

Report Quantitative validation LASLLIAM February 2022

Jeanne Kurvers, Fernanda Minuz, Rola Naeb, Lorenzo Rocca & Karen Schramm (Author group LASLLIAM)
Sanneke Schouwstra & Remco Feskens (Cito)

1 Introduction

At the beginning of the European Year of Languages in 2001, the Council of Europe officially launched the Common European Framework of Reference (CEFR). The CEFR “was designed to provide a transparent, coherent and comprehensive basis for the elaboration of language syllabuses and curriculum guidelines, the design of teaching and learning materials, and the assessment of foreign language proficiency. The CEFR has been translated into over 40 languages and is used all over Europe and in other continents”.¹ Today, the Common European Framework of Reference for Languages can be considered one of the best-known and most used Council of Europe policy instruments (Council of Europe, 2020).

The CEFR is an extensively used framework for which resources have been developed to support the implementation and in 2018 a Companion volume has been published with new descriptors, following the requests of practitioners. In the same year, a project started with the Education Policy Division of the Council of Europe to develop a European Framework of Reference for Literacy and Second language learning of non-literate and low-literate adult migrants.

This reference guide on *Literacy and Second Language Learning for the Linguistic Integration of Adult Migrants* (LASLLIAM) is linked to the CEFR and is intended to build on the CEFR Companion volume. LASLLIAM aims “to support language educators, curriculum designers and language policy makers in their endeavour to design, implement and evaluate curricula, syllabi and teaching materials tailored towards the specific needs of the target learners.”²

The LASLLIAM guide contains descriptors for levels below and up to the A1 level of the CEFR for migrant learners who either enter a new country or are refugees with no or hardly any previous schooling, or for low-literate migrants, who can read and write (in a non-alphabetic script) and learn the language(s) of the host countries; they are regularly referred to as LESLLA learners.³

¹ <https://www.coe.int/en/web/common-european-framework-reference-languages/introduction-and-context>

² *Literacy and Second Language Learning for the Linguistic Integration of Adult Migrants*, p. 13

³ See www.leslla.org

The guide is designed for the developers of curricula, teaching material and assessment tools and for teachers of literacy-and-second language learners and provides descriptors for oral and written communicative language activities and language use strategies (i.e. reception, production, and interaction) as well as for technical literacy and digital skills.

The first version of the LASLLIAM guide was thoroughly based on empirical research and already validated frameworks in some European countries and on the feedback of a group of five experts⁴ in the field on draft versions of the guide and the scales (see chapter 1-3 in the Reference Guide). After that, the validation of the descriptors and scales started using a mixed-method approach. The first validation study was a qualitative study in which feedback was collected regarding descriptors from the LASLLIAM scales. During face to face and/or online workshops feedback was collected about the clarity of the descriptors, the pedagogic usefulness and relevance to real life use. Besides, feedback was also collected on the level assignment of the Technical Literacy scales and on the assignment to categories of the Specific Communicative language activities.

The outcomes of that study revealed that overall clarity, relevance for real life and the pedagogical usefulness of most descriptors was agreed upon by nearly all participants in the workshops. The scales on Technical literacy, the Overall and Specific scales, Communicative language activities and Language use strategies, and the Digital skills scales revealed for clarity, pedagogical usefulness and relevance for real life on average over 90% agreement. Of the 478 descriptors included in the qualitative validation, at least 90% agreement was received for clarity on 77% of descriptors, for pedagogical usefulness on 80% of the descriptors and for relevance to real life on 87% of the descriptors.⁵ This indicates that the respondents in the workshop recognised the aim and the quality of most of the descriptors. Descriptors that had scores below 80% agreement and/or were commented upon by several workshops members were deleted or revised in accordance with the suggestions and a few were added; other descriptors were revised to preserve consistency in wording and the progression line in scaling. (See report on qualitative validation for details).

The next phase of the validation, the focus of this report, was a quantitative validation study on clarity and the scaling of the descriptors. Pedagogical usefulness and relevance for real life were not included for judgement anymore, because all descriptors were judged as useful and relevant by at least 80% of the respondents in the qualitative study. Regarding clarity, the data were mainly used to see whether the positive results of the qualitative workshops could be confirmed. Because the number of responses in the first round was too low to validate the scaling progression of the descriptors of several scales, and because the representativeness of European countries could be improved, a second round was added with a survey in ten more countries.

2 The quantitative validation study

The quantitative validation study was conducted in two steps between April and December 2021 among practitioners experienced in the field of literacy and/or second language learning and assessment of low-educated and non-literate learners. The aims of the quantitative validation were to corroborate the results on

⁴ These experts were: Jean Claude Beacco, Kaatje Dalderop, Bart Deygers, Cecilie Hamnes Carlsen and David Little.

⁵ The 90 descriptors from the CEFR Companion Volume that were included as well were not judged.

clarity from the qualitative validation of the descriptors and to validate the scaling progression of the descriptors of the Technical Literacy scales, the Communicative Language Activities scales and the Language use strategies. The validation consisted of teachers' assessments of the difficulty of the descriptors.

Although, as North points out, true and complete calibration and validation of the descriptors requires assessment of language behaviour of real language learners, validation with experienced literacy and second language teachers is a necessary and important part of a validation study to investigate whether "the descriptors are placed at the appropriate level on the scheme of levels, and that this placement is validated by data, and not just the opinion of the authors or individual informants." (North, 2020: 26).

Research questions

To reach the aims of the quantitative validation, checking the previous results on clarity and validating the scaling progression of the descriptors, the following research questions were formulated:

1. To which extent do the practitioners consider the descriptors of the LASLLIAM scales clear?
2. Which LASLLIAM level (1-4) do the practitioners assign to the descriptors of the LASLLIAM Technical literacy scales, the Overall Scales and the Specific Scales of Communicative language activities and to which extent do the levels as assigned corroborate with the intended levels?
3. To which extent do the practitioners consider the descriptors of the LASLLIAM Language Use Strategies Scales more or less demanding and to which extent do the levels of demandingness as assigned corroborate with the intended levels?
4. Does the assessment of the LASLLIAM descriptors depend on participants' background like age, gender, country, experience with LESLLA learners, and familiarity with CEFR levels?

The first question was included to see if the (positive) findings about clarity from the qualitative validation study could be confirmed. The second and third question aimed at validating the difficulty level and the progression line of the descriptors in the various scales. The last question was added to check whether the evaluation of descriptors might have been impacted by background characteristics of the respondents.

In the following section (section 3), we first present a description and rationale for the methodology in the study of the first step (3.1) and the second step (3.2) respectively, and the analyses of the data (3.3). In section 4 we first present an overview of the participants in the validation study (4.1) and after that the results on clarity of all descriptors (4.2), on level assignment for the Technical literacy scales and the Communicative language activities (4.3) and about the level of demandingness of the Language use strategy scales (section 4.4). In section 4 we present the outcomes of the combined data of the two rounds. In section 5 we will explain the criteria for revision of descriptors and scales and provide an overview of the adaptations made for the descriptors in all scales. The report will close with a conclusion in section 6.

3 Method

3.1 Method quantitative validation: first step

The first step of the quantitative validation was conducted with an online survey. The participants were recruited through the literacy-and-second-language networks in the different countries, the connections already established in the qualitative validation, the international LESLLA network, LIAM and ALTE networks and other relevant Council of Europe networks. They were expected to have experience with literacy and language teaching of LESLLA learners, to be familiar with the CEFR and proficient enough in English to judge descriptors in English. Participants were invited to fill out an online survey.

The survey included questions to collect background information about the participants' profiles, particularly age, gender, language(s) taught, years of experience with L2 literacy learners and familiarity with CEFR levels and language assessment. Participants were introduced to LASLLIAM and the concept of progression lines/criteria at the start of the survey. Then, they were expected to complete five tasks with descriptors to be judged as detailed in Table 1. For clarity, the respondents were asked to indicate whether the descriptor was clear (yes/no); for assignment to level, they were asked to assign each descriptor to one of the four LASLLIAM levels. For the Language use strategy scales, the respondents were asked to indicate whether a descriptor was less or more demanding, less demanding referring to the LASLLIAM levels 1 and 2, and more demanding to the levels 2 and 3. Because in the validation processes of the CEFR and CEFR Companion Volume evaluation of the levels for strategies was not included, we decided for a first step in validating the strategy descriptors. We included assessing the level of demandingness (high demanding for level 3 and 4 or low demanding for level 1 and 2) of the strategies. This can be considered a base for further validating and standard setting of the strategy descriptors. The descriptors of the Digital skills were only judged on clarity, because the digital scales are independent of the other LASLLIAM scales.

In order to avoid misunderstandings and keep the respondents familiar with the specific (literacy and-language-related) progression lines and terms, a glossary with key terms present in the descriptors (such as phoneme, sight word, synthesise, simple sentence, turn, etc.) was provided as additional material to give respondents the possibility to look up these terms.

Table 1 – Tasks, scales and requirements in the first step of the quantitative validation

Task	Scales and descriptors	Judgment
Task 1	Technical Literacy scales	Clarity (yes-no) Assignment to LASLLIAM levels (1-4)
Task 2	Overall scales	Clarity (yes-no) Assignment to LASLLIAM levels (1-4)
Task 3	Specific scales*	Clarity (yes-no) assignment to LASLLIAM levels (1-4)
Task 4	Language Use Strategies scales	Clarity (yes-no) Assignment to level of demandingness (less demanding-more demanding)
Task 5	Digital Skills scales	Clarity (yes-no)

* to allow respondents to better contextualize descriptors in the Specific scales, examples from the four domains were presented

Table 2 presents the scales and numbers of descriptors used in the first step of the quantitative validation.

Table 2: Number of LASLLIAM descriptors included in the quantitative validation (first step)

Scale		No of descriptors
Technical Literacy scales (total 78)	Language and Print Awareness	22
	Reading	31
	Writing	25
Overall scales (total 37)	Oral Reception	4
	Written Reception	8
	Oral Production	5
	Written Production	6
	Oral Interaction	5
	Written Interaction	9
Specific scales (total 131)	Oral Reception	22
	Written Reception	24
	Oral Production	13
	Written Production	13
	Oral Interaction	40
	Written Interaction	19
Language Use Strategies scales (total 139)	Oral Reception Strategies	17
	Written Reception Strategies	21
	Oral Production Strategies	23
	Written Production Strategies	18
	Oral Interaction Strategies	32
	Written Interaction Strategies	28
Digital Skills scales (total 70)	Communication and Collaboration	32
	Content Creation and Management	24
	Safety	14
Total no of descriptors		455

To keep the number of descriptors to be judged feasible, and the time needed to answer all questions within a reasonable limit, 18 different versions of the validation survey were developed. The number of descriptors in each version ranged between 65 and 70, in line with the numbers used in the CEFR Companion Volume validation (see Appendix 1 for further details). To ensure a balanced distribution of the descriptors across scales and levels in the different versions, the descriptors were randomized taking into account the following criteria:

- All the 455 descriptors were covered.
- In Task 1 a sample of the three categories of the Technical Literacy scales was always provided.
- In Task 2 and in Task 3, at least one full scale was always given.
- In Task 4 one full scale was always given.
- Within Task 2, Task 3 and Task 4 a coherent connection was provided, as follows:
 - o The sample in Task 3 taken from the Specific scales with domains' examples corresponded to the related full Overall scale given in Task 2.
 - o The sample in Task 4 taken from the Language Use Strategy scales was related to the scales rated in Task 2 and 3 (e.g. if Oral Reception was judged in Task 2 as full Overall scale, in Task 3 descriptors from Specific scales related to Oral Reception were provided, and in Task 4 descriptors from the Oral strategies). This enabled the respondents to base their judgements on the relation and consistency across descriptors present in different type of scales.

In the first three tasks, for each descriptor (in addition to judge its "clarity") participants were asked to assign a LASLLIAM level. To be able to enter the data of the version into one dataset, common descriptors were used in every version at the start of these tasks part of which were 'anchors': already calibrated descriptors from the CEFR Companion Volume and the Technical Literacy scales (see Appendix 1 for details).

The 18 versions were entered into SurveyMonkey, a cloud-based survey tool used by the Council of Europe for online surveys that enables quick collating of data and possibilities to enter the data into one statistical dataset.

3.2 Method quantitative validation: second step

In the first round, 296 respondents from 15 different countries⁶, assessed (sometimes a part of the) descriptors in one of the versions, leading to a low number of descriptors assessed in some of the scales, especially in tasks 4 and 5. In order to collect more data to validate the difficulty level of the descriptors, and to guarantee a representative sample of respondents from contexts and languages among the CoE member states, a second step of quantitative validation was conducted in an additional number of 11 countries.⁷

Because the judgments on clarity were mainly aimed at confirming the overall positive outcome of the qualitative validation study, and the first step revealed that 94% of the descriptors were considered clear by the respondents, it was decided to only ask for assignment to level or degree of demandingness in the second round. Participants of the second step were expected to have the same profile of the ones involved in the first step. The procedure was mainly the same as in the first round: participants were introduced to LASLLIAM and the concept of progression lines/criteria at the start of the survey, they were asked to judge in the same way. Because the time span to fill out the questionnaire was less, no different versions were designed for the Technical Literacy Scales and the Overall and Specific Communicative language activities scales. Only in one task, the assessment of the Language use strategies with a high number of descriptors, the descriptors were split up into two different versions, one for the descriptors about oral strategies and one for the descriptors

⁶ Most of them from Italy, UK and Ireland, Germany, Netherlands and Flanders, Norway, Spain, Greece, and a few from Finland, Sweden, Romania, Czech Republic, Lithuania, Luxembourg, Albania and Montenegro

⁷ Bulgaria, Denmark, Ireland, Slovenia, Portugal, France, Norway, Catalonia, Turkey, Spain and UK.

about written strategies. The division of scales among tasks was designed to keep the time for each task within about the same time limit. Table 3 presents the four tasks with the number of descriptors to be judged. Different from the online survey in the first round (and to keep track of the time needed), in round two a coordinator in each of the countries checked the availability of the tasks, took care of the division of task 4 among respondents, and collected the data.

Table 3: LASLLIAM quantitative validation's tasks and scales (second step)

Task	Scales and descriptors	Judgment	Nr of scales	Nr of descriptors
Task 1	Technical Literacy	assignment to LASLLIAM level (1-4)	3	69
Task 2	Communicative Language Activities (overall and specific written)	assignment to LASLLIAM level (1-4)	12	65
Task 3	Communicative Language Activities (overall and specific oral)	assignment to LASLLIAM level (1-4)	15	80
Task 4	Language Use Strategies Group A (written) Group B (oral)	assignment to level of demandingness (less-more)	9 (for each group)	(A) 68 (B) 72

The same random Id number for descriptors used in the qualitative validation, as well as in the first step of the quantitative validation, was maintained in order to be able to combine the data. Also in this step, in order to ensure a balanced distribution of the descriptors across scales and levels in the different versions, the descriptors were randomized taking into account the following criteria:

- All the categories and all the descriptors were covered.
- In all the tasks, the full scale was always given (including domains' examples, where provided).
- In the Overall and Specific scales, the descriptors that were already validated in the first round of the quantitative validation were not included anymore. This means if more than 200 hundred responses were already collected, clarity was confirmed by at least 80% agreement and mode and mean confirmed the intended level.⁸

In addition:

- The descriptors taken from CEFR Companion Volume were already validated earlier and for this reason they were not validated again.
- For task 4 the respondents were divided in two groups (A and B) in coherence with the division made between task 2 and task 3: group A was asked to work on the written dimension, group B to work on the oral dimension.

3.3 Analysis of the data

⁸ Only some of those already validated descriptors that were needed to keep all the levels of a scale in the second round were also included in the second round.

Cito, institute for educational measurement in the Netherlands, combined the data collected in the two rounds into a single data file suitable for analyses in R. R is a free software environment for statistical computing and graphics (<https://www.r-project.org/>).

Based on the common items in the first round of data collection and the ID-numbers of the descriptors, the data from the different versions and the two rounds were merged into one data-file. Some descriptors were corrected, due to minor mistakes in ID-number, spelling or punctuation (e.g. a comma or space missing in one of the versions). One descriptor (ID 28) turned out to be missing in the second round, and one (ID 226) was used in two scales.

Respondents who did not meet two of the expected background criteria (at least three years of experience with LASLLIAM learners and familiar with the CEFR) as well as one respondent who filled out the survey in an unexpected way were excluded. Respondents who had not answered one or more of the background questions were not excluded. In total 424 respondents were kept in the combined data-file.

Although North (2020) recommends using IRT/Rasch analysis as the best way to identify and exclude descriptors and respondents that behave in improbable and inconsistent ways, and to be able to generalise beyond the context(s) and group(s) from which the data was collected, there were some reasons not (to be able) to choose IRT/Rasch analysis for our aims. First of all, the data we collected are not test items of adult literacy students, but teachers' opinions about the difficulty level of descriptors for the target learners. Secondly, IRT requires a considerable number of responses (preferably more than 400 responses for each descriptor) and much variation in the data. The number of responses ranged for most descriptors between 120 and 200 (except for the overall scales), and in particular on clarity there was much agreement among respondents. Lack of variation is a promising sign for the intended scales, but not adequate for using IRT. And thirdly, CITO only uses IRT for removing test-items that do not fit with the mathematical model, but not for removing respondents with different opinions about the clarity or difficulty level of the descriptors. We consider IRT/Rasch analysis to be more relevant in a further validation process when learners' data are collected and analysed.

Therefore only descriptive statistics and classification analysis were used to analyse the data and validate the judgments of the respondents. The analysis included:

- Collation of raw ratings to percentages (for all the tasks).
- Descriptive statistics including frequencies, mode, mean and standard deviation for all the tasks.
- Comparative analyses of the assigned levels with the intended levels (for Technical Literacy and the Overall and Specific Communicative language activities) and of the level of demandingness (for the Language use strategies)

For assignment to levels, this means comparing the agreement of the respondents with the intended levels of the descriptors provided in the LASLLIAM scales. The criteria used for deciding on agreement were the following: 'Agreement on the intended level' was indicated when the mode (the most often chosen level by the respondents) and the mean of the assigned level confirmed the intended level. 'Agreement on another level' was indicated if at least 75% of the respondents judged a descriptor at another level than the intended one. 'Disagreement' was indicated when mode and mean of the assigned level differed from the intended level. For the language use strategies 'Agreement' indicated the majority of the respondents confirmed the intended level

of demandingness, 'disagreement' if the majority assigned a descriptor to another level of demandingness than the intended one and 'agreement on another level of demandingness' if at least 75% of the respondents did choose for another than the intended level of demandingness.

4. Results of quantitative validation

4.1 Participants

In total 424 respondents who assessed descriptors participated in the two rounds of the quantitative validation study, 279 in the first round, and 145 in the second round. Table 4 presents the background characteristics of the participants.

Table 4: Background data participants in the study

	Group	number	percentage
gender	Female	336	79.8
	Male	82	19.5
	Prefer not to say	3	0.7
age	Under 30	82	19.3
	31-40	145	34.2
	41-50	108	25.5
	51-60	76	17.9
	Over 60	13	3.07
Experience	Less than 3 years	92	22.5
	3-5 years	74	18.1
	6-10 years	97	23.8
	more than 10 years	145	35.5
Specific Literacy training	No	222	56.2
	Yes	173	43.8
Familiarity with CEFR	Not entirely familiar	12	2.8
	Yes substantially familiar	176	41.6
	Yes very familiar	235	55.6
Languages taught	Italian	109	25.9
	English	83	19.8
	Dutch	36	8.6
	German	33	7.9
	Norwegian	21	5.0
	French	18	4.3
	Slovene	17	4.0
	Bulgarian	16	3.8
	Danish	16	3.8
	Portuguese	14	3.3

	Spanish	11	2.7
	Catalan	9	2.1
	Greek	8	1.9
	Other*	29	6.9

*Other= Finnish, Turkish, Swedish, Czech, Japanese, Romanian, Albanian, Basque, Lithuanian, Luxembourgish, Montenegrin, Esperanto, none (each 1-5 respondents).

The vast majority of the participants is female (80%) and nearly 60% are between 31-50 years old. The younger and older age groups are slightly less represented. The participants are practitioners experienced in the field of literacy and/or second language learning and assessment: nearly 80% of the participants have at least 3 years of experience with LESLLA learners, more than one third more than 10 years of experience. In contrast, almost a quarter of the participants have less than 3 years of experience with L2 literacy learners, the remaining participants between 3 and 10 years. Despite their experience with LESLLA learners, less than half of the respondents received specialist training regarding the target group. Almost all participants indicated that they were “substantially familiar” or “very familiar” with the CEFR levels. Most participants are involved in teaching one language (87%), but some are involved in teaching 2 or more languages or none. Overall 25 different languages are reported. About a quarter of the respondents mention that the first language they teach LESLLA learners is Italian, about 20% mention English. Dutch, German, French, Norwegian and Slovenian are mentioned by between 4% and 9% of the respondents. Other languages (n=17) are mentioned by less than 4% of the participants, 10 of which only by one or two participants.

4.2 Clarity

Clarity of the descriptors was only judged in the first round of the quantitative validation, the online Survey with 279 respondents. The descriptors were divided among the 18 versions. Table 5 summarizes the results for the descriptors of the different scales. The number of CEFR descriptors that were added in the survey (as common items or to realize completeness) are not included in the summary of judgments on clarity.

Table 5: Scales and number of descriptors as rated for clarity

Category	Scale	Range responses	Number of descriptors*	Rated clear by 80-100%	Rated clear by <80%
Technical Literacy	Reading	13-279**	31	31	0
	Writing	15-279	25	25	0
	LPA	13-57	22	21	1
	<i>Total TL</i>		78	77	1
Overall scales	Overall Spoken Interaction	42-217	4	4	0
	Overall Spoken reception	217	3	3	0
	Overall Spoken production	217	4	4	0

	Overall Written Interaction	40-217	8	8	0
	Overall Written reception	31-217	7	7	0
	Overall Written Production	38-217	6	6	0
	<i>Total Overall scales</i>		32	32	0
Specific Scales	Spoken Interaction	11-41	33	33	0
	Spoken reception	11-30	18	18	0
	Spoken production	29	11	11	0
	Written Interaction	11-38	16	15	1
	Written reception	9-31	17	17	0
	Written Production	35	12	12	0
	<i>Total Specific scales</i>		107	106	1
Language use strategies	Spoken Interaction	11-18	32	29	3
	Spoken Reception	7-12	17	17	0
	Spoken Production	6-12	23	23	0
	Written Interaction	11-15	28	28	0
	Written Reception	7-11	21	18	3
	Written Production	5-16	19	18	1
	<i>Total LUS</i>		140	133	7
Digital skills	Collaboration and Co-operation	7-18	29	26	3
	Safety	11-27	13	12	1
	Content creation and management	5-15	24	24	0
	<i>Total Digital skills</i>		66	62	4
Total			423	410	13

* CEFR Companion Volume descriptors not included

** The highest number in the range of TL scales and the Overall scales indicates the number of responses for the common items in all versions

Overall, the results confirmed the outcomes of the qualitative validation study: nearly all descriptors were considered clear by the respondents (m=94%). For all descriptors we found that the majority of participants considered them clear. For many descriptors (175) all participants indicated that the descriptor was clear (m=100%). For only 13 descriptors (3%) less than 80% of the participants considered them clear and only for two descriptors less than 70% of the participants considered them clear.

Technical Literacy

The Technical Literacy Scales consisted of three subscales: Reading, Writing and Language and Print awareness. All participants evaluated four descriptors of the Reading scale and four descriptors of the Writing scale (the anchors). The other descriptors were distributed across the 18 versions of the survey. All except one of the 78

descriptors were considered clear by at least 80% of the participants, 64 of them by at least 90% of the participants.

The descriptors of Reading were considered clear. The percentage of participants that considered the descriptors clear was between 82% and 100%. The 25 descriptors of Writing were also considered clear. The percentage of participants that considered the descriptors clear was between 80% and 100%. The 22 descriptors of Language and Print Awareness were also considered clear. Only slightly fewer participants considered the descriptor “Can synthesise spoken words into short and simple sentences” clear ($m=73\%$). Between 89% and 100% of the participants considered the other 21 descriptors clear.

Communicative Language activities

Overall scales

The Overall Scales consisted of six scales: Overall Oral Interaction, Overall Oral Reception, Overall Oral Production, Overall Written Interaction, Overall Written Reception, and Overall Written Production.

The descriptors of the Overall Scales were considered clear. The percentage of participants indicating that a descriptor was clear varied between 92% and 100%.

Specific Scales

The Specific scales covered the same global categories as the Overall scales: Oral Interaction, Oral Reception, Oral Production, Written Interaction, Written Reception, and Written Production.

The descriptors were contextualized by the examples of language use in the 4 domains (personal, public, occupational and educational). In each version descriptors of only one of the six subscales were presented. Few participants evaluated the descriptors of the Specific scales (less than fifty).

Oral interaction contains 33 descriptors about Conversation, Goal-oriented co-operation, Informal discussion, Information exchange, Interviewing and being interviewed, and Obtaining goods and services.

All 33 descriptors are considered clear by at least 83% of the respondents. For 30 out of the 33 descriptors, 90% to 100 of the participants consider the descriptors clear.

Oral reception contains in total 18 descriptors: about Listening as a Member of a Live Audience, Listening to Announcements and Instructions, Listening to audio media and recordings, watching TV and video and Understanding conversation between other speakers. For all descriptors we see that at least 91% of the participants consider the descriptors clear.

Oral production contains 11 descriptors about sustained monologue: Describing experience and Giving information. At least 83% of the participants considered the descriptors clear and for 9 out of 11 more than 90%.

Written interaction contains 16 descriptors about (Offline and Online) Correspondence and (Offline and Online) Notes, Messages, Forms and Transactions. One descriptor was considered slightly less clear: “Can ask for or report personal details in areas of immediate need in an everyday context” ($m = 75\%$). All other descriptors were considered clear by at least 90% of the participants.

Written reception contains in total 17 descriptors about Reading as a leisure time activity, Reading correspondence, Reading for information, Reading for orientation and Reading instructions. At least 86% of the participants considered the written reception descriptors clear.

Written production, finally, contains 12 descriptors about Creative writing and Functional writing. More than 92% of the participants considered each of the descriptors clear.

Language Use Strategies scales

In LASLLIAM the Language Use Strategies are related to the communicative language activities for both oral and written language. Both for oral and written language, the Language Use Strategies contain 3 categories of descriptors: descriptors for Planning, Compensating and Monitoring and Repair. The Language Use Strategies scales contain in total 140 descriptors for Oral Interaction, Oral Reception, Oral Production, Written Interaction, Written Reception and Written Production. The descriptors included in each version were related to the scales in task 2 and task 3. The number of reactions for each of the descriptors was low (between 5 and 19 respondents) and these results should be interpreted with caution. Overall, 133 of the 140 descriptors (95%) were considered clear by at least 80% of the participants, and 85 descriptors (61%) were considered clear by all respondents. This also confirms the outcomes of the qualitative validation.

Oral Interaction strategies

Eleven descriptors (out of 32) are considered clear by all respondents. However, we find also more descriptors that are considered less clear than was the case with the previous scales. Less than 80% consider the following three descriptors clear: “Can use verbal imitation to build rapport” (78%), “Can use simple expressions (e.g. ‘Clear?’), repetition or translation into L1 or L3 to ensure interlocutor’s comprehension” (75%), and “Can imitate words or phrases to maintain rapport” (78%). Another seven descriptors are considered clear for 80 to 90% of the participants.

Oral Reception strategies

At least 83% of the respondents consider the 17 descriptors of oral reception clear, even 100% of the participants consider 15 of the descriptors clear.

Oral Production strategies

Also, the 23 descriptors of Oral Production are considered clear: the mean clarity is between 80 and 100%, and 11 descriptors are considered clear by all participants.

Written Interaction strategies

The descriptors of Written Interaction (28 in total) are considered clear for 83 to 100% of the participants, all except one by at least 90% and 17 by all participants.

Written Reception strategies

Fifteen of the 21 Written Reception strategy descriptors are considered clear by all reacting participants ($m = 100\%$). However, we find also three descriptors that less than 80% of the reacting participants considered clear: “Can highlight words and phrases that s/he understands well to monitor the meaning.” (71%), “Can identify unknown element in a picture (e.g. object in a picture story) to ask for the word.” (71%) and “Can mark an unclear phrase to ask for the meaning.” (71%).

Written Production strategies

Of the 19 descriptors 17 were considered clear by at least 94% of the respondents. One descriptor “Can outline the structure to write a simple text (e.g. where - when - what?)” was considered clear by 80%, and one by less than 80% of the respondents: “Can use words from their plurilingual repertoire to maintain writing in the second language” (77%).

Digital Skills scales

In LASLLIAM three Digital skills are included: Collaboration and Co-operation, Safety, and Content creation and management.

In each version one of the three Digital scales was included. The participants are only asked to indicate whether they consider the descriptors clear. Very few participants evaluated the descriptors of the Digital Skills scales (between 5 and 20 participants). The results should be interpreted with utmost caution. Overall, 62 of the 66 descriptors (94%) were considered clear by at least 80% of the participants.

Collaboration and Co-operation

Twenty-six of the 29 descriptors were considered clear by more than 80% of the respondents, 19 of them by more than 90% of the reacting participants. Three descriptors were considered clear by less than 80% of the participants. These were: “Can use simple, personally relevant software with guidance (e.g. a learning tool, web browser)” (64%), “Can produce audio-visual files with guidance” (67%), and “Can manage a contact list (e.g. sort contacts into groups) (72%).

Safety

More than 82% of the reacting participants considered 12 of the 13 other descriptors clear. One of the 13 descriptors was considered less clear (75%): “Can recognise the essential safety issues using visual clues (e.g. error message icon like a red x)”.

Content creation and management

At least 80% of the reacting participants considered each of the 24 descriptors clear. Half of the descriptors were considered clear by all (m = 100%).

Summarizing, nearly all descriptors were considered clear by more than 80% of the respondents, in most cases by more than 90% of them. On average 97% of the descriptors of all scales were considered clear. All descriptors of the overall scales were considered clear by more than 80% of the respondents, and all except one descriptor of the Technical Literacy Scales and the Specific scales were considered clear as well. For Language Use strategies and Digital skills, respectively only 7 (5%) and 5 (6%) of the descriptors were considered clear by less than 80% of the respondents.

4.3 Assignment to level

Table 6 summarizes the outcomes of the assignment to level for the Technical Literacy scales, the Overall and Specific scales of the Communicative Activities, and the level of demandingness for the Language Use strategies. The descriptors of the Digital skills scales were not judged for level of difficulty. As already indicated: ‘Agreement on intended level’ in the table means mode and mean the same as the intended level, ‘agreement on another

level’ at least 75% of the respondents judged a descriptor at another level than the intended one, and ‘disagreement on level’ if mode and mean did not equal the intended level. For the language use strategies agreement indicates the majority of the respondents judged a strategy at the same level of demandingness than the intended one, disagreement if the majority choose another level and agreement on another level of demandingness if at least 75% of the respondents did choose for another than the intended level of demandingness. As explained before, the CEFR descriptors that were included in the questionnaire are not included in the summary table. The number of ratings of the Technical Literacy descriptors ranged between 169 and 279, for the Overall scales between 180 and 383, for the Specific scales between 155⁹ and 198, and for the Language use strategies between 118 and 135 (See Appendix 2 for details).

Table 6: Scales and number of descriptors that revealed agreement or disagreement with the intended level (CEFR descriptors not included in agreement numbers)

Category	Scale	Range responses	Number of descriptors*	Agreement on intended level	Agreement on another level	Disagreement on level **
Technical Literacy	Reading	179-279	31	23	0	8
	Writing	175-279	25	21	0	4
	LPA	170-212	22	17	0	5
	<i>Total TL</i>		78	61	0	17
Overall scales	Oral Interaction	192-367	4	2	0	2
	Oral Reception	373-374	3	3	0	0
	Oral Production	185-369	4	4	0	0
	Written Interaction	190-367	8	4	0	4
	Written Reception	197-383	7	5	0	2
	Written Production	200-217	5	4	0	1
	<i>Total Overall scales</i>		31	22	0	9
Specific Scales	Oral Interaction	160-191	33	20	0	13
	Oral Reception	155-187	20	8	0	12
	Oral Production	180-183	11	6	0	5
	Written Interaction	165-193	16	5	1	10
	Written Reception	170-190	18	8	0	10
	Written Production	197-199	12	7	0	5
	<i>Total Specific scales</i>		110	54	1	55
	Oral Interaction	126-135	32	24	1	7

⁹ Except for one descriptor (nr 28) that was missing in the second round and only rated by 11 respondents in one version of the first round.

Language use strategies	Oral Reception	120-126	17	12	0	5
	Oral Production	119-127	23	19	1	3
	Written Interaction	118-126	28	23	0	5
	Written Reception	121-126	21	18	0	3
	Written Production	114-129	19	15	0	4
	<i>Total LUS</i>		<i>140</i>	<i>111</i>	<i>2</i>	<i>27</i>
Total			359	249 (70%)	3 (1%)	107 (29%)

* CEFR Companion Volume descriptors not included

** In most case mode and mean refer to adjacent level (73 out of 80 descriptors).

Overall, of the 359 descriptors rated for level or level of demandingness, 249 (70%) were rated by the largest number of the respondents (the mode) at the same level as intended (78% for the Technical Literacy scales, 74% for the Overall scales Communicative Language Activities, 50% for the Specific scales of Communicative Language activities and 86% for the Language Use strategies). As could be expected, the agreement was highest for the Language Use strategies, that only had to be rated at two difficulty levels (more or less demanding) than for rating at four levels for the other scales. There were only three descriptors (less than 1%) that were clearly rated at another level than the intended one (see below). For 107 descriptors (29%) there was disagreement among the respondents about the level. In nearly all of these cases (91%), the mode was one level above or one level below the intended level, while the second highest number of judgments was at the intended level in nearly all cases as well. Most disagreements could be noted for the Specific scales Oral and Written Interaction and Reception. The levels the respondents assigned to the 219 descriptors of the Technical Literacy Scales and the Overall and Specific Communicative Language Activities scales are on average nicely distributed between 1.24 (level 1) and 3.82 (level 4); the levels of demandingness for the 110 descriptors of the Language use strategies are nicely distributed between 0.03 (low demanding) and 0.97 (high demanding).

Technical Literacy

The levels participants attributed to the Technical Literacy Scales are on average distributed between 1.24 (level 1) and 3.43 (level 4). See Figure 1.

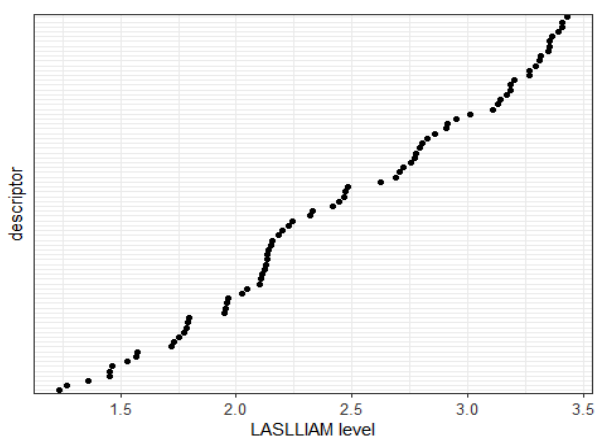


Figure 1: Level assignment descriptors Technical Literacy scales

The distribution of the average level is very similar for the three Technical Literacy Scales: Reading (1.24-3.43), Writing (1.27-3.32) and Language and Print Awareness (1.36-3.27).

Of the 31 descriptors of *Reading*, 23 were rated at the intended level. Of the 8 descriptors that were assigned differently, the mode of 4 descriptors was at one level above the intended one, of 3 one level below the intended one. Only one descriptor was rated by most respondents at level 1 (42% or level 2 (40%) instead of the intended level 4: “Can spell out relevant names letter by letter using the letter names (e.g. own name)”.

Of the 25 descriptors of *Writing* 21 were rated at the intended level. For 4 descriptors with disagreement, the mode of 3 descriptors was one level below the intended one, and once above the intended one.

Most of the 22 descriptors of *Language and Print Awareness* were also rated at the intended level, for five descriptors the mode was one level above the intended one, once below the intended one. One descriptor was rated by the largest group of respondents at level 4 instead of the intended level 2: “Can identify the stressed syllable(s) in a word.”

Overall scales

The level participants assigned to the overall scales are nicely distributed between 1.30 (level 1) and 3.42 (level 4). See Figure 2.

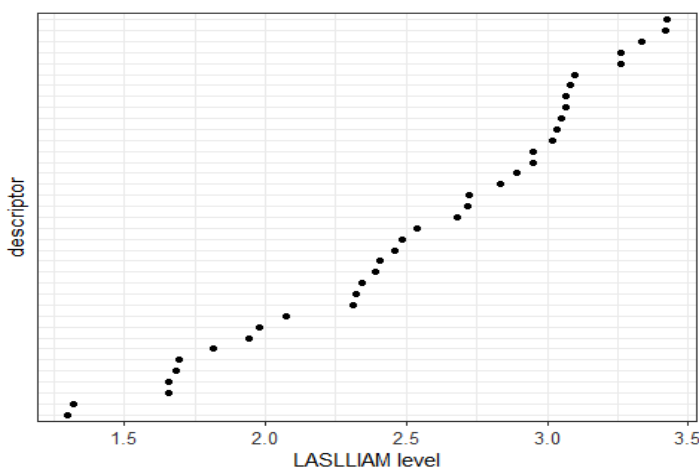


Figure 2: Average level assignment descriptors Overall scales.

The distribution among average levels is similar for the six scales: Oral Interaction (1.68-3.26), Oral Reception (1.69-2.89, level 4 was a CEFR descriptor), Oral Production (1.66-3.34), Written Interaction (1.66-3.42), Written Reception (1.32-3.42) and Written Production (1.30-3.26).

Of the 31 descriptors of the six Overall scales, the largest group of respondents agreed with the intended level for 22 descriptors. Disagreement about the intended level was observed for 9 descriptors, 2 of Oral Interaction, 4 of Written Interaction, 2 of Written Reception and 1 of Written Production. In all cases except one the most mentioned level was one level above (7 descriptors) or one level below (1 descriptor) the intended one. The exception was one descriptor of Written Reception that was rated at level 3 by 29% of the respondents, at level 2 by 27% of the respondents and at the intended level 1 by 26% of the respondents: “Can pick out a personally relevant piece of information in a text about a routine everyday topic by reading sight words and using visual and situational clues.”

Specific scales

Overall, the level participants assigned to the Specific scales are nicely distributed between 1.27 (level 1) and 3.82 (level 4). See figure 3.

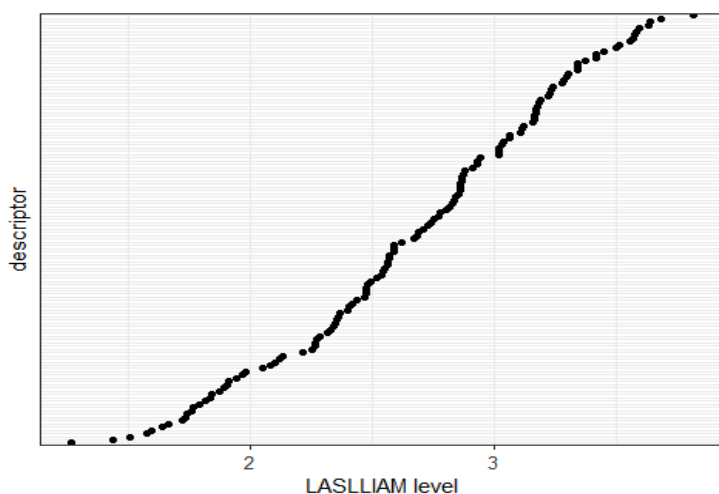


Figure 3: Average level assignment descriptors Specific scales

The distribution is more or less similar for the six scales, although the average for the lowest level is different in one of the scales: Oral Interaction (1.27-3.56), Oral Reception (1.73-3.42), Oral Production (1.72-3.62), Written Interaction (1.89-3.64), Written Reception (1.64-3.42) and Written Production (1.76-3.82).

About half of the 110 descriptors of the Specific scales (54) were rated by the largest number of respondents at the intended scale, for the other half (55 descriptors) there was disagreement and for one descriptor the respondents agreed on another level than the intended one.

Of the 33 descriptors *Oral Interaction*, 20 descriptors were assigned to the intended level, 7 descriptors were attributed to one level above the intended one, and 6 were assigned to one level below the intended one.

Of the 20 descriptors of *Oral Reception*, 8 were assigned to the intended level, 6 to one level above the intended one, and 6 to one level below the intended one.

Of the 11 descriptors of *Oral Production*, 6 were assigned to the intended level, 2 to one level above and 3 to one level below the intended one.

Of the 16 descriptors *Written Interaction*, 5 were assigned to the intended level, 9 to one level above the intended one, and 1 to one level below the intended one. One descriptor was assigned by 75% of the respondents to level 4 instead of the intended level 3: "Can write and respond to short messages (e.g., proposals, opinions, feelings) with simple sentences, phrases and formulaic expressions, sometimes using a common connector."

Of the 18 descriptors of *Written Reception*, 8 were assigned to the intended level, 7 to one level above the intended one, 1 to one level below the intended one, and for two descriptors the mode was different. The majority of the respondents (56%) assigned one descriptor at level 3 instead of the intended level 1: "Can distinguish personally relevant correspondence from irrelevant correspondence (e.g. impersonal mail)."

Of the 12 descriptors of *Written Production*, 7 were assigned to the intended level, 4 to one level above and 1 to one level below the intended one.

Language use strategies

The level of demandingness participants attributed to the Language use strategies is nicely distributed between 0.03 (attributed as low demanding by nearly all respondents) and 0.97 (attributed as high demanding by nearly all respondents). See Figure 4

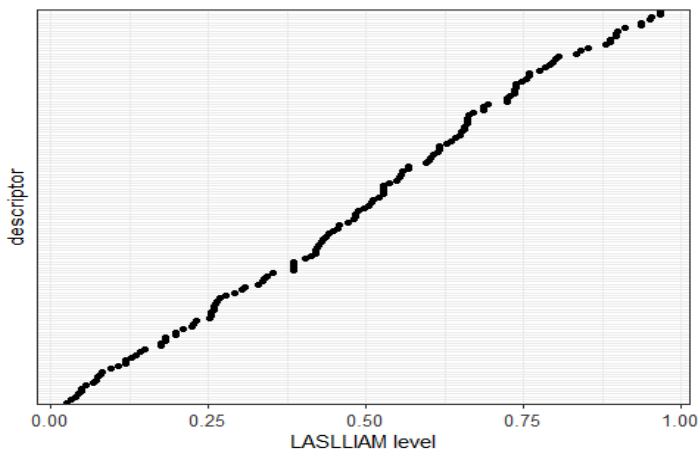


Figure 4: Average assignment level of demandingness descriptors Language use strategies

The distribution is very similar for the six scales: Oral Interaction strategies (0.04-0.90), Oral Reception strategies (0.14-0.94), Oral Production strategies (0.07-0.97), Written Interaction strategies (0.04-0.95), Written Reception strategies (0.03-0.97) and Written Production strategies (0.03-0.91).

Of the 140 Language strategies, 111 were assigned to the intended level of demandingness by the majority of the participants, 27 were assigned to another level of demandingness by a (small) majority of the respondents.

For 24 of the 32 *Oral Interaction* strategies the respondents agreed on the intended level of demandingness, 7 revealed disagreement and for 1 the respondents agreed on another level of demandingness. The descriptor “Can ask for repetition with words or phrases (e.g. “Please repeat”) to overcome problems in comprehension” was rated by 75% of the respondents as low demanding instead of the intended high demanding.

For 12 of the 17 *Oral Reception* strategies the respondents agreed on the intended level of demandingness, while for 5 descriptors they disagreed.

For 19 of the 23 *Oral Production* strategies the respondents agreed on the intended level of demandingness, for 3 descriptors they disagreed and for 1 descriptor the respondents agreed on another level of demandingness. The descriptor “Can use simple markers of self-correction (e.g. “No, no”) and body language to ease interlocutor's comprehension” was considered high demanding by 89% of the respondents instead of the intended low demanding.

For 23 of the 28 *Written Interaction* strategies the respondents agreed on the intended level of demandingness, while for 5 descriptors they disagreed.

For 18 of the 21 *Written Reception* strategies the respondents agreed on the intended level of demandingness, while they disagreed for 3 descriptors.

For 15 of the 19 *Written Production* strategies the respondents agreed on the intended level of demandingness, while for 4 descriptors they disagreed.

Summarizing, the respondents agreed with the intended level for 70% of the 359 descriptors, for 29% of the descriptors the average assignment was judged as slightly higher or lower than the intended one and for only 3 descriptors (1%) the majority of the respondents clearly judged for another level than the intended one. The slope of the lines in the Figures 1-4 clearly indicate that on average the progression line of the descriptors is as intended.

4.4 Comparison between groups of respondents

To check whether specific groups differed in their ratings we also compared the assignments on differences between gender, age group, Familiarity, experience with LASLLIAM learners, literacy training or language taught. Table 7 presents an overview of the comparative data.

Table 7: Average evaluation of clarity, level and level of demandingness by group: range respondents, mean and standard deviation

		Range N*	Clarity		Level assignment		Level of demandingness	
			Mean	sd	Mean	sd	Mean	sd
Gender	Female	223-336	0.95	0.22	2.58	1.01	0.49	0.5
	Male	55-82	0.94	0.23	2.54	1.00	0.52	0.5
Age	Under 30	9-13	0.98	0.13	2.51	0.97	0.55	0.5
	31-40	45-82	0.94	0.24	2.59	1.01	0.50	0.5
	41-50	94-145	0.95	0.22	2.56	1.01	0.49	0.5
	51-60	76-108	0.95	0.22	2.62	0.98	0.47	0.5
	over 60	55-76	0.95	0.21	2.48	1.02	0.49	0.5
Familiarity	Not entirely familiar	4-12	0.99	0.12	2.48	0.96	0.47	0.5
	Substantially familiar	111-176	0.94	0.24	2.61	1.00	0.50	0.5
	Very familiar	164-235	0.95	0.21	2.55	1.01	0.49	0.5
Experience	Less than 3 years	64-92	0.95	0.21	2.61	1.01	0.50	0.5
	3-5 years	51-74	0.95	0.23	2.56	0.99	0.49	0.5
	6-10 years	61-97	0.96	0.21	2.59	0.99	0.50	0.5
	more than 10 years	103-145	0.95	0.23	2.55	1.01	0.48	0.5
Lit Training	No	142-222	0.95	0.22	2.59	0.99	0.50	0.5
	Yes	137-173	0.95	0.22	2.55	1.02	0.48	0.5
Language**	Bulgarian***	16			2.69	0.90	0.41	0.49
	Dutch	26-36	0.96	0.20	2.54	1.04	0.55	0.50
	Danish	16			2.76	0.95	0.54	0.50
	English	45-83	0.93	0.25	2.64	1.02	0.51	0.50

	French	7-18	0.96	0.21	2.31	0.99	0.46	0.50
	German	25-33	0.96	0.20	2.45	1.06	0.54	0.50
	Italian	84-109	0.96	0.21	2.49	1.04	0.48	0.50
	Norwegian	11-21	0.92	0.27	2.61	1.06	0.44	0.50
	Slovene	17			2.62	0.96	0.49	0.50
	Other							

* Range refers to averages clarity, level assignment and assignment to level of demandingness.

**Only languages with at least 15 respondents are mentioned separately, all other languages included under other

*** Some languages were only included in the second round, so no data on clarity.

The differences between the groups are in general small for clarity, assignment to level and assignment to level of demandingness, and the standard deviations are rather similar in all groups. No significant differences between the background data gender, age groups, experience, training and familiarity with CEFR were revealed. And although the Danish assigned on average slightly higher levels to the descriptors, and the French slightly lower levels, also the differences between language groups were small overall.

5 Revision of scales and descriptors

Overall, the quantitative validation study revealed for most of the descriptors much agreement among respondents about clarity, assigned level and level of demandingness of the descriptors. After thorough discussion of the outcomes, the following criteria for keeping or deleting a descriptor, and incidentally for replacing or (slightly) revising a descriptor were used:

- A descriptor was deleted from the scales if less than 70% of the respondents considered the descriptor clear.
- Descriptors were kept if mode (the most mentioned) and mean of the level was the same as the intended level¹⁰
- A descriptor was moved to another level according to two conditions: if at least 75% of the respondents agreed on one specific level (other than the intended one) and if the moving did not affect the consistency of the scale, otherwise the descriptor was deleted.
- Descriptors that did not meet the criteria of mode and mean and were also not rated at another level by more than 75% of the respondents were deleted.
- In some cases, a similar descriptor from a related scale (e.g. production and interaction) that did meet the criteria replaced the original one. This criterion was applied when the deletion of the original one would have created a gap in a Specific scale.
- The descriptors that were taken from the CEFR Companion Volume were already validated and therefore not validated again. In total 71 descriptors from CEFR Companion Volume are integrated in the LASLLIAM scales for Communicative Language Activities. They were kept unchanged and they were completed in the Specific scales by tables of domains' example (see chapter 4 of the LASLLIAM Reference Guide).

¹⁰ Incidentally a tolerance range around the mean of about 10% was accepted.

- Incidentally a descriptor was slightly revised to correct an error or to keep consistency in the wording (e.g. message instead of text, deleting duplication or adding a missing word).

In total 85 LASLLIAM descriptors were deleted from the scales, 24 were replaced and of 32 descriptors the text was slightly revised to correct a mistake or to adapt to replacement. Appendix 3 presents an overview of all the scales and descriptors with the changes made clearly marked.

Table 8 presents an overview of the number of descriptors in each of the scales of the final version of LASLLIAM.

Table 8 – Number of descriptors in the final version of LASLLIAM

Scale		No of descriptors
Technical Literacy scales (total 59)	Language and Print Awareness	15
	Reading	24
	Writing	20
Overall scales* (total 41)	Oral Reception	6
	Written Reception	9
	Oral Production	4
	Written Production	7
	Oral Interaction	7
	Written Interaction	8
Specific scales* (total 143)	Oral Reception	24
	Written Reception	29
	Oral Production	12
	Written Production	12
	Oral Interaction	49
	Written Interaction	17
Language Use Strategies scales (total 119)	Oral Reception Strategies	14
	Written Reception Strategies	19
	Oral Production Strategies	21
	Written Production Strategies	18
	Oral Interaction Strategies	26
	Written Interaction Strategies	21
Digital Skills scales (total 63)	Communication and Collaboration	27
	Content Creation and Management	24
	Safety	12
Total no of descriptors		425

*including descriptors from CEFR Companion Volume

All in all, after careful revisions in several rounds of consultations with experts, qualitative validation workshops with 91 groups from 10 different countries and quantitative validation by more than 400 teachers in about 20 different countries, the 568 descriptors at the beginning of the validation process led to the 425 descriptors, with 354 new descriptors validated by experienced teachers teaching about 24 languages and completed with 71 descriptors already validated from the CEFR Companion Volume.

6 Conclusion

This study was conducted to validate the LASLLIAM descriptors that intended to describe progression line in L2 language and literacy development of non-literate and low-literate adult migrants. The study consisted of evaluation of the clarity of all descriptors in the LASLLIAM scales for Technical Literacy, Communicative Language activities (Overall, Specific and Language use strategies) and Digital skills, the rating of the levels of the scales for Technical Literacy and Communicative language activities and the level of demandingness of the Language use strategy descriptors. A major finding was that the descriptors in general were considered clear and that this outcome confirmed the findings of the qualitative study. The second major finding was that on average the intended scaling progression in the scales was recognized by the respondents. In cases where individual descriptors were not considered clear by at least 80% of the respondents, or respondents disagreed about the difficulty level, the descriptors were deleted or adapted, taking into account the outcomes of the quantitative validation. The whole procedure resulted in a final revised version of the LASLLIAM scales with 425 descriptors, including 71 descriptors from the CEFR Companion volume level pre-A1 and A1.

It should be noted that this study presented a validation based on ratings of experienced teachers. As North and Schneider (1998: 26) point out: “A claim to absolute objectivity in the scaling would, however, be overstated. What is actually being scaled here is the teachers’ collective interpretation of the difficulty of these descriptors.” Although this is considered not necessarily synonymous with objectivity, it is a very important step in validating the progression line in the scales. Further calibration and validation can be reached by using learner data. Besides, the descriptors of the Language use strategies were only evaluated on level of demandingness as a first step and the descriptors of the Digital skills were only rated on clarity. Future research is recommended to continue validating LASLLIAM.

References

- Council of Europe (2001), *Common European Framework of Reference for Languages: learning, teaching, assessment*, Cambridge University Press, Cambridge.
- Council of Europe (2020), *Common European Framework of Reference for Languages: learning, teaching, assessment: companion volume*, Council of Europe, Strasbourg (www.coe.int/lang-cefr).
- Minuz, F., Kurvers, J., Naeb, R. Rocca, L. & Schramm, K. (2022). *Literacy and second language learning for the linguistic integration of adult migrants (LASLLIAM)*. Council of Europe, Strasbourg.
- North, B. (2020). *The CEFR illustrative descriptors: Validation reference paper for researchers*. Council of Europe, Strasbourg.

North, B. & Schneider, G. (1998) Scaling descriptors for language proficiency scales. *Language Testing*, 15 (2) 217–263.

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Appendices

1. Design tasks in steps 1 and 2 quantitative validation
2. Mode, mean, standard deviation and number of respondents for all descriptor ratings for level and level of demandingness
3. Revised descriptors with changes marked.