

CEFR-Based Assessment in Sign Languages: The Case of LSE and Perspectives for LIS

Maria Grazia Marrocu

Ca' Foscari University of Venice, Rey Juan Carlos University of Madrid
mariagrazia.marrocu@unive.it, mg.marrocu.2025@alumnos.urjc.es

Abstract

This study examines the application of the Common European Framework of Reference for Languages (CEFR) to Spanish Sign Language (LSE) in a university context, with reference to the Italian situation (Council of Europe, 2020). In Spain, CEFR descriptors are already integrated into academic programmes for the assessment of LSE, whereas in Italy the context remains uneven due to the lack of shared criteria for the teaching and assessment of Italian Sign Language (LIS). The research project, conducted jointly by Ca' Foscari University of Venice and Rey Juan Carlos University of Madrid, adopts a longitudinal and comparative design focusing on the first three CEFR proficiency levels (A1, A2, B1) of LSE among L2M2 learners (second language second modality). A mixed-methods approach combining classroom observations, self-assessment instruments, and standardised assessment rubrics is used to analyse the alignment between students' self-assessments and instructors' external evaluations, with particular attention to linguistic and metacognitive awareness. The findings show increasing accuracy in self-assessment as proficiency develops, alongside recurring issues such as the overestimation of receptive skills and the underestimation of productive competence. These results highlight the need for targeted assessment interventions and contribute to the development of CEFR-consistent evaluation practices for sign languages.

Keywords: Sign Languages, CEFR, Assessment

1. Introduction

The growing attention devoted to sign languages within European academic institutions has prompted research in language pedagogy to confront new methodological and assessment-related challenges. Among these, one of the most pressing challenges concerns the application of shared and standardized instruments for teaching and evaluation, such as those proposed by the *Common European Framework of Reference for Languages* (CEFR), updated in 2020, which for the first time systematically incorporated sign languages (Council of Europe, 2020; De Monte, 2022).

The present contribution forms part of a broader comparative research project conducted under joint supervision between Ca' Foscari University of Venice and Rey Juan Carlos University of Madrid. The overall objective is to analyze and compare teaching and assessment practices in university courses of Italian Sign Language (LIS) and Spanish Sign Language (LSE), with particular attention to the assessment of linguistic competences in the first three CEFR levels (A1, A2, B1), and to the relationship between students' self-assessment and instructors' external evaluation. The findings emerging from the first phase of the study, carried out at Rey Juan Carlos

University during the 2024/2025 academic year, will be presented, and the didactic and metacognitive implications of these results will be discussed.

2. The CEFR for Sign Languages (SL) - CEFR4LS

The CEFR provides a reference framework for the description of language competences, articulated across six levels (from A1 to C2), and organizes competences not only in terms of "what one knows," but also in terms of how one acts in and through language, and with which strategies (Council of Europe, 2020). Chapter 6 of the 2020 Companion Volume (CEFR4LS), the outcome of nearly a decade of research and implementation projects, extends the principles of the CEFR to sign languages through descriptors structured around three macro-competences: Linguistic,

Sociolinguistic, and Pragmatic (Figure 1) (Haug & Keller, 2012; Leeson et al., 2016, 2019, 2020; Council of Europe, 2020; ECML, n.d., a, b).

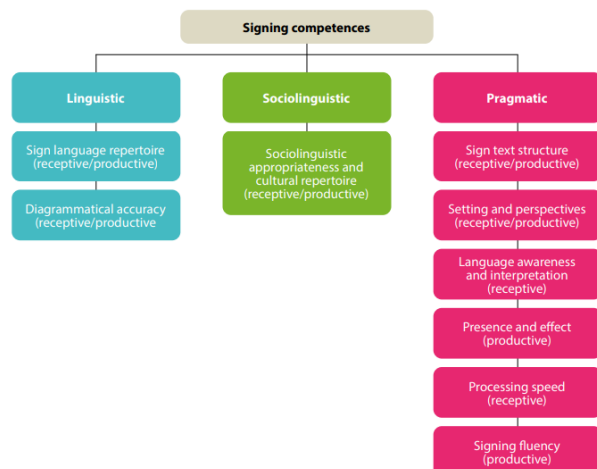


Figure 1: Signing Competences (CEFR, 2020:144)

These macro-competences are further subdivided into specific sections with the aim of ensuring coherent and functional assessment. Each is analyzed according to the four modes of communication: reception, production, interaction and mediation. In the case of sign languages, these descriptors are particularly relevant for capturing both formal and pragmatic aspects of language use, moving beyond the mere ability to produce isolated signs (Barni et al., 2024).

3. Teaching Context

The dissemination of sign language teaching across Europe currently presents a markedly heterogeneous landscape, characterized by substantial differences among national contexts. In the present study, the analysis focuses specifically on a comparison between Spain and Italy. The selection is motivated by the strong similarities between their respective sign languages and Deaf communities, which, however, have developed within distinct educational and regulatory frameworks (Marrocu, 2023). In Italy, the educational provision remains fragmented and is often independently defined by individual associations, institutions, or universities. There is no inter-university network of university lecturers capable of fostering the necessary dialogue and critical reflection on teaching practices (Branchini et al., 2024; Cardinaletti, 2018; Cardinaletti & Mantovan, 2023; Mantovan, 2021). In Spain, by contrast, the integration of the CEFR into the teaching of Spanish Sign Language (LSE) has been in place

¹ (Ortega & Özyürek, 2013; Pichler & Koulidobrova, 2015; Marschark et al., 2015; Ortega & Morgan, 2015).

for several years and is supported by official curricular instruments. Since 2010, specific syllabi have been available for instructional planning and, notably, the language received official recognition as early as 2007 (BOE, 2007; Aroca, 2012; Díez, 2014; Fundación CNSE, 2010). At present, the only comprehensive four-year undergraduate degree programme covering the full range of LSE proficiency levels (A1-C1) is offered by Rey Juan Carlos University of Madrid, thus representing a case study of particular interest for the analysis of standardization practices in sign language education (CNLSE, n.d.).

4. The Research Project

4.1 Objectives and Methodology

The study presented here forms part of a doctoral research project conducted under joint supervision between Ca' Foscari University of Venice and Rey Juan Carlos University (URJC) of Madrid. It is a longitudinal investigation whose ultimate objectives are:

- to construct a comparative framework of the teaching and assessment pathways of LIS and LSE in academic contexts;
- to analyze and compare the strategies, pedagogical approaches, and assessment methods employed, considering the guidelines provided by the CEFR (2020);
- to conduct an in-depth examination of the instruments used to assess linguistic competences, with particular attention to the initial stages of language learning.

The first phase of data collection, defined as Corpus 1, was carried out during the 2024/2025 academic year at URJC. The sample comprised three groups of hearing university students approaching a sign language for the first time (L2M2 - second language, second modality¹): 29 students at level A1 (mean age 20.1 years), 25 at level A2 (mean age 21.8 years), and 22 at level B1 (mean age 22.3 years). The teaching staff consisted of six university lecturers: four Deaf signing instructors and two hearing signing instructors, including one CODA² lecturer and one professional LSE interpreter.

The methodological approach adopted was mixed methods, combining qualitative and quantitative instruments designed to provide a triangulated perspective on the teaching and assessment process. Two questionnaires were administered: one at the beginning and one at the end of the course, comprising both closed-ended Likert

² Children Of Deaf Adults

scale items and open-ended questions (Arroyo Hernández, 2021).

At the end of the course, students were also asked to complete a linguistic self-assessment based on the descriptors of the 2020 CEFR, which are organized by competences in production and reception. To further explore the teaching perspective, individual semi-structured interviews were conducted with the instructors, with the aim of clarifying attitudes, perceptions, and practices related to teaching and assessment that had emerged from a previously administered mixed-format questionnaire.

In addition, systematic classroom observation was carried out in order to document instructional dynamics, the type and frequency of activities performed, recurring errors, and levels of student participation. The collected data were analysed using both descriptive and inferential statistical tools: quantitative responses were processed with the software packages Excel and RStudio, while analyses of variance (ANOVA) and significance tests, such as Chi-square test and correlation analyses, were conducted using JASP. Interviews and observational data were subjected to qualitative analysis aimed at identifying recurring patterns, discontinuities, and emerging perspectives.

In this contribution, particular attention is devoted to the comparative analysis between students' linguistic self-assessment and the external evaluation provided by lecturers (Björkstrand et al., 2019, Lugarini, 2010). The objective is to investigate the degree of alignment between subjective perception - based on standardized CEFR descriptors - and expert judgment grounded in the same guidelines. This relationship is further examined through methodological triangulation integrating classroom observation data, in order to verify the consistency among self-reported performance, assessed performance, and performance observed within the instructional context. Such an approach not only deepens reflection on learners' linguistic and metacognitive awareness but also makes it possible to identify any systematic discrepancies between self-perception and actual language behaviour, with significant implications for the development of more accurate assessment and pedagogical instruments for sign languages.

4.2 Data Analysis

The data analysis was conducted based on the self-assessments completed by students in the A1, A2, and B1 groups at the end of the semester. Students filled in self-assessment grids structured according to the sections and descriptors of the 2020 CEFR for sign languages (Consejo de

Europa, 2020). The scale employed ranged from 1 = "not at all" to 4 = "very much," depending on the level of ability perceived. In parallel, instructors carried out an external evaluation of students' competences using an analytic scoring grid (0 - 10), divided into the following areas: grammatical, lexical, discourse, and pragmatic competence. This framework was generally applied in the final examination assessments, which included comprehension and production tests administered twice during the semester, a group presentation delivered in the final weeks of the course, and a theoretical examination during the official exam session.

To compare the two datasets, the assessment grids were made compatible both in terms of their sections and through the alignment of scoring scales to a common 0 - 10 metric. This procedure made it possible to calculate mean scores and the differences between students' self-assessments and instructors' evaluations. It also included the computation of mean differences across the three groups, together with the verification of statistical significance through ANOVA and post hoc tests.

4.3 Results

This section presents the results emerging from the analysis of the collected data, providing an overview of the observed trends and the differences identified across the various proficiency levels, with particular attention to students' self-assessments and their comparison with lecturers' evaluations.

4.3.1 Students' Self-Assessments

The analysis of the A1 group's self-assessments, collected through grids structured according to the 2020 CEFR descriptors for sign languages, reveals a distribution of responses that highlights recurring and differentiated trends across the various sections. As illustrated in Figure 1, many students positioned themselves within the categories "quite" and "very" with a peak in Linguistic Awareness (Conc_Ling) and Sociolinguistic Appropriateness (Aprop_Socio), that is, competences related to the relational and informal dimension of communication, in which students report feeling more at ease. By contrast, more structurally oriented sections, such as Schematic Precision in both reception and production (P.Esq_REC and P.Esq_PRO), Text Structure (Texto), and Presence and Effects (Presencia), display greater dispersion in

responses, with relatively high values also in the categories “a little” and “not at all.”

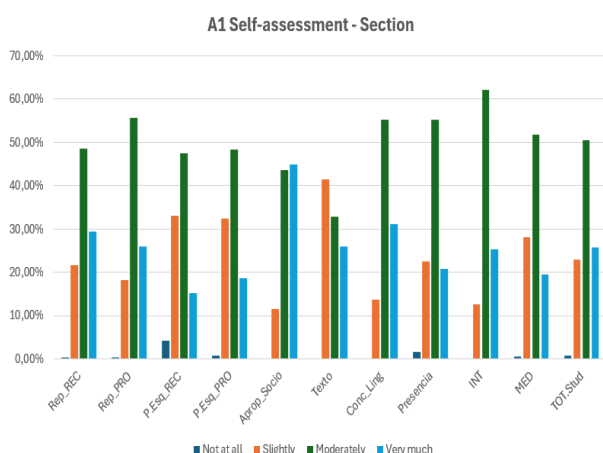


Figure 2: A1 Group Self-Assessment by CEFR section

These findings suggest an early awareness of morphosyntactic and metalinguistic difficulties, consistent with an initial level profile which, although relying on visual communicative strategies, has not yet achieved full formal control of the language.

For the A2 group, the self-assessments reveal a predominance of the category “quite” across almost all sections, indicating a perception of intermediate and functional competence that is, however, not yet fully consistent with this linguistic profile. Responses in the category “very” are less frequent but tend to cluster particularly in Sociolinguistic Appropriateness in reception (Soc_REC), Fluency (Fluidez), Presence and Effects (Presencia), and Interaction (INT), suggesting that the development of linguistic competence does not always correspond to a heightened sense of confidence in more complex communicative situations.

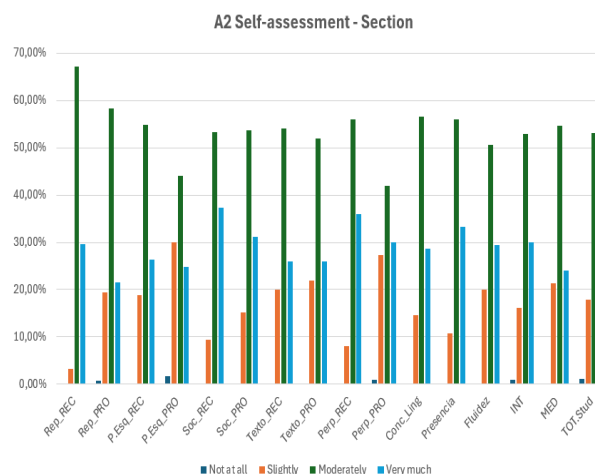


Figure 3: A2 Group Self-Assessment by CEFR section

In the B1 group, a more articulated profile emerges compared to the previous levels (Figure 3), with a less polarized distribution of responses. The category “quite” continues to prevail across many sections; however, an increase in “very” responses is observed, particularly in Schematic Precision in reception (P.Esq_REC) and Sociolinguistic Appropriateness in production (Soc_PRO), indicating greater mastery and awareness of one’s competences. Nevertheless, critical issues persist in sections related to integrated production and the discursive management of language. Indeed, responses in the categories “a little” and “not at all” increase in Sociolinguistic Appropriateness in reception (Soc_REC), Fluency (Fluidez), Presence and Effects (Presencia), and Interaction (INT), suggesting that the development of linguistic competence does not always correspond to a heightened sense of confidence in more complex communicative situations.

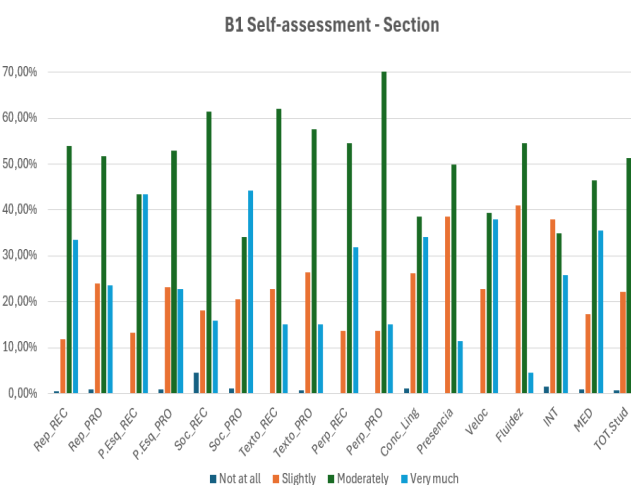


Figure 4: B1 Group Self-Assessment by CEFR section

From this perspective, self-assessment confirms its value as a metacognitive tool, useful not only for identifying strengths and areas requiring further development, but also for guiding future instructional interventions in a more targeted and informed manner.

4.3.2 Comparison between students' self-assessments and lecturers' evaluations

Comparing students' self-assessments with lecturers' evaluations makes it possible to capture competences in a multi-level and cross-cutting manner and constitutes an essential parameter for assessing the degree of alignment or potential discrepancy between what students believe they know and what they demonstrate. This comparison enabled the identification of significant divergences across the different competence areas, differentiated by proficiency level.

In the A1 group, a marked overestimation is observed in comprehension (+2,5) and, to a lesser extent, in lexical competence (+0,3). By contrast, all other competences are underestimated, most notably grammatical competence (-1,10), suggesting a critical perception of productive and formal abilities. The overall judgment value is slightly negative, pointing to a general tendency toward underestimation.

In the A2 group, although an overestimation remains in comprehension (+1,57) and lexical competence (+1,05), the discrepancy is smaller than in A1, indicating greater exposure to the language and a gradual internalization of its structures. Some competences are still underestimated, such as discourse competence (-0,52), whereas the overall evaluation approaches that of the lecturers (-0,41), indicating an improvement in self-assessment ability.

The B1 group displays a more advanced level of metalinguistic awareness: the mean overall difference is virtually null (+0,06), suggesting a strong alignment between subjective perception and objective performance. Nevertheless, a substantial overestimation emerges in grammatical competence (+3,2), likely attributable to a subjective sense of increased fluency and communicative naturalness.

A comparative analysis across the three groups further shows that five of the six competences examined present statistically significant differences between levels. The only competence

that did not yield significant differences was lexical competence. These results highlight a trajectory of linguistic development that runs in parallel with a substantial change in students' capacity for self-assessment.

4.4 Discussion of the Results

The findings outline a gradual and significant development of linguistic and metacognitive awareness across the three groups analyzed.

In the A1 group, a clear tendency toward underestimation emerges in productive and formal competences, accompanied by a marked overestimation of comprehension. This dynamic is confirmed both by the lecturers' objective evaluations and by classroom observations, which revealed effective informal communicative performances alongside persistent difficulties in the articulation of signs. The high level of engagement observed in this group suggests that initial motivation does not immediately translate into confidence in one's own competences.

In the A2 group, an overall reduction in underestimation is observed, although a certain degree of overestimation in comprehension persists. Difficulties nonetheless emerge in relation to lexical complexity and in the transition from predominantly receptive competence to more active production. In this respect, the frequent adoption of non-verbal compensatory strategies appears to facilitate communication, yet tends to mask only partial mastery of linguistic structures. Reduced participation, as documented through classroom observation, may have negatively affected both confidence consolidation and the accuracy of self-assessment.

The B1 group presents a more mature and balanced profile: the correspondence between self-assessment and external evaluation is almost complete, indicating well-developed metalinguistic awareness. However, the significant overestimation of grammatical competence reveals a specific phenomenon known as *fluency overconfidence*, whereby increased communicative fluency leads learners to perceive formal competences as fully acquired. This effect highlights the need to develop further metacognitive tools capable of enabling learners to assess accurately the formal quality of their production, even at more advanced levels (Sanchez & Dunning 2020). The comparative analysis confirms that linguistic level substantially influences the quality of self-assessment. While at initial levels oscillation between overestimation and insecurity predominates, more advanced

levels display greater stability, albeit with the risk of overestimating aspects related to structured production. Comprehension, by contrast, appears to be systematically overvalued at all levels, albeit with progressively smaller discrepancies. This represents a critical area requiring targeted metacognitive and pedagogical attention. The Dunning-Kruger effect becomes particularly evident when individual cases are examined: learners with lower competences tend to overestimate themselves in more intuitive tasks, whereas those with genuinely higher abilities may incur a general underestimation, maintaining a more realistic appraisal of their shortcomings (Carter & Dunning, 2008).

Overall, these results suggest the value of self-assessment not only as a diagnostic instrument, but also as a formative lever to promote self-regulation and self-awareness in language learning processes.

5. Limitations

Several methodological aspects warrant careful consideration when reflecting on the limitations of the present study. First, the conversion scale applied to self-assessments - transformed from the original 1-4 range to a 0-10 scale - may have introduced potential distortion effects. A further limitation concerns the small group sizes, particularly at the B1 level. This makes the interpretation of the observed effect more complex and reduces the robustness of the statistical data, thereby limiting the generalizability of the findings.

From the students' perspective, a relatively low level of engagement was observed. This is compounded by the possible influence of social desirability bias, which may have led students to orient their self-assessments toward levels perceived as more acceptable or aligned with external expectations.

Finally, the timing of the data collection should be considered. The self-assessments were administered during a period of heightened stress associated with the examination session. Such a context may have influenced students' self-perception and, consequently, the responses provided.

6. Conclusion

Overall, the study confirms the value of self-assessment as a formative practice, provided that it is supported by tools fostering linguistic reflection and anchored in a shared reference

framework. The integration of quantitative and qualitative data proved essential for capturing the complexity of language learning in L2M2 students, thereby enabling more precisely targeted instructional interventions. The evidence gathered offers practical insights for the continuation of the comparative project within the Italian context. In anticipation of data collection in Italy, it will be crucial to maintain the same methodological framework while also considering the structural differences between the two academic systems. For this reason, one of the primary objectives of the second phase will be to examine the extent to which the assessment practices adopted in Spain may be adapted or proposed as a model within the Italian context.

In this regard, the adoption of the CEFR descriptors for sign languages has proven to be an invaluable instrument. The availability of shared guidelines not only enhances the comparability of educational pathways, but also facilitates the development of a common evaluative language between students and lecturers (Serragiotto, 2017). This is particularly significant in a visual-manual domain such as sign languages, where competences must be assessed in their multimodal and pragmatic dimensions.

7. Acknowledgments

Sincere thanks are extended to the entire teaching staff and the students of Rey Juan Carlos University, and in particular to Professor Francisco Vera Villaverde. Without your support, this work would not have been possible.

8. Bibliographical References

Anna Cardinaletti. (2018). La lingua dei segni italiana a Ca' Foscari: Didattica, ricerca e progetti sull'accessibilità. In A. Cardinaletti, L. Cerasi, and P. Rigobon (Eds.), *Le lingue occidentali nei 150 anni di storia di Ca' Foscari* (pp. 341-353). Edizioni Ca' Foscari. <http://doi.org/10.30687/978-88-6969-262-8/016>

Anna Cardinaletti and Lara Mantovan. (2023). Le lingue dei segni nel volume complementare e l'insegnamento della LIS nelle università italiane. *Italiano LinguaDue*, 14(2). <https://doi.org/10.54103/2037-3597/19575>

BOE. (2007). Ley 27/2007, de 23 de octubre. <https://www.boe.es/boe/dias/2007/10/24/pdfs/A43251-43259.pdf>

- Carmen Sanchez and David Dunning. (2020). Decision fluency and overconfidence among beginners. *Decision*, 7(3), 225-237. <https://doi.org/10.1037/dec0000122>
- Chiara Branchini, Anna Cardinaletti, A. and Lara Mantovan. (Eds.). (2024). *Insegnare la LIS e la LISt all'università. Esperienze, materiali, strumenti per la didattica e per la valutazione. Edizioni Ca' Foscari*. <http://doi.org/10.30687/978-88-6969-827-9>
- CNLSE. (n.d.). *Cursos de formación LSE en España*. <https://cnlse.es/es/recursos/otros/formacion/universidades>
- Consejo de Europa. (2020). *Marco común europeo de referencia para las lenguas: Aprendizaje, enseñanza, evaluación. Volumen complementario*. Consejo de Europa.
- Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment. Companion volume*. Council of Europe.
- Deborah Chen Pichler and Helen Koulidobrova. (2015). *Acquisition of sign language as a second language (L2)*. Gallaudet University Press.
- ECML [European Center for Modern Languages of the Council of Europe] (n.d., a). *Promoting excellence in sign language instruction*. <https://www.ecml.at/Thematicareas/SignedLanguages/ProSign>
- ECML [European Center for Modern Languages of the Council of Europe] (n.d., b). *Sign languages*. <https://www.ecml.at/thematicareas/signedlanguages>
- Edoardo Lugarini (Ed.). (2010). *Valutare le competenze linguistiche (Collana Giscel, 11)*. FrancoAngeli.
- Eva Aroca. (Coord.). (2012). *Propuesta curricular de la lengua de signos española. Nivel usuario independiente B1*. Fundación CNSE.
- Fundación CNSE (Coord.). (2010). *Propuesta curricular de la lengua de signos española. Nivel usuario básico A1-A2*. Fundación CNSE.
- Gerardo Ortega and Asli Özyürek. (2013). *Gesture-sign interface in hearing non-signers' first exposure to sign*. In *Proceedings of the Tilburg Gesture Research Meeting*. Tilburg University.
- Gerardo Ortega and Gary Morgan. (2015). *Input processing at first exposure to a sign language*. *Second Language Research*, 31(4), 443-463. <https://doi.org/10.1177/0267658315576822>
- Graziano Serragiotto. (2017). *La valutazione per l'apprendimento delle lingue straniere dal punto di vista dei docenti*. *Rassegna Italiana di Linguistica Applicata*, 49(1), 7-19.
- Ignacio Arroyo Hernández. (2021). *Atteggiamenti verso la pronuncia della lingua straniera: Il questionario di Ca'Foscari (2019) nel panorama attuale di ricerca*. *Rassegna Italiana di Linguistica Applicata*, 1-2, 165-172.
- Lorraine Leeson and Beppie Van den Bogaerde. (2019). *Sign languages*. In D. Newby, F. Heyworth, and M. Cavalli (Eds.), *Changing contexts, evolving competences* (pp. 100-107). Council of Europe.
- Lorraine Leeson, Christian Rathmann, Beppie Van den Bogaerde and Tobias Haug. (2016). *Sign languages and the CEFR: Common reference level descriptors*. European Centre for Modern Languages of the Council of Europe.
- Lorraine Leeson, Jordan Fenlon, Johanna Mesch, Sarah Sheridan and Carmel Grehan. (2020). *Corpora in sign language pedagogy*. In R. Rosen (Ed.), *The Routledge handbook of sign language pedagogy* (pp. 339-352). Routledge.
- Lara Mantovan. (2021). *Insegnare la lingua dei segni italiana all'università: Esperienze consolidate e direzioni future*. *QuaderniCIRD*, 22, 33-49.
- Marc Marschark, Linda Spencer, Andreea Durkin, Georgianna Borgna, Carol Convertino, Elizabeth Machmer, William G Kronenberger and Alexandra Trani. (2015). *Understanding language, hearing status, and visual-spatial skills*. *Journal of Deaf Studies and Deaf Education*, 20(4), 310-330. <https://doi.org/10.1093/deafed/env025>
- Maria Aránzazu Díez. (Coord.). (2014). *Propuesta curricular de la lengua de signos española. Nivel usuario independiente B2*. Fundación CNSE.
- Maria Grazia Marrocu. (2023). *Espressioni idiomatiche: Confronto interlinguistico tra la LSE*

e la LIS (Tesi di laurea magistrale). Università Ca' Foscari Venezia.

Maria Tagarelli De Monte. (2022). Using the Common European Framework of Reference for Italian Sign Language education: Lessons from research and practice. *Momento – Diálogos em Educação*, 31(2), 394-416. <https://doi.org/10.14295/momento.v31i02.14502>

Monica Barni, Claudio Ferrara, Sabina Fontana and Maria Roccaforte. (2024). Certificare la competenza linguistico-comunicativa nella LIS: Riflessioni, raccomandazioni e proposte. *LEND. Lingua e nuova didattica*, 1, 8-21.

Thomas Björkstrand, Erin Balkstam and Josephine Willing. (2019). Sign language dictionary as a digital tool in L2 teaching: Sentence scoring on CEFR levels A1-B2. Department of Linguistics, Stockholm University.

Tobias Haug and Jorg Keller. (2012). ESF exploratory workshop on development of theoretical and practical guidelines for the adaptation of the CEFR to sign languages: Scientific report. European Science Foundation.

Travis J Carter and David Dunning. (2008). Faulty self-assessment: Why evaluating one's own competence is an intrinsically difficult task. *Social and Personality Psychology Compass*, 2(1), 346-360. <https://doi.org/10.1111/j.1751-9004.2007.00031.x>