura, 1997). Understanding these variables is crucial for recognizing the value of different assessments and enhancing meaningful classroom assessments. Recent studies, such as those by Stanja et al. (2022-2023), highlight the potential of learning analytics to support formative assessment practices by providing detailed insights into students' conceptions and learning processes. By addressing these gaps, this study aims to provide empirical evidence on the evolution of assessment conceptions and self-efficacy among early-career teachers, highlighting the practical implications for teacher education programs. Insights from mid-year assessments may suggest improvements for optimizing intern skills and conducting formative assessments with students.

Hopfenbeck et al. (2023) emphasize the importance of integrating AI to overcome challenges in implementing formative assessment and enhancing personalized feedback. AI can assist in providing timely, individualized feedback and managing the workload associated with formative assessments, thus supporting teachers in their instructional practices. These insights further underscore the necessity of incorporating advanced technologies in teacher education programs to improve assessment literacy and efficacy.

Enhancing teachers' assessment literacy and self-efficacy through formative assessment can lead to more effective teaching practices and improved student outcomes. By providing a clearer understanding of these dynamics, this study aims to inform teacher education programs and policymakers about the importance of supporting formative assessment practices during the induction phase.

Teacher induction programs worldwide

The induction phase is crucial for shaping teachers' professional identity but is also one of the most challenging periods in their careers. New teachers often face both professional and personal difficulties, such as managing classes, planning curricula, and evaluating students with limited experience. They also experience significant pressure and anxiety due to workload, role conflicts, and a

sense of loneliness, exacerbated by the gap between their expectations and reality (Anthony et al., 2019). These challenges lead many novice teachers to leave the profession within their first year (Sutcher et al., 2019).

To address these issues, many education systems have developed teacher induction programs. These programs support early-career teachers by bridging the gap between their studies and actual teaching. They aim to enhance professional knowledge, pedagogic skills, and class management abilities, while also familiarizing new teachers with the educational system and school culture.

Typically lasting about a year, induction programs include individual mentorship, which involves support from veteran teachers who model teaching practices, share responsibilities, and engage in regular interactions (Ovadia, 2018; Whitaker et al., 2019). Effective mentorship includes professional training for mentors and encourages reflective practices, fostering productive dialogue between mentors and inductees (Zilbershtrom, 2013; Guterman, 2014).

Mentors help novice teachers integrate into the school community, promoting collaboration and responsible membership. Additional peer support allows new teachers to discuss challenges in a safe environment, further enhancing their coping strategies.

Entering teaching: Israel's teacher induction program

Two major differences can be identified between the induction programs in Israel and other countries. First, in Israel, training institutions are more involved in the induction process by operating workshops and handling administrative tasks with the Ministry of Education. In most other countries, such involvement is minimal or non-existent (Nasser-Abu Alhija et al., 2011).

In Israel, the induction phase is a distinct training process due to the unique characteristics of both the teacher and the educational system (Zilbershtrom, 2013). This phase, outlined in the Ministry of Education's Director General circular, lasts three years, with the first year being the induction year. Successful completion grants a "license to engage in teaching" and the status of "novice teacher," akin to professions like medicine and social work (Sperling, 2015).

Teacher inductees are graduates of teacher training programs employed in the education system in at least one-third of a full-time position (Ministry of Education, 2020). They fully integrate into school staff and undertake all teaching responsibilities, including interactions with supervisors, principals, teachers, students, and parents (Ovadia, 2018).

The induction program has two key components: a 60-hour peer workshop at a teacher training institution and personal mentoring at the school (Ministry of Education, 2020). The workshops, led by a training institution moderator, provide group support and address both

emotional and professional issues through the analysis of real-world events (Zilbershtrom, 2013).

Personal mentoring, the second component, is provided by an experienced teacher at the school. Mentorship supports inductees emotionally, helping them manage challenges and reduce feelings of loneliness and pressure (Whitaker et al., 2019). The mentor also introduces the school's organizational culture and aids professional socialization by cultivating teaching skills and effective practices (Schatz-Oppenheimer et al., 2014).

The mentor's role involves balancing support with evaluation for licensure purposes (Schatz-Oppenheimer et al., 2014). Mentors meet inductees weekly, focusing on overcoming work difficulties in the first semester and developing reflective thinking in the second. Assessment and feedback are crucial, with formative assessment mid-year and summative assessment at the end, which influences licensure decisions. This dual role requires mentors to balance emotional support and critical evaluation (Schatz-Oppenheimer et al., 2014).

Assessment processes within teacher induction programs

The literature distinguishes between two major approaches to evaluating teachers based on the assessment's aims: formative assessment and summative assessment. Formative assessment is designed to help teachers and enhance their ability to perform their roles, which is particularly significant due to their lack of professional experience (Sperling, 2016). The most important component of this type of assessment is the provision of detailed, informative feedback on how the inductee or novice teacher meets standards, defined requirements, or measurable goals. The most effective feedback is that which is provided regularly and continuously, highlighting strengths in inductees' teaching practices, areas needing improvement, and specific recommendations for enhancing their work.

Effective feedback for novice or inductee teachers has been found to enhance their motivation at work, improve their teaching skills and the quality of their teaching, reinforce their sense of teaching efficacy (Guterman, 2014), and improve their students' achievements. High-quality, constructive, and meaningful feedback can also significantly promote reflective processes that allow teachers to plan and achieve teaching goals (Taylor and Tyler, 2012). Additionally, it contributes to inductees' ability to critically observe and react to their teaching, which can improve their teaching practices and shape their teaching conceptions.

Beyond the significance of feedback in the process of inductee assessment, some studies focus on unique tools for evaluating inductees, such as using a personal portfolio. Many induction programs also include self-assessment components.

In contrast, summative assessment emphasizes accountability for the teacher's performance or screens and filters out teachers who are not suitable for teaching or who have a low chance of persevering beyond their first year. In many education systems, the assessment of novice teachers includes both formative assessment, which involves receiving effective feedback intended to improve their teaching skills, and summative assessment, which is intended to screen and filter out teachers who are not suitable for teaching (Sperling, 2016). In the case of inductees, summative assessment determines whether they will receive a license to engage in teaching.

Evaluating students at school

Evaluating students is a practice performed by teachers and/or their students. It provides evidence of students' achievements to examine the effectiveness of learning and teaching processes and student performance in the studied discipline (Birenbaum, 2018). This evaluation also helps in making decisions to shape future teaching-learning processes.

There are two main types of assessment processes that parallel those used for teacher inductees: summative assessment, also called assessment of learning, and formative assessment, which is assessment for the purpose of learning (Birenbaum, 2018). The summative assessment aims to evaluate the student's knowledge and understanding for grading, categorizing, or reporting (Yan and Brown, 2021) to stakeholders within or outside the school and for making classroom decisions. This type of assessment emphasizes results and grades, judging students against existing measurable standards (Levy-Feldman and Libman, 2022). If feedback is provided by the teacher, it is usually brief and limited and therefore does not significantly improve the student's learning process or their acquisition of effectiveness and independence. Summative assessment often disregards the learning process, treating teaching, learning, and assessment as separate entities. Typically, students are not involved in the assessment process, even though involving students can help them develop an awareness of its impact, as well as reflective abilities and self-assessment skills. These processes have been found to improve students' learning and achievements (Birenbaum, 2018).

The main aim of formative assessment, also called assessment for learning and considered a type of alternative assessment (Levy-Feldman and Libman, 2022), is to advance students' learning processes and enhance their effectiveness and independence (Yan and Brown, 2021). This type of assessment is usually informal and authentic, conducted in natural conditions that are linked to the student's world, utilizing an array of qualitative and alternative assessment techniques. Assessment tools include portfolios of student assignments (Birenbaum, 2018), tasks performed, projects, and exhibitions. The

formative assessment process often involves several stages: defining aims, goals, and expected results; planning milestones for achieving these goals; collecting data; gauging the gap between desired results and actual achievements; applying intervention programs to close these gaps; performing evaluations to examine the efficacy of the programs while stressing aspects for improvement and preservation; and finally, identifying additional needs and beginning the process anew (Friedman, 2020).

Despite the growing understanding that to advance students' achievements, teachers must develop formative assessment processes that include providing learning-promoting feedback and focusing on the learning process rather than just the products, the literature shows that the most common assessment method among schoolteachers remains summative assessment (Yan and Brown, 2021).

Assessment literacy refers to teachers' knowledge, skills, tools, and thorough understanding of various assessment processes. This includes knowing what each type evaluates, when and why different assessments are used, the meaning of their results, and the effectiveness or contribution of various methods, as well as their shortcomings and costs. Literacy in evaluating achievements is a major aspect of teacher training because, from the moment they become teachers, they need to know how to use these assessments effectively.

Assessment self-efficacy

One element that influences the assessment methods chosen by teachers, closely linked to the level of assessment literacy, is assessment self-efficacy. According to Bandura's (1977) Social Cognitive Theory, self-efficacy relates to individuals' beliefs regarding their ability to successfully perform specific tasks. A sense of efficacy can be developed through observing and learning from the experiences of others.

In education, a teacher's sense of assessment self-efficacy is affected by their level of assessment literacy and their previous experience with performing assessments. Studies examining the relationship between assessment self-efficacy and the application of assessment methods have found a positive correlation between high levels of assessment self-efficacy and the use of formative assessment (Yan and Brown, 2021). Conversely, a low level of assessment self-efficacy is associated with minimal use of alternative assessment methods. The assessment methods teachers choose to apply in the classroom, as well as their use of formative assessment, are influenced by their general perception of student assessment.

Teacher conceptions of student assessment

Teacher conceptions of student assessment relate to their

beliefs about the aims and effectiveness of different methods of evaluating student learning (Opre, 2015). A common distinction between teacher conceptions of student assessment differentiates among four possible assessment aims: improving teaching and learning; student accountability; accountability of teachers and schools; and the belief that assessment is irrelevant. These conceptions, except for the fourth, can be seen as existing on a continuum (Barnes et al., 2017), and teachers may hold several conceptions simultaneously.

The first conception ascribes considerable significance to assessing students' achievements in a way that contributes to each student by enhancing the efficiency of their learning process and adapting teaching methods to meet students' needs and address difficulties. This conception also includes the belief that formative assessment provides teachers with information to make their teaching more effective (Opre, 2015).

The second conception of assessment, identified by Brown (2006), views it as essential for promoting student accountability for learning. This conception focuses on measured achievements rather than the learning process (Opre, 2015).

The third conception of student assessment, which aligns with the second, sees assessment as a means to compare students and evaluate their achievements against measurable standards.

Teachers who subscribe to the fourth conception see no need for a structured assessment process, believing they can evaluate their students' understanding through direct interaction. This view holds that assessment negatively impacts teachers' autonomy and professionalism and distracts from students' learning. Moreover, this conception argues that assessment can harm students by causing anxiety and damaging self-esteem (Harris et al., 2008). It rejects the idea of assessment, contending that it is unreliable and more harmful than beneficial for both teachers and students.

Studies suggest a link between teachers' conceptions regarding evaluating achievements and their use of assessment in the classroom (Birenbaum, 2018). A positive association was found between teachers' favourable views of formative assessment for improving learning processes and making teaching decisions, and their intentions to apply such assessments in the classroom (Yan and Brown, 2021). However, Davis and Neitzel (2011) found that teachers often used assessment tools imposed by external authorities, contrary to their conceptions of assessment.

The current quantitative study utilizes a longitudinal research design to explore changes in teaching inductees' assessment conceptions over their induction year.

The current study has two main aims:

- to examine changes in inductees' conceptions of student evaluation over the course of their internship year.
- to identify factors that influence the assessment methods

inductees apply with their students by the end of the induction year.

The study is guided by three main research questions:

1. To what extent do the conceptions of teacher inductees on student evaluation and their assessment self-efficacy change over the course of the induction year?
2. Do the components of the assessment process that inductees were exposed to during their induction affect the assessment methods they utilize with their students?
3. Are teacher inductees' assessment self-efficacy and conceptions of student assessment associated with the assessments they choose to use with their students?

The research hypotheses:

- Over the course of the induction year, inductees' conception of assessment as a formative process will increase, while the conception of assessment as an irrelevant process will decrease.
- Inductees' assessment self-efficacy will increase over the induction year.
- An association will be found between teachers' conceptions of student assessment at the end of the year and their use of assessment methods in class. Stronger perceptions of assessment as a formative process at the end of the year will be positively associated with inductees' use of performance assessments and non-achievement-based grading, and negatively associated with the use of traditional exams, administering standard exams, and grading. A reverse pattern of associations will be found concerning conceptions of assessment as a summative or harmful/irrelevant process.
- An association will be found between assessment self-efficacy and the use of assessments in class. A positive association will be found between assessment self-efficacy at the end of the year and the use of formative assessment methods.

METHODOLOGY

Participants

Data was collected at two-time points during the induction year: after a formative assessment in January and after a summative assessment in June. At the first time point, participants included 152 elementary school teachers in their first year of full-time work in the state religious education system. The sample consisted of 119 women (78.3%), with ages ranging from 21 to 55 years ($M = 31.67$, $SD = 8.26$). Most participants were married (101, 66.4%) and had up to five children ($M = 1.38$, $SD = 1.40$). The majority held a Bachelor's degree.

At the end of the school year, the follow-up survey was completed by 98 of the original participants, resulting in a

dropout rate of 36%. This dropout rate aligns with typical rates observed in longitudinal studies (Goodman and Blum, 1996). A comparison of the demographic profiles of the two groups showed no statistically significant differences.

Procedure

Participants were recruited in November of the 2021/22 academic year through appeals to the heads of research authorities at colleges, heads of induction and teaching units, and instructors of internship workshops at colleges and universities. The recruitment process was approved by the Ethics Committee at the host university and the Chief Scientist's Office at the Ministry of Education.

Questionnaires were distributed online to participants who agreed to partake in both phases of the study. This study employed convenience sampling. Participants were informed that they would need to complete questionnaires at two-time points: the midpoint and the end of the year. To facilitate follow-up, participants provided their email addresses and the last four digits of their phone numbers, which allowed the researchers to match their responses across the two-time points. Participants were assured that their contact information would be used solely for the purpose of the study.

To examine the research variables, the following self-report questionnaires were utilized:

Background questionnaire

Participants were asked to provide several details: gender, age, marital status, number of children, level of education, function during the induction year (homeroom teacher and subject teacher, or only subject teacher), the town where the school was located, educational stream (state, state religious, ultra-orthodox), size of the school by the number of students (small – up to 200 students, medium – from 200 to 500 students, large – more than 500 students), average number of students in classrooms taught, number of classrooms taught in the induction year, level of classrooms taught (1st-2nd grade, 3rd-4th grade, 5th-6th grade, 7th-8th grade), extent of position at the school (full position, half position, third position), main subject taught, previous experience in teaching, type of training institution, and whether the participant had taken a course on assessment as part of the training. Additionally, participants were asked to indicate who was present in the feedback session (mentor, principal, mentor and principal, other staff member from the school), the number of observations conducted by the mentor and principal before the feedback session, and the frequency of encounters with the mentor during the first semester (once a week, once every two weeks, once every three weeks, once a month).

Conceptions of assessment questionnaire

To measure participants' conceptions of student assessment, a short version of the Conceptions of Assessment (CoA) questionnaire, developed by Brown in 2002 and revised in 2006, was used (Brown, 2006). This version, translated into Hebrew by Levy-Vered (2013), includes 21 statements reflecting seven common conceptions of the aim of assessment: improving learning, improving teaching (both formative), summarizing and describing learning, school accountability (both summative), validity and reliability, assessment as unfair and inaccurate, and assessment as irrelevant or harmful. Participants ranked their agreement with each item on a 6-point Likert scale, from 1 (strongly disagree) to 6 (strongly agree). Scores were calculated for each conception and overall by averaging the relevant items, with negative conceptions reversed. The Hebrew questionnaire demonstrated high construct validity and internal reliability, with Cronbach's alpha ranging from .70 to .85 (Levy-Vered, 2013). In the current study, reliability ranged from $\alpha = .72$ to $.89$.

Assessment self-efficacy

This questionnaire, developed by Levy-Vered (2013), examines inductees' assessment of self-efficacy. It comprises 20 statements representing specific assessment operations in five areas of assessment literacy: recognizing, choosing, and developing appropriate assessment tools; scoring and developing valid assessment criteria; interpreting assessment findings to make decisions; reporting assessment findings; and ensuring validity and ethics in assessment. Participants rated their ability to perform each operation on a 6-point Likert scale ranging from 1 (completely incapable) to 6 (completely capable). A total self-efficacy score was calculated for each participant by averaging the ratings of all items. Higher scores indicate higher assessment self-efficacy. In the original study, the internal reliability of the questionnaire was $\alpha = .97$ (Levy-Vered, 2013). In the current study, reliability was $\alpha = .95$ for the first measurement and $\alpha = .96$ for the second measurement.

Student assessments in class

To measure the use of various student assessment methods, the Assessment Practices Inventory (API) devised by Zhang and Burry-Stock (2003) and translated into Hebrew by Amasha (2014) was used. This questionnaire was administered only at the end of the year to avoid potential bias in responses.

The original version of the questionnaire examined both teacher conceptions of assessment and the degree to which assessment techniques were applied in practice. In

this study, the focus was solely on the inductees' use of assessment methods. According to factor analysis by Zhang and Burry-Stock (2003), the questionnaire items reflect six assessment components: administering traditional tests; administering standard tests, correcting tests, and improving teaching; reporting assessment results, ethics, and scoring; using performance assessment; non-achievement-based grading; and validity and reliability tests. Participants rated their use of these assessment components on a 6-point Likert scale from 1 (do not use at all) to 6 (use always). The short questionnaire developed by Amasha (2014) includes 34 items based on the same six factors as the original study. Scores were calculated for each participant for both total use and each factor. Amasha (2014) found an internal reliability of $\alpha = .93$ for the full questionnaire. In the current study, reliability for the questionnaire was: traditional exams, $\alpha = .67$; applying traditional assessment, $\alpha = .84$; implementing test results, $\alpha = .82$; utilizing performance assessment, $\alpha = .75$. After removing item 21, reliability for applying grades was $\alpha = .95$, and for applying validity, $\alpha = .79$.

Statistical methods

The data analysis employed both descriptive and inferential statistical methods. Descriptive statistics, including means and standard deviations, were used to summarize the demographic characteristics of the sample and the main study variables. To address the research questions and hypotheses, paired-sample t-tests were conducted to compare changes in assessment self-efficacy and conceptions of assessment between the two-time points. This method was chosen because it allows for the comparison of means from the same group of participants at different times, which aligns with the longitudinal nature of the study. Additionally, Pearson correlation coefficients were calculated to examine the relationships between assessment self-efficacy, conceptions of assessment, and the use of assessment methods. The choice of Pearson correlation was based on its suitability for assessing the strength and direction of linear relationships between continuous variables.

Potential limitations and biases

One potential limitation of the study is the use of convenience sampling, which may limit the generalizability of the findings. To mitigate this limitation, efforts were made to recruit a diverse sample of participants from various colleges and universities within the state religious education system. Another limitation is the reliance on self-report questionnaires, which may be subject to social desirability bias and inaccuracies in self-assessment. To address this, the questionnaires included validated scales

with established reliability and validity, and participants were assured of the confidentiality of their responses to encourage honest reporting. Furthermore, the dropout rate of 36% could introduce attrition bias. However, comparisons of demographic profiles between those who completed both surveys and those who did not showed no statistically significant differences, suggesting that attrition did not systematically bias the results.

RESULTS

Preliminary analyses

Quantitative data were collected at two time points: after the mid-year assessment (February-April 2022) and after the formative assessment (June-July 2022). Initially, the association between background variables and research variables was explored to determine if they should be included in some analyses. An ANOVA test was used to

compare groups with different categorical background characteristics for each research variable (Table 1).

A t-test for independent samples was conducted to compare teachers with previous teaching experience and those without. Significant differences were found in several areas:

- Regarding the conception of assessment as harmful, teachers with previous experience ($M = 4.83$, $SD = 0.90$) scored higher than those without previous experience ($M = 4.49$, $SD = 0.96$) ($t(96) = 1.70$, $p = .04$).
- For validity and reliability tests, teachers without previous experience ($M = 5.15$, $SD = 0.68$) scored higher than those with previous experience ($M = 4.74$, $SD = 1.00$) ($t(96) = -2.40$, $p = .01$).
- In reporting assessment results, teachers without previous experience ($M = 5.28$, $SD = 0.67$) scored higher than those with previous experience ($M = 5.02$, $SD = 0.71$) ($t(96) = -1.80$, $p = .04$).

Table 1. Differences between inductees with previous experience in teaching and those with no experience, in each of the research variables.

	Teachers with previous experience	Teachers with no previous	<i>t</i>
	(<i>N</i> = 37)	experience (<i>N</i> = 61)	
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	
Formative assessment –t1	4.50 (.93)	4.31 (1.04)	.90
Formative assessment –t2	4.65 (1.15)	4.88 (1.12)	-.90
Summative assessment –t1	3.79 (.89)	3.77 (.80)	.10
Summative assessment –t2	4.07 (.99)	4.29 (.93)	-1.10
Assessment as harmful –t1	4.83 (.90)	4.49 (.96)	1.70*
Assessment –as harmful - t2	4.63 (1.15)	4.98 (1.1)	-1.50
Assessment validity - t1	3.95 (1.15)	3.82 (1.29)	.50
Assessment validity –t2	4.27 (1.25)	4.59 (1.26)	-1.20
Assessment self-efficacy –t1	4.66 (.78)	4.57 (.85)	.50
Assessment self-efficacy –t2	4.97 (.85)	5.08 (.65)	-.70
Conducting traditional exams	4.06 (.68)	4.26 (.65)	-1.40
Administering standard exams	4.65 (.95)	4.78 (1.08)	-.60
Reporting assessment results	5.02 (.71)	5.28 (.67)	-1.80*
Use of performance assessments	4.75 (1.17)	5.03 (.85)	-1.30
Non-achievement based grades	4.93 (1.11)	4.96 (.97)	-.10
Validity and reliability tests	4.74 (1.00)	5.15 (.68)	-2.40*

* $p < .05$.

Additionally, a comparison was conducted between teachers who had taken a course on assessment during their studies and those who had not, using a t-test for independent samples. Table 2 presents this comparison.

Significant differences were found in several areas. Teachers who had taken a course in assessment used standard exams more frequently ($M = 4.99$, $SD = .77$) than those who had not taken such a course ($M = 4.45$, $SD = 1.21$) ($t(96) = -1.41$, $p = .80$). Similarly, teachers who had taken an assessment course reported assessment results

more frequently ($M = 5.32$, $SD = .60$) compared to those who had not ($M = 5.05$, $SD = 0.78$) ($t(96) = -1.80$, $p = .37$).

At the end of the year, the conception of assessment as formative was higher among teachers who had taken an assessment course ($M = 4.97$, $SD = 1.12$) than among those who had not ($M = 4.59$, $SD = 1.13$) ($t(96) = -0.93$, $p = .17$). The conception of assessment as harmful was also higher among teachers who had taken a course in assessment ($M = 5.07$, $SD = 1.05$) compared to those who had not ($M = 4.60$, $SD = 1.17$) ($t(96) = -1.45$, $p = .74$).

Furthermore, assessment validity at the end of the year was higher among teachers who had taken a course in

assessment ($M = 4.75$, $SD = 1.19$) than those who had not ($M = 4.15$, $SD = 1.28$) ($t(96) = -1.21$, $p = .11$).

Table 2. Differences between teachers who had taken a course in assessment and those who had not, in each of the research variables.

	Took assessment course ($N = 52$)	Did not take assessment course ($N = 46$)	<i>t</i>
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	
Formative assessment – t1	4.32 (1.14)	4.45 (.81)	-.67
Formative assessment – t2	4.97 (1.12)	4.59 (1.13)	1.67*
Summative assessment – t1	3.72 (.83)	3.84 (.84)	-.70
Summative assessment – t2	4.97 (1.12)	4.59 (1.13)	1.00
Assessment as harmful –t1	4.62 (.93)	4.62 (.99)	-.01
Assessment as harmful –t2	5.07 (1.05)	4.60 (1.17)	2.09
Assessment validity - t1	3.72 (1.38)	4.04 (1.03)	-1.28
Assessment validity –t2	4.75 (1.19)	4.15 (1.28)	2.23*
Assessment self-efficacy –t1	4.63 (.80)	4.56 (.86)	.41
Assessment self-efficacy –t2	5.15 (.62)	4.91 (.83)	1.56
Conducting traditional exams	4.13 (.65)	4.24 (.69)	-.86
Administering standard exams	4.99 (.77)	4.45 (1.21)	2.65*
Reporting assessment results	5.32 (.60)	5.04 (.78)	1.99*
Use of performance assessments	5.02 (.95)	4.81 (1.02)	1.02
Non-achievement based grading	5.00 (.98)	4.90 (1.06)	.45
Validity and reliability tests	5.10 (.85)	4.45 (.80)	1.33

* $p < .05$.

A comparison was conducted between teachers who studied at a college and those who studied at a university for each of the research variables, using t-tests for independent samples. Table 3 presents this comparison.

As evident from Table 3, the findings show that teachers who studied at a college displayed higher levels of

formative and summative assessment conceptions, assessment validity, and assessment self-efficacy at Time 2, as well as higher levels of reporting assessment results, use of performance assessments, and validity and reliability tests, compared to teachers who studied at a university.

Table 3. Differences between teachers who studied at a college and those who studied at a university, in each of the research variables.

	Studied at college ($N = 83$)	Studies at university ($N = 15$)	<i>t</i>
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	
Formative assessment – t1	4.39 (.96)	4.32 (1.22)	.27
Formative assessment –t2	4.91 (1.12)	4.13 (1.01)	2.71*
Summative assessment – t1	3.77 (.83)	3.80 (.87)	-.11
Summative assessment – t2	4.28 (.95)	3.78 (.90)	1.93*
Assessment as harmful – t1	4.57(1.00)	4.89 (.57)	-1.18
Assessment as harmful – t2	4.89 (1.18)	3.93 (1.08)	.83
Assessment validity –t1	3.84 (1.24)	4.00 (1.22)	-.43
Assessment validity –t2	4.56 (1.27)	3.93 (1.08)	2.02*
Assessment self-efficacy – t1	4.57 (.86)	4.76 (.55)	-.80
Assessment self-efficacy – t2	5.12 (.68)	4.58 (.85)	2.34*
Conducting traditional exams	4.20 (.64)	4.08 (.80)	.63
Administering standard exams	4.76(1.01)	4.58 (1.17)	.62
Reporting assessment results	5.27 (.65)	4.72 (.79)	2.50*
Use of performance assessments	5.04 (.90)	4.26 (1.19)	2.40*
Non-achievement based grading	4.94 (1.04)	5.00 (.92)	-.18
Validity and reliability tests	5.06 (.79)	4.63 (.98)	1.60*

* $p < .05$.

Pearson correlations were calculated between the quantitative background variables – the number of classrooms in which the teachers were teaching and the number of mentor observations – and each of the research variables. Table 4 presents these correlations.

The study found that only the number of classrooms was significantly associated with the research variables.

Significant associations were found with conceptions of formative assessment and assessment validity at both measurements, conceptions of summative assessment and assessment as harmful at the end of the year, self-efficacy at the midpoint of the year, reporting of assessment results, and the use of performance assessments.

Table 4. Correlations between the number of classrooms and mentor observations, and the research variables.

	Number of classrooms taught	Number of mentor observations
Formative assessment – t1	20.*	-.05
Formative assessment – t2	-.29**	-.02
Summative assessment – t1	.06	-.08
Summative assessment – t2	.31**	-.06
Assessment as harmful – t1	.09	.01
Assessment as harmful – t2	-.25*	.03
Assessment validity – t1	21.*	-.12
Assessment validity – t2	-.27**	-.05
Assessment self-efficacy – t1	.20*	.09
Assessment self-efficacy – t2	-.17	.16
Conducting traditional exams	-.01	.18
Administering standard exams	-.14	.15
Reporting assessment results	-.30**	.15
Use of performance assessments	-.21*	.18
Non-achievement based grading	-.09	.01
Validity and reliability tests	-.18	.07

* $p < .05$; ** $p < .01$.

Changes in teacher conceptions of student assessment and assessment self-efficacy over the course of the year

The third and fourth research hypotheses related to changes in teacher conceptions regarding student assessment and assessment self-efficacy across two-time

points. To examine these changes, an ANOVA for repeated measures was conducted. Table 5 presents this comparison.

The results indicate that, aside from the conception of assessment as harmful, there was a significant increase in all levels of assessment conceptions and assessment self-efficacy between the two measurements.

Table 5. Comparison between conceptions regarding student assessment and assessment self-efficacy in the middle and end of the year (N = 98).

	Mid-year (M/SD)	End of year (M/SD)	F
Formative assessment	4.39 (1.00)	4.83 (1.11)	8.17***
Summative assessment	3.77 (.83)	4.20 (.96)	11.37***
Assessment as harmful	4.62 (.95)	4.85 (1.10)	2.10
Assessment validity	3.95 (.86)	4.46 (1.20)	9.68***
Assessment self-efficacy	4.60 (.82)	5.04 (.73)	15.38***

*** $p < .001$.

Associations between teacher conceptions of student assessment and implementation of assessment methods

The first research hypothesis proposed that associations

would be found between teacher conceptions regarding student assessment and assessment self-efficacy and the implementation of assessments in class. To examine this hypothesis, Pearson tests were conducted to explore the relationships between each of the four conceptions and

assessment self-efficacy, and the use of various assessment implementation techniques. This hypothesis was investigated twice: once with conceptions of assessment and self-efficacy as measured in the middle of the year, and again at the end of the year (Table 6).

As evident from Table 6, the data collected in the middle

of the year showed negative associations between the conception of assessment as harmful and the use of performance assessments, as well as between the conception of assessment as harmful and validity and reliability tests in assessment.

Table 6. Pearson correlations between conceptions regarding student assessment as measured in the middle of the year and manners of implementing assessments.

	Formative assessment	Summative assessment	Assessment as harmful	Assessment validity	Assessment self-efficacy
Conducting traditional exams	.01	-.01	-.18	.04	.02
Administering standard exams	-.02	.09	-.02	.08	-.05
Reporting assessment results	-.09	.04	-.14	-.04	-.06
Use of performance assessments	-.05	-.06	-.20*	-.13*	-.11
Non-achievement based grading	.03	-.01	.06	.11	-.03
Validity and reliability tests	-.13	-.12	-.21*	-.12	-.02

* $p < .05$; ** $p < .01$.

Table 7 shows the Pearson correlations between the conceptions of assessment and self-efficacy and each of the applications of assessment in the classroom, as measured at the end of the year.

As evident from the data collected at the end of the year,

administering standard exams, reporting assessment results, using performance assessments, and conducting validity and reliability tests were statistically significantly associated with each of the assessment conceptions and with assessment self-efficacy.

Table 7. Pearson correlations between conceptions of student assessment and assessment self-efficacy as measured at the end of the year, and manners of implementing assessments.

	Formative assessment	Summative assessment	Assessment as harmful	Assessment validity	Assessment self-efficacy
Conducting traditional exams	.06	.11	-.02	.07	.11
Administering standard exams	.40*	.34**	.52***	.46***	.35***
Reporting assessment results	.48***	.38***	.56***	.53***	.53***
Use of performance assessments	.38***	.38***	.35***	.43***	.45***
Non-achievement based grading	.14	.17	.17	.19	.01
Validity and reliability tests	.36***	.32***	.50***	.39***	.69***

* $p < .05$; *** $p < .001$.

Predicting the use of assessment methods from assessment conceptions and assessment self-efficacy

The second research aim was to examine the extent to which certain conceptions of assessment and assessment self-efficacy predict how inductees apply various assessment techniques in practice. There was no specific hypothesis for this aim, beyond the hypotheses regarding associations between the conception and self-efficacy

variables and the assessment implementation variables. To assess predictive ability, we performed linear regression with each assessment method implemented, using the four assessment conceptions and assessment self-efficacy as independent variables.

To determine whether this analysis could be conducted with the current data, Pearson tests were first conducted between the independent variables – among the conceptions of assessment and self-efficacy. Strong associations were found between the measures. These

findings contradict the assumption of the variables' independence and significantly increased concerns of multicollinearity. For this reason, no regression analyses were performed.

In summary, assessment self-efficacy and beliefs in all the assessment conceptions increased from the middle of the induction year to its end. Assessment self-efficacy and all assessment conceptions were found to be associated with the actual implementation of summative assessment (administering standard exams and reporting assessment results), formative assessment (use of performance assessments), and validity and reliability tests of assessment.

DISCUSSION

This study examined changes in assessment conceptions and assessment self-efficacy of inductees following their formative assessment, using quantitative data collected at two points: the middle and end of the year. The first research hypothesis suggested that belief in assessment as a formative process would strengthen and belief in assessment as harmful/irrelevant would weaken by the end of the year, while conceptions related to summative assessment would remain unchanged. This hypothesis was partially confirmed, as all assessment conceptions, except for the harmful/irrelevant conception, showed an increase between the two measurement points.

These findings support the constructivist theoretical framework. The increase in belief in formative assessment aligns with Vygotsky's and Bruner's emphasis on social interaction and discovery learning, indicating that formative assessment can enhance teachers' ability to guide students in constructing knowledge. This shift suggests that the constructivist approach is effective not only in teaching but also in assessment practices, reinforcing the importance of formative assessment in promoting student motivation and self-efficacy.

Previous studies have not examined the impact of assessment processes during teacher induction on teachers' assessment conceptions and changes over the induction year. Hence, the hypothesis was based on the assumption that after experiencing significant mid-year formative assessments, inductees would recognize its value, enhancing their belief in its necessity. This aligns with studies on teacher conceptions, indicating that the induction environment, teaching experiences, professional development, and reflection processes shape their conceptions of teaching, learning, and assessment. The findings showed an increase in belief in formative assessment by the end of the year.

Regarding summative assessment, the findings were more complex. The second measurement occurred after inductees had received summative assessments at the end of the year, which determined their transition to regular teaching. This likely heightened their awareness of

summative assessment's role in decision-making and comparative evaluations at year-end. These results add to the literature on assessment processes during induction, demonstrating that a combination of formative and summative assessments during the induction year influences teachers' conceptions of their combined use.

The second research hypothesis predicted a rise in inductees' assessment self-efficacy during the second half of the year, which was confirmed. These findings are consistent with studies showing the contribution of formative assessment to teaching skills and self-efficacy in teaching capabilities (Guterman, 2014). This supports Bandura's Social Cognitive Theory, which posits that self-efficacy is developed through observing and learning from the experiences of others. The increase in assessment self-efficacy indicates that formative assessment processes, supported by mentor feedback, significantly contribute to inductees' confidence in conducting assessments.

Factors affecting assessment methods implemented by inductees

To identify factors influencing inductees' choice of assessment methods, the study examined associations between beliefs in four assessment conceptions and their self-efficacy at mid-year and year-end, and the student assessment methods they implemented. The first hypothesis within this aim posited a positive association between the use of alternative evaluation techniques (e.g., performance assessment and non-achievement-based grading) and the conception of assessment as formative, and a negative association with summative or harmful/irrelevant assessment conceptions. This hypothesis was partially confirmed; a positive association was found between the end-of-year conception of assessment as formative and the use of performance assessments. However, this method was only positively associated with the conception of assessment as formative. These associations were not evident at mid-year, likely because beliefs in these conceptions intensified over time.

Overload was a significant factor hindering the implementation of formative assessments. Inductees teaching more classes used performance assessments less frequently. This finding aligns with existing research on the challenges faced by early-stage teachers, who often have the same responsibilities as veteran teachers (Ministry of Education, 2020; Ovadia, 2018). Overload creates pressure and limits time for alternative assessments (Maskit and Dorfberger, 2018). This suggests that to enhance formative assessment practices, induction programs should address workload management and provide additional support for new teachers.

A similar pattern was found for summative assessment

methods, including traditional exams, reporting grades, and conducting validity and reliability tests. The hypothesis was that these methods would be positively associated with summative assessment conceptions and negatively with formative assessment conceptions. These associations were not evident at year-end. A possible explanation is that inductees felt obligated to follow Ministry of Education requirements, limiting their flexibility to apply personal assessment beliefs (Davis and Neitzel, 2011).

Findings regarding associations between inductees' formative or summative assessment conceptions and their implementation of assessment methods are consistent with prior research (Birenbaum, 2018). However, the strong positive association between formative and summative assessment conceptions at year-end presents a novel insight. This indicates that inductees might not clearly differentiate between these assessment types and may not fully grasp their distinct purposes and potential contradictions. This interpretation is further supported by data showing that inexperienced teachers reported higher levels of assessment results and validity and reliability tests compared to their experienced counterparts, suggesting that experienced inductees feel less compelled to adhere strictly to procedures and exhibit more flexibility in aligning assessment methods with their conceptions. Additionally, inductees who had taken a course on assessment reported greater use of standard exams, reflecting a more comprehensive understanding of diverse assessment methods.

Inductees may also demonstrate flexibility in their assessment conceptions by integrating both formative and summative assessments. This is corroborated by studies indicating that teachers possess multi-dimensional conceptions of assessment (Opre, 2015). Mitigating the adverse effects of summative assessment through sensitive feedback, positive reinforcement, and the use of varied assessment tools can enhance student motivation and learning outcomes. Inductees observed that employing diverse methods helps students perform better and reduces competitiveness and pressure in the classroom. This unique finding suggests that the assessment process during the induction year helps inductees develop a nuanced understanding of assessment, combining summative and formative methods to better support student development.

Two assessment methods, non-achievement-based grading and traditional exams, were not associated with any assessment conception at either measurement point. This might be due to the end-of-year timing when inductees were tested for their teaching license, leading to an increased inclination to report traditional exams and a decreased inclination to report non-achievement-based grading. Further research after an additional year of full-time teaching might reveal different findings.

The second hypothesis predicted a positive association between assessment self-efficacy and each assessment

method implemented at year-end. The findings supported this hypothesis for all four assessment methods and showed associations between assessment conceptions and methods: formative assessment (use of performance assessments) and three summative methods (standard exams, reporting results, and validity and reliability tests). These findings are consistent with studies indicating a high positive association between assessment implementation and teacher self-efficacy. Teachers with higher self-efficacy are more inclined to implement formative assessments (Yan and Brown, 2021), while lower self-efficacy leads to limited formative assessment use.

Practical recommendations for teacher induction programs

Based on these findings, specific recommendations for teacher induction programs can be made. First, incorporating more comprehensive training on formative assessment practices is essential. This training should include hands-on workshops, real-world examples, and opportunities for inductees to practice and receive feedback on formative assessment techniques. Additionally, providing ongoing mentorship throughout the induction year can reinforce these practices and help inductees navigate challenges. Mentor teachers should receive training on diverse assessment methods and be equipped to support inductees in balancing formative and summative assessments.

Implications for future research and teacher training programs

The implications of these findings extend to future research and the design of teacher training programs. Further studies should explore the long-term impact of formative assessment training and mentorship on teachers' professional growth and sustained assessment practices. Research could also examine how different components of induction programs, such as mentor feedback and workload management, influence the implementation of formative assessments. Additionally, developing assessment literacy programs that emphasize the integration of formative and summative assessments can provide a more holistic approach to teacher training.

CONCLUSIONS

This study aimed to examine how the assessment process experienced by inductees during their induction year affects their assessment self-efficacy and conceptions of student assessment, and how these are associated with their actual assessment practices. The findings suggest that mentorship programs should demonstrate how to

combine formative and summative assessments effectively, highlighting the benefits and shortcomings of each method. Mentor teachers should receive training on diverse assessment methods to provide relevant information and feedback to inductees.

Evaluating inductees should focus on enhancing teaching through formative processes rather than solely serving as a tool for summative assessment. Emphasis should be placed on observations accompanied by pedagogic discourse to enhance teaching.

A new approach to year-end assessment processes for inductees should be considered, viewing them not just as summative assessments for licensure decisions but as formative assessments with summative aspects. End-of-year feedback conversations involving principals, mentors, and inductees can help inductees utilize insights from the induction year to improve their work and progress.

Furthermore, the study highlights the importance of balancing the number of classes inductees teach to prevent overload, which hinders the use of formative assessments. Integrating formative and summative assessments can provide a more comprehensive evaluation that supports both student learning and teacher development.

Future research could explore the long-term impact of these assessment experiences on teachers' professional growth and their sustained assessment practices. Understanding how teachers evolve in their assessment practices beyond the induction year can provide deeper insights into the effectiveness of current training and support programs.

REFERENCES

- Amasha, M. (2014). *Classroom evaluation: Application characteristics and influencing factors among teachers teaching in Arab schools* [Unpublished doctoral dissertation, Tel Aviv University]. [Hebrew]
- Anthony, A. B., Gimbort, B. G., Luke, J. B., and Hurt, M. H. (2019). Distributed leadership in context: Teacher leaders' contributions to novice teacher induction. *Journal of School Leadership*, 29(1), 54-83. <https://doi.org/10.1177/1052684618825086>
- Bandura, A. (1997). *Self-efficacy – The exercise of control*. Freeman.
- Barnes, N., Fives, H., and Dacey, C. M. (2017). U.S. teachers' conceptions of the purposes of assessment. *Teaching and Teacher Education*, 65(4), 107–116. doi: 10.1016/j.tate.2017.02.017
- Birenbaum, M. (Ed.) (2018). *Research and application in assessment for learning (AfL) of students and teachers in Israel*. Tel Aviv: Mofet Institute. <https://kotar.cet.ac.il/KotarApp/Index/Book.aspx?nBookID=106208462> [Hebrew]
- Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged instrument. *Psychological Reports*, 99, 166–170. <https://doi.org/10.2466/pr0.99.1.166-170>
- Bruner, J. S. (1983). Education as Social Invention. *Journal of Social Issues*, 39(4), 129–141. <https://doi.org/10.1111/j.1540-4560.1983.tb00179.x>
- Davis, D. S., and Neitzel, C. (2011). A self-regulated learning perspective on middle grades classroom assessment. *Journal of Educational Research*, 104, 202–215. <https://doi.org/10.1080/00220671003690148>
- Friedman, I. (2020). *Sampling in education and social research. Principles, strategies, and performance processes*. The Henrietta Szold Institute. <https://www.szold.org.il/product>
- דגימה במחקר-חינוך וחברה-עקרונות אסטר [Hebrew]
- Goodman, J. S., and Blum, T. C. (1996). Assessing the non-random sampling effects of subject attrition in longitudinal research. *Journal of Management*, 22(4), 627–652. [https://doi.org/10.1016/S0149-2063\(96\)90027-6](https://doi.org/10.1016/S0149-2063(96)90027-6)
- Guterman, K. (2014). The path to educational leadership: The missing link in teacher evaluation. *Hachinuch Vesvivo*, 36, 141–160. [Hebrew]
- Harris, I., Irving, S. L., and Peterson, E. (2008). *Secondary teachers' conceptions of the purpose of assessment and feedback*. Paper presented to the Australian Association for Research in Education (AARE), Brisbane, Australia
- Hopfenbeck, T. N., Zhang, Z., Sun, S. Z., Robertson, P., and McGrane, J. A. (2023, November). Challenges and opportunities for classroom-based formative assessment and AI: a perspective article. In *Frontiers in Education* (Vol. 8, p. 1270700). Frontiers Media SA. <https://doi.org/10.3389/feduc.2023.1270700>
- Levy-Feldman, I. (2020). From educator to mentor: The good teacher for the twenty-first century. *Gilui Daat*, 16, 43–69. [Hebrew]
- Levy-Feldman, I., and Libman, Z. (2022). Between positivistic measurement and constructivist evaluation: The courage to bring about a change. *Gilui Daat*, 19, 13–14. [Hebrew]
- Levy-Vered, A. (2013). *Are future teachers really prepared for the challenge of learner assessment? The contribution of training, perceptions, and assessment self-efficacy to the assessment literacy of beginning teachers* [published dissertation, Tel Aviv University]. https://infocenters.co.il/mofet/notebook_ext.asp?book=12651&lang=heb&site=mofet [Hebrew]
- Libman, Z. (2013). Constructivism in education. In Z. Libman (Ed.), *Learning, understanding, knowing: Exploring pathways to constructivist teaching* (pp. 13–54). Hakibbutz Hameuchad and Mofet. [Hebrew]
- Maskit, D., and Dorfberger, S. (2018). Practicum teacher training in the students' perspective: Associate-Academic model and traditional model. Mofet Institute – Academia-Kita, 43– 45. [Hebrew]
- Ministry of Education (2020). *Letter to the principal about mid-year feedback for teaching interns 2021*. https://infocenters.co.il/mofet/notebook_ext.asp?book=11874&lang=eng&site=mofet [Hebrew]
- Nasser-Abu Alhija, F., Fresko, N., and Reichenberg, R. (2011). Teacher internship process– overview. In O. Schatz-Opppenheimer, D. Maskit, & S. Zilbershtrom (Eds.), *Being a teacher – in the teaching induction lane* (pp. 55-87). Mofet Institute and the Ministry of Education, Teacher Training and Professional Development Administration, Internship and Induction Department. [Hebrew]
- Opre, D. (2015). Teachers' conceptions of assessment. *Procedia. Social and Behavioral Sciences*, 209, 229–233. <https://doi.org/10.1016/j.sbspro.2015.11.222>
- Ovadia, M. (2018). Contribution of the internship workshop to the support and empowerment of intern teachers. *Mofet Institute Bulletin*, 62, 34–44. <https://ebook.macam.ac.il/read/6/bitaon62#39> [Hebrew]
- Schatz-Opppenheimer, O., Mandel, B., and Zilbershtrom, S. (2014). *Mentoring and accompanying interns and new education workers*. Ministry of Education. Education Workers Administration, Division A, Internship and Induction Department. <https://meyda.education.gov.il/files/staj/ChoveretChonchim.pdf> [Hebrew]
- Sperling, D. (2016). *Processes of licensing and evaluating new teachers. Information review*. Mofet Institute. http://library.macam.ac.il/study/pdf_files/d12274.pdf [Hebrew]
- Stanja, J., Gritz, W., Krugel, J., Hoppe, A., and Dannemann, S. (2023). Formative assessment strategies for students' conceptions—The potential of learning analytics. *British Journal of Educational Technology*, 54(1), 58-75. <https://doi.org/10.1111/bjet.13288>
- Sutcher, L., Darling-Hammond, L., and Carver-Thomas, D. (2019). Understanding teacher shortages: An analysis of teacher supply and demand in the United States. *Education Policy Analysis Archives*, 27(35), 1–40. <https://files.eric.ed.gov/fulltext/EJ1213618.pdf>
- Taylor, E. S., and Tyler, J. H. (2012). Can teacher evaluation improve teaching? *Education Next*, 12(4), 79–84.
- Vygotsky, L. (1987). *The collected works of L.S. Vygotsky.* Volume 1. *Thinking and Speaking*. New York, NY: Plenum Press.
- Whitaker, T., Whitaker Good, M., and Whitaker, K. (2019). How principals can support new teachers. *Educational Leadership*, 77(1), 50–54.

- <https://www.ascd.org/el/articles/how-principals-can-support-new-teachers>
- Yan, Z., and Brown, G. T. L. (2021). Assessment for learning in the Hong Kong assessment reform: A case of policy borrowing. *Studies in Educational Evaluation*, 68, 100985. <https://doi.org/10.1016/j.stueduc.2021.100985>
- Zhang, Z., and Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, 16(4), 323–342. <https://www.esf.edu/assessment/documents/assessmentpracticesandtskills.pdf>
- Zilbershtrom, S. (2013). The policy of teaching internship and induction department. In: S. Shimoni & A. Abinadab-Unger (Eds.), *The sequence. Training, specialization and professional development of teachers – policy, theory and practice* (pp. 95–100). Mofet Institute and the Ministry of Education. [Hebrew]

Citation: Tal-Amsili, K., Milevsky, A., Gerkerova, A., and Davidovitch, N. (2024). Assessment literacy: The process of evaluating early-career teachers and its association with assessment self-efficacy. *African Educational Research Journal*, 12(3): 210-223
