Language assessment literacy: Understanding the construct from Norwegian EFL teachers' perspective

Henrik Bøhn, University of South-Eastern Norway
Dina Tsagari, Oslo Metropolitan University

The question of what language assessment literacy (LAL) entails continues to preoccupy the research community. In the current study empirical evidence from Norway was gathered in order to inform our understanding of the construct. The focus of the investigation was twofold: First, it explored how Norwegian EFL teachers understand LAL. Secondly, it examined to what extent the same teachers found the dimensions in Taylor's (2013) LAL model relevant in the Norwegian context. To that end semi-structured interviews were carried out with 10 teachers at the lower and upper secondary levels in Norway. The results showed that the teachers attributed a lot of weight to *disciplinary competence* and *formative assessment*, and they agreed with Taylor on several of her hypothesized dimensions, such as the importance given to *language pedagogy*. However, they also disagreed with some of the other dimensions, such as test-specific skills (e.g., how to design and validate multiple-choice tests), which they believed to be less relevant in the Norwegian context. The findings point to the contextualized nature of teacher LAL and provide important knowledge on how teacher LAL can be conceptualized and how Taylor's model can be developed and refined.

Key words: assessment literacy, language assessment literacy, formative assessment, language teachers

Introduction

Language assessment literacy (LAL), understood as the knowledge and skills required by stakeholders involved in language assessment practices, has attracted a lot of attention in recent years (Tsagari, 2020). One reason for this is the growing importance attributed to formative assessment (Inbar-Lourie, 2017). LAL research has taken different directions, but a major concern which continues to preoccupy the research

community is the question of what the LAL construct comprises (Fulcher, 2020; Kremmel & Harding, 2019; Levi & Inbar-Lourie, 2020). This question involves issues such as the relationship between LAL and the more general concept of assessment literacy (AL), whether LAL is generic or contextual, and what kind of LAL is required for different kinds of stakeholder groups (teachers, administrators, students etc.) (Harding & Kremmel, 2016; Inbar-Lourie, 2017; Yan & Fan, 2021). In this study the focus was directed towards LAL for teachers.

Different conceptualizations of LAL exist (e.g., Brindley, 2001; Fulcher, 2012; Inbar-Lourie, 2008b; Pill & Harding, 2013). One conceptualization which is widely quoted in the research literature is Taylor's (2013) LAL model (Baker, 2016; Deygers & Malone, 2019; Giraldo, 2018; Hildén & Fröjdendahl, 2018; Kremmel & Harding, 2019).[1] This is a hypothesized model consisting of eight different components found to be relevant for four different stakeholder groups: classroom teachers, test writers, university administrators and professional language testers. What makes the model particularly interesting is that it scales the different competences required along different dimensions (cf. Figure 1, below). It thus suggests which knowledge and skills are important and less important. For teachers, it theorizes, for example, that *Language pedagogy* is highly important, whereas *Knowledge of theory* is less important. However, not all the dimensions in Taylor's model are equally clearly defined (Bøhn & Tsagari, 2021), and although the model has been empirically explored and tested (e.g., Baker & Riches, 2018; Bøhn & Tsagari, 2021; Kremmel & Harding, 2019; Yan et al., 2018), there are still questions regarding the importance of the different dimensions for teacher LAL, how some of the dimensions interrelate, and to what extent the model is relevant across contexts. Moreover, as calls have been made for the exploration of the construct from different perspectives and in different settings (Harding & Kremmel, 2016; Inbar-Lourie, 2017), there is need for more empirical evidence from various contexts.

Against this backdrop, the present study examined the concept of LAL from the point of view of English as a foreign language (EFL) teachers in Norway. Taking Taylor's (2013) model as a starting point, it used data from semi-structured interviews including closed-response and open-ended questions with 10 Norwegian secondary EFL teachers to explore the issues of how LAL can be understood and to what extent Taylor's model presents a relevant picture of LAL in their context. The purpose of the study was to inform our understanding of this construct and clarify what kind of knowledge and skills teachers need in order to carry out proper assessments in different settings.

---

[1] Taylor did not use the term *model* to describe her conceptualization of LAL. However, since "model" can be taken to mean "a stylized description of a target system" (Frigg & Hartmann, 2020), we shall use that term here.

The Norwegian context

In Norwegian primary and secondary education, English foreign language (EFL) teaching and learning is curriculum-based and compulsory for all students from first grade onwards (age six). The curricula are goal-oriented, specifying a number of competence aims which students are to work with, and against which they are to be assessed. Formative assessment holds a strong position in the educational system, having been given statutory status in the Regulations to the Education Act in 2009. These specify that:

> [t]he purposes of assessment […] are to promote learning […] and to provide information about [the students'] competence during the learning process, and at the end of instruction. […] Continuous assessment means that the students […] are to:
> a) participate in the assessment of their own work and reflect on their own learning and development;
> b) understand what they are to learn and what will be expected of them;
> c) be informed of what they have achieved;
> d) obtain guidance on how they can continue working to enhance their competence.

(Norwegian Ministry of Education and Research, 2006/2019, § 3-1 & 3-12, our translation)

With the introduction of the new national curriculum in 2020, stipulations regarding formative assessment were also laid down in the different subject curriculum texts. They reiterate many of the points made in the Regulations quoted above, emphasizing that students are to be involved in the assessment of their own learning, that teachers are to give feedback that can help the students progress, and that teachers need to "adapt their instruction to enable pupils to use the guidance provided to develop their [English] skills" (Norwegian Directorate for Education and Training, 2019a, p. 9, our translation). These are all provisions which are in line with recommendations made in the formative assessment literature (e.g., Wiliam, 2018).

In addition to these legislative specifications, the government also made Assessment for Learning (AfL) a nationwide priority area in 2010 (Norwegian Directorate for Education and Training, 2019b). Financial resources were therefore allocated to universities and teacher training colleges, as well as to local educational authorities, in order to organize courses, seminars, and online resources to help boost teachers' knowledge and skills in the area of formative assessment. The stated aim of the programme was "to make school owners (i.e., local governments) [and] schools […] further develop an assessment culture and assessment practices in which learning is the goal" (Norwegian Directorate for Education and Training, 2015, our translation). An evaluation of the programme carried out in 2018 found that teachers in Norway

overall had increased their knowledge of AfL. At the same time, the evaluation showed that they needed to continue to develop their skills in this area, including their knowledge of how to strengthen the relationship between formative and summative assessment practices (Norwegian Directorate for Education and Training, 2019b).

As for educational testing, Norway has traditionally not had a strong testing culture (Sjøberg, 2014). However, since the turn of the millennium, accountability testing, through international tests such as PISA, as well as national diagnostic testing, have become commonplace in the educational system. Although such testing is welcomed by some, it has also been met with considerable criticism. Among other things, critics maintain that this type of evaluation has negative washback effects, that quality is best evaluated by teachers' professional judgement, and that national testing is more about external control than internal development (Engh, 2011; Gunnulfsen, 2018). As for negative washback effects, for example, there are those who hold that the national tests make teachers spend too much time preparing their students for the tests, rather than focusing more strongly on the learning objectives of the curriculum (Sjøberg, August 14, 2019).

As for classroom-based assessment, no grades are awarded to students in primary school (age 6-12), which means that assessment is solely formative at this level. At the lower and upper secondary levels (age 13-19), however, assessment is required to be both summative and formative. Summative assessment is chiefly given in the form of "overall achievement marks", i.e., grades given by each subject teacher on the basis of multiple forms of classroom assessment. However, 30 per cent of the students in 10th grade (last year in lower secondary school) and 1st and 2nd year, upper secondary school, are randomly selected to sit for a written exam, and five per cent are randomly selected to take an oral exam.

The written exams are currently in the process of being overhauled. The educational authorities state that the new exams will be digital, that tasks will be more varied, and that the exams must more rigorously be built on the theoretical principles of test validity and reliability (Norwegian Directorate for Education and Training, 2021). Hence, with the strong focus on various types of assessment in the Norwegian educational context, in a system that is presently in transition, and where reports state that teachers need to continue to improve their AL, it is highly relevant to inquire into the teachers' perceptions on what assessment literacy in this system entails.

Conceptualizations of LAL

Models of LAL derive from the more general notion of assessment literacy (AL), a term introduced by Stiggins (1991). According to Stiggins, AL denotes "a basic understanding of the meaning of high- and low-quality assessment and [the ability] to

apply that knowledge to various measures of student achievement" (p. 545). As the definition implies, AL comprises both a knowledge component and a skills component and recent descriptions of AL and LAL have continued to include both (Abell & Siegel, 2011; DeLuca, 2012; Fulcher, 2012; JCSEE, 2015; Taylor, 2013; Xu & Brown, 2016). Some models also include a *principles* component, which refers to awareness of the use and impact of language tests, including aspects such as ethics, fairness and professionalism (Davies, 2008; Fulcher, 2012; Taylor, 2013).

Traditionally, according to Inbar-Lourie (Inbar-Lourie, 2008b), LAL had a strong psychometric orientation, focusing on the knowledge and skills related to traditional language testing. However, the growing focus on formative assessment in the late 1990s opened up the inclusion of more learning-focused assessment components in the descriptions of LAL (Brindley, 2001; Fulcher, 2012). An important development in our understanding of the construct was the shift from more componential notions of LAL (Brindley, 2001; Inbar-Lourie, 2008a) to the view that LAL is multidimensional and evolving (Fulcher, 2012; Pill & Harding, 2013), suggesting that different stakeholders would need various types of knowledge and skills at different levels. Taylor's (2013) model is a manifestation of this view. It includes eight dimensions which describe various types of skills, knowledge and principles, and five literacy levels, ranging from 0, no literacy required, to 4, very high literacy required. The LAL profile for teachers, illustrated in Figure 1, shows what kinds of levels of proficiency they would need with regard to the different dimensions.
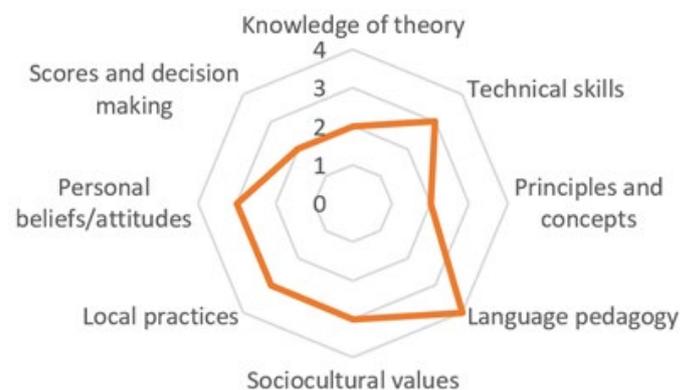
Figure 1. Teacher LAL according to Taylor (2013)

As can be seen in Figure 1, *Language pedagogy* is hypothesized to be the dimension that is most important for teachers, being located at level 4 on the profile, whereas *Knowledge of theory*, *Principles and concepts* and *Scores and decision making* are assumed to be the least important, all being located at level 2.

Regarding research using Taylor's model, few studies have actually operationalized and empirically tested the dimensions in the model. Two exceptions are Kremmel and Harding (2019) and Bøhn and Tsagari (2021). In Kremmel and Harding's study Taylor's

dimensions were operationalized in a survey questionnaire and distributed widely to teachers, researchers, policy-makers, test takers and other stakeholders globally. A total of 1086 respondents completed the questionnaire. Using exploratory factor analysis in the examination of the data, in order to determine the latent constructs underlying the participants' responses, Harding and Kremmel found that some of Taylor's dimensions needed to be reconceptualized and that certain levels of LAL proficiencies, as hypothesized by Taylor, were somewhat different from her suggestions. Among other things they found that the *Technical skills* dimension should be divided into two dimensions: "Developing and administering language assessments" and "Statistical and research methods". Similarly, the *Language pedagogy dimension* should be split into "Assessment in language pedagogy" and "Washback and preparation". In addition, they suggested that the components *Sociocultural values* and *Personal beliefs and attitudes* should be conflated and labelled *Assessment policy and local practices.* As for the importance of the different dimensions, Kremmel and Harding's findings and Taylor's hypothesized dimensions are not directly comparable, due to the modifications made to some of the dimensions. Nevertheless, some patterns emerged. For example, there is evidence that some of Taylor's hypotheses hold (e.g., that *Personal beliefs and attitudes* were important for teachers). However, some deviations from the original model also appeared. *Language pedagogy*, *Local practices* and *Sociocultural values*, for instance, were not found to be as significant as Taylor thought.

In Bøhn & Tsagari's (2021) study, Taylor's dimensions were operationalized in the form of questionnaire items and presented to five Norwegian EFL teacher educators in individual, semi-structured interviews. Although the operationalization of the dimensions turned out to be challenging (cf. below), the findings showed that the teacher educators agreed with many of Taylor's assumptions. One such example is *Language pedagogy*, which they believed to be a fundamental competency. However, the teacher educators agreed with the respondents in Kremmel and Harding's (2019) study that *Knowledge of theory* was more important than what Taylor predicted. They also reported that *Principles and concepts* was considerably more important and *Technical skills* were less important than what Taylor hypothesized.

In the same study, Taylor's dimensions were also operationalized by a close reading of the sources upon which the model is directly, or indirectly, built (Fulcher, 2012; Jeong, 2013; Malone, 2013; O'Loughlin, 2013; Pill & Harding, 2013; Scarino, 2013). This examination revealed some seemingly unclear interrelationships between some of the dimensions, particularly regarding the dimensions *Language pedagogy* and *Scores and decision making*, on the one hand, and *Principles and concepts* and *Knowledge of theory*, on the other. As for the former relationship, Taylor seems to include matters related to formative assessment in the component *Language pedagogy*. However, some types of "decision making" could also well be attributed to this dimension, as decisions may include judgements teachers make, on the basis of assessment results, of whether or not

to adjust their instruction to improve their students' learning (Bachman & Palmer, 2010; Black & Wiliam, 1998). When it comes to the relationship between *Principles and concepts* and *Knowledge of theory*, it appears that the former dimension is taken from Fulcher (2012), who employs this term in his own LAL framework. Fulcher mentions "test fairness" and "ethics of testing" (p. 115) as examples of "principles" and argues that "concepts" are related to "*theory* […] especially where statistics are introduced" (p. 124, italics added). Consequently, one may argue that "concepts" logically belongs to the *Knowledge of theory* dimension, rather than to the *Principles and concepts* dimension. For more details on how Taylor's dimensions were operationalized in this study, see Appendix B.

Thus, Taylor's model presents a very interesting theoretical conceptualization of what LAL can be, but empirical evidence of its relevance is limited. Kremmel and Harding (2019) recruited many respondents from the US, the UK and China, but few from the Scandinavian countries. Bøhn and Tsagari (2021) interviewed only five informants, none of whom were teachers. Given that LAL may be contextual, it would therefore be relevant to investigate how Norwegian teachers understand the LAL construct. Against this backdrop, the present study investigated how English secondary school teachers in Norway understand teacher LAL and to what extent they find Taylor's (2013) model relevant. The following two research questions (RQs) were therefore developed:

RQ1. What is Norwegian secondary school EFL teachers' general understanding of LAL in the Norwegian context?

RQ2. To what extent do Norwegian secondary school EFL teachers find the dimensions in Taylor's (2013) model relevant?

Method

Participants

To answer the RQs, 10 EFL teachers from 10 different lower and upper secondary schools in Norway were recruited as interview informants. Purposive sampling (Creswell, 2013) was used for selecting participants, in order to ensure variation in the sample with regard to geographical spread and type of school (lower and upper secondary, i.e. age 13-16 and 16-19, respectively). The informants were between 25 and 56 years old (mean=43), and they had from six months to 30 years of work experience (mean=16.9). They were all fully qualified English teachers with different types of educational specialization. All except one of the informants were university graduates, with two having completed a BA and seven having either obtained or been enrolled for Master's degrees in English, Education or Applied Linguistics. Their identities have been anonymized in this paper, and they will be referred to as "T1", "T2", etc. (which are acronyms for "Teacher 1", "Teacher 2", and so on).

Data collection

Using Taylor's (2013) model as starting point, we developed a pilot interview guide, which was distributed to two lower-secondary school teachers. Both open-ended questions and closed-response items were included (cf., interview guide, Appendix A). After some background questions on age, experience and so on, we started the pilot interview with open-ended questions concerning RQ1, i.e., how the informants understood the idea of teacher LAL in the Norwegian context. The first open question was formulated as follows: "What kinds of skills and knowledge does an English teacher in secondary school need in order to carry out good assessments?". Subsequently, we went on to present closed-response items regarding the relevance of Taylor's model, with accompanying six-point Likert scales going from "Strongly disagree" to "Strongly agree" (see Appendix A), in order to measure the teachers' views of the importance of the model's dimensions for the Norwegian context (cf. RQ2). A six-point Likert scale was chosen as it was assumed to reduce uncertain and neutral responses (Matell & Jacoby, 1971). The reason for starting out with open questions was that we wanted the teachers' unguided views on the construct. Such a procedure could potentially also mitigate the problem of conformity in responses to closed questions, as respondents are likely to report that everything is important (Fulcher, 2012).

The closed-response items were developed on the basis of our interpretation of the dimensions in Taylor's model. These items, then, served as our operationalization of the model. Their specifications can be found in Appendix B. When presenting the closed-response items to the informants (both the pilot informants and the teachers in the study), we welcomed them to elaborate on their answers in case they had comments on the different items, or they wanted to explain or moderate their reasons for responding the way they did. The reason for this was to obtain possible fuller accounts of how the teachers understood the construct and Taylor's way of representing it (Bryman, 2016).

The final interview guide was slightly modified to obtain some more data. More specifically, we added two items on the importance of summative and formative assessment (items 16a and 16b, cf. Appendix A) as they were thought to strengthen our findings regarding the *Language pedagogy* and the *Scores and decision making* dimensions. The participants were interviewed via telephone (*n*=5) and face-to-face (*n*=5) for periods of between 20 minutes and one hour. Nine interviews were carried out in Norwegian, and one was done in English, as this was the first language of the informant. Before the interviews the teachers were informed of the purpose of the study, and they signed a consent form to allow us to use their data anonymously.

Analysis

The interviews were transcribed in the original language by the first author, resulting in nine Norwegian and one English transcript. After transcription they were then sent back to the interviewees for respondent validation (Bryman, 2016). Two interviewees wrote back to us, commenting on some minor issues, such as the educational specialization they had taken and the kinds of students they were presently teaching. Apart from that, the informants did not report any inaccuracies in the accounts. The nine Norwegian transcripts were then translated into English, again by the first author, as we wanted to conduct the data analysis in English.

The answers to the open questions, as well as the follow-up responses, were analysed through four consecutive cycles, using the computer software QSR NVivo 12. In the first cycle, the transcripts were read through to get an overall take on the material. In the second cycle, In Vivo coding (Miles et al., 2014) was used to generate initial codes, or categories, and the material was annotated in cases where conspicuous statements or significant explanations were made. In the third cycle, the In Vivo codes were compared internally to ensure that they did not overlap, and to determine whether statements were appropriately classified in their respective category. In the fourth cycle, some of the categories were further refined in order to provide more accurate and coherent descriptions of the different sub-constructs and create more consistency in the analysis. For example, a category which was initially labelled "Knowledge of the curriculum" and subsumed under the broader construct of "Disciplinary competence" was later split up into the categories *Knowledge of communicative competence*, *Knowledge of texts* and *Knowledge of literature, society and culture*.

In addition to the qualitative analysis of the answers to the open-ended questions, the responses to the closed questions were analysed by feeding the data into SPSS Statistics 26 and calculating average measures, such as mean and median. Only descriptive statistics were used as the design of the study was mainly qualitative and involving a limited number of participants (n=10). Hence, the point was to obtain a general idea of how the informants deemed the relevance of Taylor's dimensions, rather to make more advanced calculations regarding for example correlation coefficients.

Results

Results for RQ1

The analysis of the responses to the initial, open-ended question concerning RQ1, i.e. what teacher LAL comprises, yielded five components:

- *Disciplinary competence*

- *Assessment-specific competence*
- *Pedagogical competence*
- *Collaboration competence*
- *Metacognitive skills*

The first component, *Disciplinary competence*, concerns subject-related matter such as knowledge of texts, communicative competence, and knowledge of literature, culture and society. In the Norwegian context, this would relate to issues such as text composition, genre and register, and literature, society and culture pertaining to the English-speaking world. Three of the teachers actually used this term to describe the dimension, and it relates to both the teachers' knowledge of these issues, and their skills, particularly regarding communicative competence, both their students' and their own. This is the reason why we decided to use the term *competence* to describe the component, rather than "knowledge", as competence can be defined as comprising both knowledge and skills (European Commision, 2013) and that the concept is a focal point in the Norwegian educational system (Norwegian Directorate for Education and Training, 2019b). The following statement, from informant T4, gives an example:

> Disciplinary knowledge [entails] culture and… language knowledge and… er… Because giving feedback on for example grammar requires an understanding of how language is grammatically put together… And the same thing goes for text structure. The teacher needs to have this subject knowledge.

The second dimension, *Assessment-specific competence*, includes a range of different skills and know-how concerning assessment and how to carry out sound assessment practices. Examples of these are knowledge of different purposes and types of language assessment, what and how to assess, knowledge of assessment theory, such as validity and reliability, and an understanding of institutional requirements and limitations. The following quotes from two informants can serve as examples:

> [Y]ou need to have a basis upon which you are able to award that mark, and that needs to be linked to those reflections you have made, and the stipulations made [government documents, assessment guidelines], and the research and the theory which is the basis of what we're doing, in a similar way to all the other things we're doing. [Informant T6]

> [One needs to have] awareness […] of what one is assessing... It is important that you make sure to assess the different competences… in the students… not just written, or not just oral, but also other competences… And this is linked to the purpose of assessment… And you need to use different forms of assessment to do that. [Informant T7]

The third dimension, *Pedagogical competence*, was developed from statements such as "you have to be a good teacher in order to carry out relevant assessments" [informant

T4]. This includes the ability to create good relations with the students, to convey curricular content so that learners understand, and to communicate assessment results in a way that is comprehensible to them. The ability to create good relations requires empathy and social skills. Finally, this dimension implies knowledge of formative assessment and the corresponding ability to clarify goals and criteria, involve the students in their own learning processes, and give feedback which may improve learning. Two additional quotes further illustrate the informants' reflections on this dimension:

> [I]t is important to be able to adapt topics… so that they understand. […] But the teacher also needs to be able to communicate with the students… both before assessment… to let them know what you're after… what the students are supposed to show, and also be able to give relevant feedback which the students can make use of. [Informant T4]

> And then you need to have knowledge of how to help the student improve, right? What should I focus on now in my conversation with the student [about his/her performance]? What is required in order for this student to move on… to develop his or her knowledge and skills in the subject? [Informant T6]

The fourth dimension, *Collaboration competence*, relates to the teachers' ability to collaborate with colleagues in the development of an assessment culture. This includes for example discussing criteria, grading students' papers together, and sharing ideas on fruitful ways of assessing student competence. This component was explained in the following way:

> [C]ollaboration with other English teachers involving the sharing of experiences [is important]. We have, for a number of years, been grading students' written work in pairs, right… involving reading each others' [i.e. each others' students'] texts and discuss how to grade them. [Informant T3]

The fifth dimension, *Metacognitive skills*, relates to the teachers' ability to reflect on the usefulness and relevance of their own assessment methods, avoiding mechanical practices which do not contribute to valid, reliable and fair assessments. This was touched upon by three teachers.

Informant T4 addressed the issue in the following way:

> [It is important] that you are conscious of what you want to do in a particular assessment situation... relating to... for example, the ability to be able to distinguish between different forms of assessment on the basis of what kinds of results that you want. [Informant T4]

Results for RQ2

RQ2, regarding the teachers' views on the relevance of Taylor's (2013) LAL model in the Norwegian context, was answered with data from the closed-response items. The responses were analysed by calculating mean, median, and standard deviation values. As for which items measured the different components of the model, see Appendix B. Overall, the analysis showed that the teachers scored the different items above average (Mean = 4.62), but with some clear differences concerning which dimensions they found to be particularly important. The items and their corresponding dimensions that received the highest scores are listed in Table 1.

Table 1. The items receiving the highest scores by the informants on a scale from 1 ("Strongly disagree") to 6 ("Strongly agree")

Dimension	Item no.	Item	Mean	Median	St. dev.
Language Pedagogy	16a	*It is important to have knowledge of formative assessment*	5.80	6.00	.32
Language pedagogy	12	*It is important to have knowledge of how assessment can promote learning*	5.80	6.00	.42
Local practices	24	*It is important to have knowledge of the national curriculum*	5.70	6.00	.67
Language pedagogy	21	*It is important to have knowledge of how assessment can be used to motivate students.*	5.60	6.00	.52
Language pedagogy	30	*It is important to have knowledge of how to communicate assessment results in appropriate ways*	5.50	6.00	.71
Personal beliefs/ attitudes	32	*It is important to have knowledge of how one's own view on assessment (values, outlook) may affect assessment.*	5.50	6.00	.71

As can be seen from Table 1, the highest scored statements were items 12 and 16a. Both concern formative assessment and indicate that the teachers found the component *Language pedagogy* to be essential for teacher LAL. This finding was corroborated by the fact that item 21, regarding motivation, and item 30, how to communicate assessment results, also received very high scores. The third highest rated item was no. 24, knowledge of the national curriculum. This suggests that the dimension *Local practices* was also highly valued. However, as item 26, which also tapped into this dimension, was not deemed equally important (see Table 2, below), *Local practices* ended up with a lower overall rating (cf., Table 3, below). Finally, it can be seen in Table 1 that item 32, knowledge of how one's own values, norms and perceptions may affect assessment was also awarded a high score. Since this item was the only one tapping into the *Personal beliefs/attitudes dimension*, it shows that the teachers found it to be a central component of teacher LAL.

When it comes to the lowest-scored items, Table 2 gives an overview of the results.

Table 2. The items given the lowest scores by the informants on a scale from 1 ("Strongly disagree") to 6 ("strongly agree")

Dimension	Item no.	Item	Mean	Median	St. dev.
Technical skills	15d	*It is important to have knowledge of large-scale testing*	3.20	3.00	1.42
Technical skills	15a	*It is important to have knowledge of how testing can be used appropriately in the English language classroom*	3.30	3.00	1.49
Technical skills	15c	*It is important to have knowledge of statistical measures in order to interpret results from multiple-choice tests*	3.40	3.50	1.05
Technical skills	15b	*It is important to have knowledge of how good items are created*	3.70	3.50	1.06
Local practices	26	*It is important to have knowledge of local issues which may be relevant for assessment* (e.g., local educational authorities' rating guidelines)	3.80	4.00	1.23
Sociocultural values	31	*It is important to have knowledge of how values in society may affect assessment*	3.80	4.00	1.23
Knowledge of theory	27	*It is important to have knowledge of assessment/testing theory (issues such as «validity», «reliability» etc.).*	3.90	4.00	1.37

As Table 2 shows, items 15a-d were the statements that scored the lowest. All of them tap into the *Technical skills* dimension. In addition, as mentioned above, item 26 (local issues) was scored fairly low, comparably speaking. Consequently, the component *Local practices* did not obtain a very high overall score, despite the fact that item 24, knowledge of the curriculum, was found to be very important (see Table 1 and Table 2). Furthermore, item 31, knowledge of societal values, was rated lower than most other statements. Consequently, *Sociocultural values* received a lower overall rating. Finally, item 27, knowledge of assessment theory, was also deemed somewhat less relevant. However, since the other item measuring the same dimension, i.e. no. 22b, knowledge of theories of communicative competence, was found to be slightly more important, the overall score for *Knowledge of theory* turned out to be higher.

By aggregating the responses to all the items regarding each dimension, we could calculate average scores for all the eight dimensions. However, since our closed-response items were scored on a six-point scale, we found it relevant to translate these scores to the five-point scale used in Taylor's model. In order to do so we used the median scores, as they may be more relevant with ordinal data (Jamieson, 2004), and applied the formula 0,8x – 0,8 (IBM Support, n.d.). The overall results of the answers to the closed-response items, represented both in the values reported by the informants (middle column, below)

and in the values converted to Taylor's model (right column below), are reported in Table 3.

Table 3. Conversion of median scores from the six-point scale of the present study to the five-point scale of Taylor's model.

Dimensions in Taylor's model	Original median values	Conversion to Taylor's model
Knowledge of theory	4.25	2.60
Technical skills	3.25	1.80
Principles and concepts	4.00	2.40
Language pedagogy	5.50	3.60
Sociocultural values	4.00	2.40
Local practices	5.00	3.20
Personal beliefs/attitudes	6.00	4.00
Scores and decision making	4.75	3.00

Follow-up answers to the closed-response items

In order to explain the findings related to RQ2, above, we analysed the follow-up comments that the teachers made on the responses to the closed questions. Two findings are notable, namely the results pertaining to *Language pedagogy*, which received the second highest overall scores, and *Technical skills*, which received the lowest.

Regarding *Language pedagogy*, several interesting comments were made on the highest scored items, 12 and 16a, which both relate to formative assessment (cf., Table 1). Seven informants pointed out that this was the most important form of assessment, since the ultimate goal of education is student learning. The following excerpt illustrates this sentiment:

> Formative means… to help students progress. Yes. So it's supposed to point to the future. And summative [assessment] is actually just a matter of awarding a mark. And to justify it, of course. But in that case, the justification is perhaps less important than actually awarding the mark. So formative assessment is absolutely the most important [form of assessment]. [Informant T7]

As for the relatively lower ratings given to *Technical skills*, the following exchange between the first author and informant T1 provides an interesting explanation. This exchange started with the informant responding to item 15a, "knowledge of language

testing", by giving it a 2. Explaining this, she said that: "I may know too little about this, but I don't think that's important at all. I don't understand how that can promote learning, but I'm sure that it can…". Following this, she went on to score item 15c, "knowledge of statistics", a 5. The following exchange between the interviewer and the informant ensued:

> I: But would you like to change what you answered previously? You said that knowledge about language testing is not important in a Norwegian context. You said 2.
>
> T1: (Pause) (Sighs). Well… it is a world that I am not that familiar with… So… in order to test… that's something that other people do, in my opinion. I don't.
>
> I: I understand. But still, you say that it is important to have knowledge of… sort of… to know something about standard deviation and…
>
> T1: Yes, because you mentioned national tests, and that's of course… Because I thought that this has got nothing to do with me. I thought about… multiple choice and items and so on. But when you mentioned that [i.e., understanding statistics to interpret results from national tests], then it gets much more interesting.

Two issues are worth commenting on in these extracts. First, the remark "I don't understand how [language testing] can contribute to learning, but I am sure that it can" indicates that the informant has a strong focus on learning-oriented assessment and seems to ignore, or at least to tone down, testing-oriented practices. Second, the same statement indicates uncertainty regarding the efficacy of language testing in classroom contexts. Support for this assumption is found in the remark "[language testing] is a world that I am not familiar with". This statement suggests that she has little knowledge of the field. Also, the fact that she first declares knowledge of language testing to be less important, and then goes on to judge knowledge of statistics as very important supports the same conclusion. Statements from three other teachers were of the same kind. For example, when informant T8 was asked about the relevance of having knowledge about language testing, standard deviation and measurement errors, he responded: "Well, that was Greek to me, so I think I'd have to say [the score] 1".

However, other informants pointed to the nature of ELT in Norway as a possible reason why language testing was found to be of less value to the teachers. This relates to the tradition of (oral and written) text production as an important part of the English subject. For instance, when asked why she thought language testing was less widespread in Norway than in the U.S., informant T3 declared:

> We are more focused on the writing of texts, or to have an oral presentation. So there must be coherence [between the learning goals, activities and assessment], I believe.

This view ties in with a position taken by informant T7, a non-Norwegian teacher who had a master's degree in applied linguistics, and who was familiar with language testing from his home country. He scored classroom language testing a 2. When asked whether he thought that multiple choice could be relevant for diagnosing students, he said: "No, I don't think so. You can do that in different ways as well". On the question of how he diagnosed new students at the start of the school year, he said:

> I feel that's very easy. Because I start by letting them write a bit, in one of the first sessions… write a short text, and then you'll be able to spot, very quickly, what the situation looks like.

In other words, this informant seemed to de-emphasize multiple-choice language testing, both because there is a stronger tradition for text production in Norway, but also because he thought that there are other means of gathering information about student achievement than multiple-choice tests.

Discussion

Findings pertaining to RQ1

The first RQ of this study, i.e., Norwegian English teachers' general understanding of the LAL construct, showed that the teachers were concerned about five dimensions: *Disciplinary competence*, *Assessment-specific competence*, *Pedagogical competence*, *Collaboration competence*, and *Metacognitive skills*. Compared to Taylor's (2013) model, which consists of eight dimensions, the present conceptualization is broader. Aspects such as language testing, ethics, values and assessment theory were not touched upon by the informants and have therefore not been included here, although the teachers confirmed in their responses to RQ2 that they found several of these aspects important. One example of this is *Sociocultural values*; another is *Personal beliefs and attitudes* (cf. Table 3, above). This may indicate that professional stakeholders are not always conscious of, nor able to explicitly explain, the different types of knowledge and skills they need in order to perform tasks.

When it comes to aspects of LAL not specifically mentioned in Taylor's model, it is worth noting the many references the teachers made to *Disciplinary competence*, such as language knowledge and other curriculum-related features. This is a dimension that has also been found to be important in other descriptions of teachers' assessment

literacy (Bøhn & Tsagari, 2021; Scarino, 2013; Xu & Brown, 2016). Hence, there are good reasons for including it in a teacher LAL model in the Norwegian context.

Two other dimensions not included by Taylor are *Collaboration competence* and *Metacognitive skills*. The former was referred to by only two informants in the present study. Still, it is a meaningful dimension in the Norwegian setting since it can be useful in the development of an assessment culture, as pointed out by the Norwegian educational authorities (cf., Literature review section, above) and it is referred to in other LAL studies (Bøhn & Tsagari, 2021). Concerning *Metacognitive skills*, this component was mentioned by three informants. Interestingly, such skills have been mentioned elsewhere as a necessary aspect of teachers' competence, both in terms of general pedagogy and in terms of assessment (Bøhn & Tsagari, 2021; Jiang et al., 2016; Wang, 2020). We therefore find it relevant for inclusion in a teacher LAL model.

As pointed out in Bøhn and Tsagari (2021), there are good reasons for arguing that some of the aspects in Taylor's model need to be clarified. In addition, some elements need to be rearranged, and some dimensions should be renamed (see also Kremmel & Harding, 2019). For example, we suggest that aspects related to formative assessment should be moved from *Score and decision making* to *Language pedagogy*, and the former dimension should be renamed "Scoring" and reserved for summative assessment. Moreover, we suggest that the *Principles and concepts* component be relabelled *Principles*, and aspects related to "concepts", which arguably concern theoretical matters, should be transferred to *Knowledge of theory* (cf., Literature review section, above). We would furthermore restrict the latter dimension to assessment-related theory, and rename it *Assessment theory*, since other dimensions, such as *Language pedagogy*, also entail theoretical knowledge. Furthermore, the inclusion of *Disciplinary competence* as a separate component would mean that aspects such as knowledge of the curriculum could be removed from the *Local practices* dimension and subsumed under *Disciplinary competence*. Hence, *Local practices* would merely concern knowledge of specific local issues, like for example local criteria. An overview of the revised model can be found in Figure 2, below.

Findings pertaining to RQ2

The results of the analyses regarding RQ2 showed that the teachers to some extent agreed with the hypothesized levels in Taylor's dimensions. However, there was also some disagreement. As can be seen in Table 3, the teachers were most in agreement with Taylor's scores given to *Local practices, Language pedagogy,* and *Principles and concepts*. The fact that the teachers found *Language pedagogy* to be highly important may come as no surprise, given the fact that this dimension was partly operationalized as knowledge of formative assessment, and that formative assessment is both a statutory requirement in the Norwegian system, as well as having been made a focus area by

the Norwegian educational authorities for many years. Interestingly, this finding was not supported by Kremmel & Harding's (2019) study of international teachers, but it is backed by Norwegian teacher educators (Bøhn & Tsagari, 2021). Hence, it may, as mentioned, be a context-specific feature.

The largest discrepancy between Taylor's hypothesized dimensions and the teachers' valuation of them was found in *Personal beliefs and attitudes*, *Scores and decision making*, and *Technical skills*. As for *Personal beliefs and attitudes*, which the teachers scored much higher than Taylor, it may be noted that the teachers' descriptions of what we categorized as *Metacognitive skills* (cf. Results for RQ1, above) were in many ways similar to our operationalization of *Personal beliefs and attitudes* (cf., Appendix B). Hence, there could be an overlap between these two dimensions.

Interestingly, in Kremmel and Harding's (2019) study, however, the teachers were in agreement with Taylor's hypothesized score, i.e. that *Personal beliefs and attitudes* was not so important. Regarding *Scores and decision making*, it is likely that the teachers in the present study scored this dimension higher than what Taylor suggested since the issue of summative assessment is also an issue that has been in focus in the Norwegian context, not least because of its inclusion in the Regulations to the Education Act. This dimension was found to be of similar importance by the teachers in Kremmel and Harding's (2019) study. Finally, as for *Technical skills*, we find here the clearest difference between Taylor's model and the Norwegian teachers' scores. The teachers' relatively low score on this dimension was similar to that of the Norwegian teacher educators in Bøhn & Tsagari's (2021) study. However, in Kremmel & Harding (2019) the teachers reported finding it more relevant to have knowledge of multiple choice, the use of statistics to evaluate scores, etc. One reason why Norwegian teachers and teacher educators appear to be more skeptical of such knowledge may be that there has traditionally been a weak testing culture in Norway, which means that there is no tradition for using multiple-choice in the classroom (as discussed in the Norwegian context section, above). Another, and related reason, may be that ELT in Norway has been fairly focused on the writing of texts, both in order to learn English and as a way of measuring competence in written exams, as pointed out by three of the informants in this study. As for the latter aspect, it is of course possible to assess content knowledge and various forms of communicative language competences (Council of Europe, 2001) on the basis of students' written texts.

Concluding remarks on findings

Combining the findings from the analysis relating to RQ1 and RQ2, we find that there are good reasons for arguing that Taylor's model should be modified, as well as be expanded for the Norwegian context. By modifying and expanding the model, as well as entering the values indicated by the teachers in their responses to RQ2, we would

like to suggest the following visualization (note that the values regarding *Disciplinary competence*, *Collaboration competence*, and *Metacognitive skills* are collated results, based on the teachers' qualitative responses to RQ1):
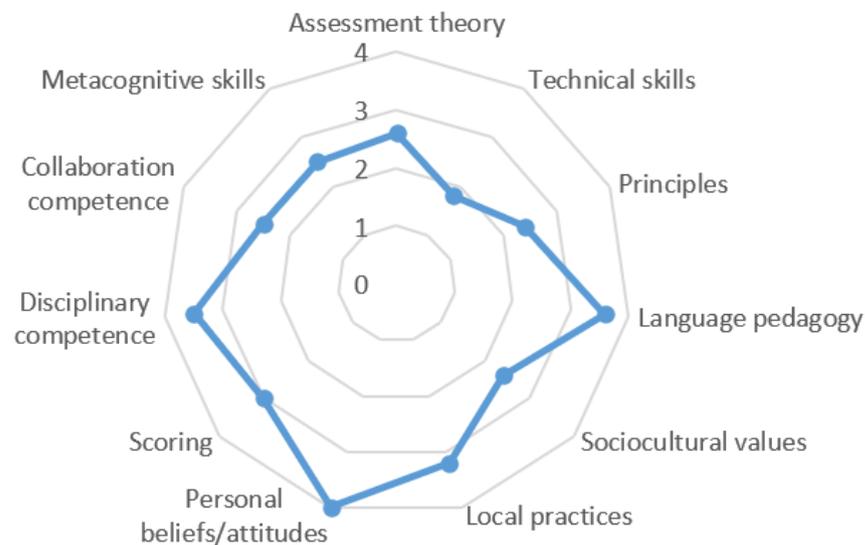
Figure 2. Teacher LAL according to the informants of this study

The focus placed on *Language pedagogy, Personal beliefs/attitudes and Disciplinary competence* by the Norwegian teachers in this study is to some extent supported by the Norwegian teacher educators in Bøhn & Tsagari (2021). Both participant groups agree that *Language Pedagogy* and *Disciplinary competence* are highly important for teacher LAL. However, the teachers put more emphasis on *Personal beliefs/attitudes* than what the teacher educators did. Another interesting similarity is that both teachers and teacher educators found *Technical skills*, such as use of multiple choice tests and knowledge of statistics, to be less important. A final aspect to be noted is that the informants in this study believed *Principles* to be considerably less important (score 2.40), whereas the teacher educators in Bøhn & Tsagari's (2021) study believed it to be of utmost importance (score 4.00). This is an indication that the teacher educators are more concerned with issues related to ethics and fairness.

Considering the fact that a new curriculum has been introduced in Norway, in which formative assessment has been even more strongly emphasized (cf., Norwegian context section, above), the findings in this study suggest that teachers are well aware of the importance attributed to learning-oriented assessment by the educational authorities, and the importance of *Language pedagogy* for LAL. It is also interesting to note the emphasis they place on *Personal beliefs/attitudes*, i.e., awareness of how their own personal outlook may affect assessment. This could indicate that they are concerned with assessments that are valid, dependable and fair, and in line with the stipulations made by the educational authorities. Interestingly, their devaluation of *Technical skills*, as the lowest-scoring dimension, may be somewhat at odds with

suggestions made by an expert committee's evaluation of summative assessments in Norway. The committee recommends that the final examination system, which is currently in the process of being revised, should be based on psychometrics and more strongly focused on validity and reliability (Blömeke et al., 2020).

Some limitations of this study should be noted. First of all, this is a predominantly qualitative study with a limited number of respondents (*n*=10). Hence, generalizations are problematic. Secondly, the operationalization of Taylor's dimensions is, as mentioned, a challenging task and is based on our understanding of their meanings. Thirdly, and similarly, the results of the analysis of the open-ended question (RQ1) rests on our interpretation of the aggregated responses to our questions. Hence, each of our informants may not have understood the construct in exactly this way. Yet, we believe that the components presented above provide a valid representation of the informants' collective understanding of teacher LAL in Norway.

Conclusions and further research

The results presented here have various implications for our understanding of teacher LAL. Firstly, they support the view that teacher LAL is localized (Harding & Kremmel, 2016; Inbar-Lourie, 2017). Secondly, they suggest that there are important LAL dimensions that have not been sufficiently highlighted in current models of LAL (Fulcher, 2012; Kremmel & Harding, 2019; Taylor, 2013). More specifically, in the Norwegian context, there are indications that disciplinary knowledge, collaboration competence, and metacognitive skills are important aspects of a language teacher's ability to carry out good assessment practices.

In addition, the findings show that the informants agreed with several of the hypotheses put forward by Taylor, particularly the importance attributed to *Language pedagogy*. However, they also found that *Scores and decision making* and *Personal beliefs* were more important than what Taylor assumed, and that *Technical skills* were less critical than predicted. On a more general level, the data revealed that this may be attributed to characteristics in the Norwegian educational system, where formative assessment has been strongly promoted in recent years, whereas language testing has been de-emphasized. This is in itself an interesting finding, as calls have been made for bringing psychometrics into the Norwegian exam system (Blömeke et al., 2020). Thus seen, one may hypothesize that teachers' understanding of the construct may change as external factors impact the system in which they work.

As for further avenues for research, we suggest that Taylor's (2013) conceptualization of the construct be further explored, particularly with a view to how the different dimensions should be operationalized, and how they should be delineated. Both the current study, Bøhn and Tsagari (2021), and Kremmel and Harding (2019) may provide

relevant points of departure in such a venture. Moreover, given the mainly qualitative nature of this study, it would be pertinent to look further into the question of how important larger cohorts of teachers find disciplinary competence, collaboration competence and metacognitive skills to be, both in Norway and elsewhere. The teacher LAL construct appears to be multifaceted and complex, and further research is needed to grasp more fully what it entails in different educational settings.

References

Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do. In D. Corrigan, J. Dillon, & R. Gunstone (Eds.), *The professional knowledge base of science teaching*. Springer.

Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford University Press.

Baker, B. A. (2016). Language assessment literacy as professional competence: The case of Canadian admissions decision makers. *The Canadian Journal of Applied Linguistics, 19*(1), 63-83.

Baker, B. A., & Riches, C. (2018). The development of EFL examinations in Haiti: Collaboration and language assessment literacy development. *Language Testing, 35*(4), 557-581. https://doi.org/https://doi.org/10.1177/0265532217716732

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74. https://doi.org/https://doi.org/10.1080/0969595980050102

Blömeke, S., Blikstad-Balas, M., Eggen, P.-O., Eira, K.-I., Fjørtoft, H., Gamlem, S. T. M., Gilje, Ø., Helgesen, R., Johannessen, S., Minken, M., Skillinghaug, S., Tveit, S., Waage, A., & Walker, M. J. (2020). *Vurderinger og anbefalinger om fremtidens eksamen [Evaluations and recommendations concerning future exams]*. https://www.udir.no/eksamen-og-prover/eksamen/vurderinger-og-anbefalinger-fremtidens-eksamen/

Brindley, G. (2001). Language assessment and professional development. In C. Elder, A. Brown, K. Hill, N. Iwashita, T. Lumley, T. McNamara, & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 126-136). Cambridge University Press.

Bryman, A. (2016). *Social research methods* (5th ed.). Oxford University Press.

Bøhn, H., & Tsagari, D. (2021). Teacher educators' conceptions of language assessment literacy in Norway. *Journal of Language Teaching and Research, 12*(2), 222-233. https://doi.org/https://doi.org/10.17507/jltr.1202.02

Council of Europe. (2001). *The Common European Framework of Reference for Languages: Learning, teaching, assessment*. Council of Europe. https://rm.coe.int/1680459f97

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

Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, *25*(3), 327-347. https://doi.org/10.1177/0265532208090156

DeLuca, C. (2012). Preparing teachers for the age of accountability: Toward a framework for assessment education. *Action in Teacher Education*, *34*(5-6), 576-591. https://doi.org/10.1080/01626620.2012.730347

Deygers, B., & Malone, M. E. (2019). Language assessment literacy in university admission policies, or the dialogue that isn't. *Language Testing*, *36*(3), 347-368. https://doi.org/10.1177/0265532219826390

Engh, R. (2011). *Vurdering for læring i skolen [Assessment for learning in school]*. Høyskoleforlaget.

European Commision. (2013). *Supporting teacher competence development for better learning outcomes*. Author.

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, *9*(2), 113-132. https://doi.org/10.1080/15434303.2011.642041

Fulcher, G. (2020). Operationalizing language assessment literacy. In D. Tsagari (Ed.), *Language assessment literacy: From theory to practice* (pp. 8-28). Cambridge Scholars Publishing.

Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development*, *20*(1), 179-195. https://doi.org/https://doi.org/10.15446/profile.v20n1.62089

Gunnulfsen, A. E. (2018). *Micro policy making in schools: Use of national test results in a Norwegian context* [Doctoral dissertation, University of Oslo], DUO Vitenarkiv. https://www.duo.uio.no/bitstream/handle/10852/61908/1/PhD-Gunnulfsen-2018.pdf

Harding, L., & Kremmel, B. (2016). Teacher assessment literacy and professional development. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 413-427). De Gruyter Mouton.

Hildén, R., & Fröjdendahl, B. (2018). The dawn of assessment literacy – exploring the conceptions of Finnish student teachers in foreign languages. *Apples: Journal of Applied Language Studies*, *12*(1), 1-24. https://doi.org/10.17011/apples/urn.201802201542

IBM Support. (n.d.). *Transforming different Likert scales to a common scale*. https://www.ibm.com/support/pages/node/422073

Inbar-Lourie, O. (2008a). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, *25*(3), 385-402. https://doi.org/10.1177/0265532208090158

Inbar-Lourie, O. (2008b). Language assessment culture. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed., Vol. 7, pp. 285-289). Springer.

Inbar-Lourie, O. (2017). Language assessment literacy. In E. Shohamy, I. G. Or, & S. May (Eds.), *Language testing and assessment* (3rd ed.). Springer. https://doi.org/https://doi.org/10.1007/978-3-319-02261-1

Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical education*, *38*(12), 1212-1218. https://doi.org/https://doi.org/10.1111/j.1365-2929.2004.02012.x

JCSEE (Joint Committee on Standards for Educational Evaluation). (2015). *The classroom assessment standards for preK-12 Teachers. https://www.amazon.com/Classroom-Assessment-Standards-PreK-12-Teachers-ebook/dp/B00V6C9RVO.*

Jeong, H. (2013). Defining assessment literacy: Is it different for language testers and non-language testers? *Language Testing*, *30*(3), 345-362. https://doi.org/10.1177/0265532213480334

Jiang, Y., Ma, L., & Gao, L. (2016). Assessing teachers' metacognition in teaching: The Teacher Metacognition Inventory. *Teaching and Teacher Education*, *59*, 403-413. https://doi.org/https://doi.org/10.1016/j.tate.2016.07.014

Kremmel, B., & Harding, L. (2019). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*, 1-21. https://doi.org/10.1080/15434303.2019.1674855

Levi, T., & Inbar-Lourie, O. (2020). Assessment Literacy or Language Assessment Literacy: Learning from the teachers. *Language Assessment Quarterly*, *17*(2), 168-182. https://doi.org/10.1080/15434303.2019.1692347

Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, *30*(3), 329-344. https://doi.org/10.1177/0265532213480129

Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert-scale items? Effects of testing time and scale property. *Journal of Applied Psychology*, *56*(6), 506-509. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.417.8264&rep=rep1&type=pdf

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Sage.

Norwegian Directorate for Education and Training. (2015). *Vurdering for læring - om satsingen [Assessment for learning - a note on the priority area].* https://www.udir.no/laring-og-trivsel/vurdering/nasjonal-satsing/om-satsingen/

Norwegian Directorate for Education and Training. (2019a). *Kunnskapsgrunnlag for vurdering av eksamensordningen [Knowledge base for the evaluation of the examination system].* https://www.udir.no/tall-og-forskning/finn-forskning/rapporter/Kunnskapsgrunnlag-for-evaluering-av-eksamensordningen/

Norwegian Directorate for Education and Training. (2019b). *Erfaringer fra nasjonal satsing på vurdering for læring (2010-2018) [Experiences from the national priority programme Assessment for learning (2018-2019)].* https://www.udir.no/tall-og-forskning/finn-forskning/rapporter/erfaringer-fra-nasjonal-satsing-pa-vurdering-for-laring-2010-2018/

Norwegian Directorate for Education and Training. (2021). *Endringer i eksamen etter nye læreplaner [Changes in the exams following the introduction of new national curricula]. https://www.udir.no/eksamen-og-prover/eksamen/slik-endrer-vi-eksamen/.*

Norwegian Ministry of Education and Research. (2006/2019). *Forskrift til Opplæringslova [Regulations to the Education Act].* https://lovdata.no/dokument/SF/forskrift/2006-06-23-724

O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing, 30*(3), 363-380. https://doi.org/10.1177/0265532213480336

Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing, 30*(3), 381-402. https://doi.org/10.1177/0265532213480337

Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing, 30*(3), 309-327. https://doi.org/10.1177/0265532213480128

Sjøberg, S. (2014). PISA-syndromet: Hvordan norsk skolepolitikk blir styrt av OECD [The PISA-syndrome: How Norwegian educational policies are governed by the OECD]. *Nytt Norsk Tidsskrift, 31*(1), 30-43. https://www.uis.no/getfile.php/13217918/HF/PISA-Syndromet_Sj%C3%B8berg_Nytt_Norsk_Tidsskrift_1-2014.pdf

Sjøberg, S. (August 14, 2019). *Nasjonale prøver [National tests].* https://snl.no/nasjonale_pr%C3%B8ver

Stiggins, R. J. (1991). Assessment Literacy. *The Phi Delta Kappan, 72*(7), 534-539. http://www.jstor.org/stable/20404455

Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing, 30*(3), 403-412. https://doi.org/10.1177/0265532213480338

Tsagari, D. (2020). Language assessment literacy: Concepts, challenges and prospects. In S. Hidri (Ed.), *Perspectives on language assessment literacy: Challenges for*

improved student learning (pp. 13-33). Routledge. https://doi.org/https://doi.org/10.4324/9781003016083

Wang, J. (2020). Exploring the perceived integrations between assessment and metacognition: A qualitative inquiry of three award-winning teacher educators' conceptions of assessment in a Hong Kong university context. *Frontiers in Education*, *4*(157), 1-15. https://doi.org/10.3389/feduc.2019.00157

Wiliam, D. (2018). *Embedded formative assessment* (2nd ed.). Solution Tree Press.

Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149-162. https://doi.org/https://doi.org/10.1016/j.tate.2016.05.010

Yan, X., & Fan, J. (2021). "Am I qualified to be a language tester?": Understanding the development of language assessment literacy across three stakeholder groups. *Language Testing*, *38*(2), 219-246. https://doi.org/10.1177/0265532220929924

Yan, X., Zhang, C., & Fan, J. J. (2018). "Assessment knowledge is important, but …": How contextual and experiential factors mediate assessment practice and training needs of language teachers. *System*, *74*, 158-168. https://doi.org/https://doi.org/10.1016/j.system.2018.03.003

Appendix A – Interview guide

RQ: What does language assessment literacy mean according to English language teachers in Norway? What kind of skills and knowledge does a teacher need in order to be able to carry out good assessment in ELT?

Background questions

1. Age?
2. What is your specialization?
3. How long have you been employed in teacher education?
4. Have you taught any assessment courses or assessment topics in teacher education (English)? What did these cover?
5. Have you otherwise been involved in language assessment (e.g., marking/consulting for/ researching TOEFL, IELTS, etc)?
6. Have you published any articles, books, reports (or other types of publications) on the topic of assessment (peer or non-peer reviewed)? Which ones?
7. Do you yourself have any experience teaching and assessing students in English foreign language classes in secondary school (in Norway or elsewhere)?
8. On a scale from one to six, how knowledgeable do you consider yourself to be when it comes to the topic 'language assessment'?

Language Assessment Literacy

General questions
9. What kinds of skills and knowledge does an English teacher in secondary school need in order to carry out good assessments?

10. As for the things you mentioned in your response to question 9., would you say that there are some skills or knowledge aspects that are important than others, or are they all equally important?

11. On a scale from one to six, how important would you say that assessment is in English teaching and learning?

Specific questions

To what extent are the following aspects important for an English teacher's assessment competence? On a scale from one to six, where one is the lowest and six is the highest, please judge the following questions/statements:

12. It is important to have knowledge of how assessment can promote learning 1 2 3 4 5 6
13. It is important to have knowledge of self-assessment   1 2 3 4 5 6

14. It is important to have knowledge of peer-assessment   1 2 3 4 5 6

15. It is important to have knowledge of:

     a. how language testing can be used appropriately in the English

        language classroom, e.g. multiple choice   1 2 3 4 5 6

     b. how good items are created   1 2 3 4 5 6

     c. statistical measures in order to interpret results

        from multiple choice tests (e.g. mean, standard deviation,

        measurement error)   1 2 3 4 5 6

     d. large-scale testing   1 2 3 4 5 6

16. How do you understand the terms 'formative' and 'summative' assessment?
On a scale from one to six, score the following statements:

     a.   It is important to have knowledge of formative assessment   1 2 3 4 5 6

     b.   It is important to have knowledge of summative assessment   1 2 3 4 5 6

17. It is important to have knowledge of different types of assessment   1 2 3 4 5 6

18. It is important to have knowledge of how assessment can be used as a
     diagnostic tool.   1 2 3 4 5 6

19. It is important to have knowledge of different purposes of assessment
     (diagnostic, ranking, promoting learning)   1 2 3 4 5 6

20. It is important to have knowledge of large scale testing, such as national
     tests, PISA etc.   1 2 3 4 5 6

21. It is important to have knowledge of how assessment can be used to
     motivate students.   1 2 3 4 5 6

22. a. It is important to have knowledge of

     a.   languages and language learning   1 2 3 4 5 6

     b.   theories of communicative competence   1 2 3 4 5 6

23. It is important to have knowledge of language frameworks, such as CEFR   1 2 3 4 5 6

24. It is important to have knowledge of the national curriculum (the general
part and the English subject curriculum)   1 2 3 4 5 6

25. It is important to have Knowledge of the Regulations to the Education Act   1 2 3 4 5 6

26. It is important to have knowledge of local issues which may be relevant for
assessment (e.g. the local educational authorities' assessment criteria,
how teachers in your local district assess student performance etc.)   1 2 3 4 5 6

27. It is important to have knowledge of assessment/testing theory (related
 to «validity», «reliability» etc.)   1 2 3 4 5 6

28. It is important to have knowledge of ethical issues (fairness, use of
assessment results for purposes other than what was intended etc.)   1 2 3 4 5 6

29. It is important to have knowledge of how to use rating scales / scoring
rubrics   1 2 3 4 5 6

30. It is important to have knowledge of how to communicate assessment
results in appropriate ways (e.g. how to explain to parents the results
from national tests)   1 2 3 4 5 6

31. It is important to have knowledge of how values in society may affect
assessment   1 2 3 4 5 6

32. It is important to have knowledge of how one's own view on assessment
(values, outlook) may affect assessment   1 2 3 4 5 6

33. It is important to have knowledge of the history of language assessment   1 2 3 4 5 6

Additional questions:

34. Would you say that knowledge is more/less important than skills, or are they equally
important (What is knowledge? What are skills?)

35. Do you think teacher education prepares students well for carrying out good assessment
practices?

36. If you answered "no" in the previous question, what do you think should be changed in
teacher education regarding the topic of assessment?

37. Is knowledge and skills in the area of assessment more/less important
in English than in other subjects? If so, why?

38. Is there anything you'd like to add?

Appendix B – Operationalization of the Taylor model dimensions

Dimension	Item no.	Item
Knowledge of theory	22b	It is important to have knowledge of theories of communicative competence.
	27	It is important to have knowledge of assessment/testing theory (issues such as «validity», «reliability» etc.).
Technical skills	15a	It is important to have knowledge of how language testing can be used appropriately in the classroom
	15b	It is important to have knowledge of how good items are created
	15c	It is important to have knowledge of statistical measures in order to interpret results from multiple choice tests (e.g. mean, standard deviation, measurement error)
	15d	It is important to have knowledge of large scale testing
Principles and concepts	28	It is important to have knowledge of ethical issues (fairness, use of assessment results for purposes other than what was intended etc.)
Language pedagogy	12	It is important to have knowledge of how assessment can promote learning
	13	It is important to have knowledge of self-assessment
	14	It is important to have knowledge of peer-assessment
	16a	It is important to have knowledge of formative assessment
	18	It is important to have knowledge of how assessment can be used as a diagnostic tool.
	21	It is important to have knowledge of how assessment can be used to motivate students.
	22a	It is important to have knowledge of languages and language learning
	30	It is important to have knowledge of how to communicate assessment results in appropriate ways (e.g. how to explain the results from national tests).
Sociocultural values	31	It is important to have knowledge of how values in society may affect assessment

Local practices	24	It is important to have knowledge of the national curriculum (the general part and the English subject curriculum).
	25	It is important to have knowledge of the Regulations to the Education Act.
	26	It is important to have knowledge of local issues which may be relevant for assessment (e.g. the local educational authorities' assessment criteria, how teachers in your local district assess student performance etc.)
Personal beliefs/attitudes	32	It is important to have knowledge of how one's own view on assessment (values, outlook) may affect assessment.
Scores and decision making	16b	It is important to have knowledge of summative assessment
	29	It is important to have knowledge of how to use rating scales / scoring rubrics.
Items not included in the analysis	17	It is important to have knowledge of different types of assessment
	19	It is important to have knowledge of different purposes of assessment (diagnostic, ranking, promoting learning)
	20	It is important to have knowledge of large scale testing, such as national tests, PISA etc.
	23	It is important to have knowledge of language frameworks, such as CEFR.
	33	It is important to have knowledge of the history of language assessment