

APPENDIX 1: Comparison of studies conducted in naturalistic environment

The table also briefly summarises the methods used to measure students' UA, and ways in which authors operationalise the AO.

Author	N	Brief description	Length of stay	Study focus	UA measurement method	AO definition
Piske et al., (2002)	64	Native Italian subjects living in Canada using English as L2 daily; Mean age at testing = 35, mean age overall = 48 (no standard deviations available), pure tone hearing	Minimum 18 years	Production of vowels	One at a time, repeating English and Italian words based on the recorded stimuli	Age of arrival
Patkowski (1980)	67	Mixed-nationality, highly educated, middle class, various ages	Minimum 5 years	Syntax command	Syntactic ratings of 5-minute long samples of recorded oral interviews, assigned by two judges	Age of arrival
Hyltenstam (1992)	24	Adolescent native-like active bilinguals of Swedish, 12 for whom L1 was Finnish, 12 for whom it was Spanish	Not controlled	Accentedness, lexical and grammatical accuracy	Retelling of four prepared texts (two read and two heard) plus a composition about a movie shown beforehand; no time limit; accuracy judged by the teacher	Not specified, comments implying both age of arrival and age of beginning of L2 instruction can be found in the paper
Jia and Aaronson, (2003)	10	Native Chinese immigrant children who were between 5 and 16 when they immigrated; 7 of them spoke only Mandarin, and 3 spoke Mandarin and another Chinese dialect; university-educated, varied income levels	Not controlled	Use of grammar The impact of L2 development on L1 proficiency	<input type="checkbox"/> GJT <input type="checkbox"/> L1 to L2 oral translation task	Age of arrival mixed with age of commencing education in a US school

Hammer and Dewaele (2015)	149	Polish-English immigrants with a university degree, mean age 31, (SD=4.7), majority were female (86 to 14 ratio); sequential bilinguals	Not controlled	Impact of AO on acculturation and self-perceived L2 proficiency level	Participants' own perception of native-like capability	Age of exposure to instruction
Bolibaugh and Foster (2021)	35	Polish-English immersed bilinguals residing in West London, whose AO ranged ranged between 1 and 35 years; 18 of the participants were pre-instructed (mean 2.66 years, SD=3.92)	Not controlled	Grammatical accuracy as a function of AO, phonological short-term memory and implicit statistical learning	GJT, a 40-minute-long listening task; the recording consisted of 5 practice items and 110 test sentences; participants' answers were assessed against answers of 30 adult monolingual native English speakers.	Age of arrival
Qureshi (2021)	84	University students in UAE learning English; 61 of them attended EMI instruction from primary school, while 23 started learning English at tertiary level; participants' proficiency judged based on IELTS score (average 6-band), 2 participants spoke both English and Arabic at home, 11 participants attended English language centre.	Not controlled, "early" learners were in fact EMI learners, and "late" learners were students whose immersion started when they entered university	Grammatical accuracy in terms of AO and (implicitly) type of instruction in early age	Paper-based GJT, 114 items on 12 rules and an editing task – text containing 24 errors to be corrected; reliability coefficient KR-20 = 0.89	Age of exposure to EMI

Table 1: Comparison of studies conducted in naturalistic environment

APPENDIX 2: Comparison of studies in conducted in instructed environment

This table includes comparison of empirical studies conducted in minimal-input and immersed instructed settings.

Author	N	Brief description	Length of instruction	UA measurement method	AO definition
Cenoz (2002)	564	Basque-Spanish bilinguals learning English as L2 for 564 hours (six years). Early starters were tested at 13, late starters at 16		Comprehensive test including: <ul style="list-style-type: none"> <input type="checkbox"/> Cloze test, <input type="checkbox"/> Written composition <input type="checkbox"/> Oral narration 	Early starters = 8 Late starters = 11 years old
Muñoz (2006) BAF project	1928	Catalan-Spanish bilinguals from state schools in Barcelona; low-middle class, middle class and professionals; Some subjects had more exposure due to extracurricular classes	200 hours (first test) 416 hours (second test) 726 hours (third test)	Extensive test battery including: <ul style="list-style-type: none"> <input type="checkbox"/> Dictation (in English, Catalan, and Spanish) <input type="checkbox"/> Cloze (in English, Catalan, and Spanish) <input type="checkbox"/> Listening comprehension <input type="checkbox"/> Grammar <input type="checkbox"/> Written composition <input type="checkbox"/> Oral narrative <input type="checkbox"/> Oral interview <input type="checkbox"/> Phonetic imitation <input type="checkbox"/> Phonetic discrimination <input type="checkbox"/> Role-play 	Beginning of instruction; very early = 2-6 early starters = 8 late starters = 11 very late = 14 adult = 18+
Jaekel et al. (2017)	5130	German primary school students from 31 grammar schools learning English as L2, who were tested in Year 5 and in Year 7	Early starters = 3.5 years / 5.5 years (240/549 hours) Late starters = 2 years / 4 years (140 / 444 hours)	Receptive language skills test (listening and reading); <ul style="list-style-type: none"> <input type="checkbox"/> Listening: picture recognition and sentence completion in German <input type="checkbox"/> Reading: multiple choice and open questions 	Beginning of instruction; early starters = 6-7 years (Year 1); late starters= 8-9 years (Year 3)
Pfenninger (2017)	200	Swiss students learning English, tested at 13 and at 18	Early starters = 11 years Late starters = 5 years	<ul style="list-style-type: none"> <input type="checkbox"/> Listening comprehension task, <input type="checkbox"/> Argumentative and narrative essays <input type="checkbox"/> GJT <input type="checkbox"/> Vocabulary size test <input type="checkbox"/> Productive vocabulary size test <input type="checkbox"/> Oral recount and spot-the-difference task 	Beginning of instruction; early starters = Grade 1 (German) Grade 3 (English) Grade 5 (French) late starters = 13 years (English)

Pfenninger & Singleton (2019)	636	Swiss secondary students aged 13-14 years at first collection and 18-19 at second collection learning English as L2, L3, or L4	Early starters = 5.5 year Late starters = 6 months	CEFR B1-B2 test including: <input type="checkbox"/> Listening task <input type="checkbox"/> Receptive vocabulary test <input type="checkbox"/> Productive vocabulary test <input type="checkbox"/> Written lexical richness, syntactic complexity, fluency, and accuracy <input type="checkbox"/> Oral lexical richness, syntactic complexity, fluency, and accuracy <input type="checkbox"/> GJT	Beginning of instruction; early starters = 8 years; late starters = 13 years
Baumert et al. (2020)	19857	German students from different types of federal schools learning English as L2; tested at 15-16 (Year 9)	Early starters = 638 hours Mid starters = 561 hours Late starters = 471 hours	<input type="checkbox"/> Reading and listening comprehension tests; <input type="checkbox"/> Levels of difficulty of items ranging between CEFR A1 to C1	Beginning of instruction; early starters = 6-7 years; mid starters = 8-9 years; late starters = 10 years
Pfenninger (2020)	91	4 groups of Swiss children form a private (pre) primary school receiving 50-50 German-English CLIL	Early starters = 8 years Mid starters = 6 years Late starters = 4 years	<input type="checkbox"/> Students' writing complexity, accuracy and fluency (holistic); <input type="checkbox"/> Analysis of repeated features in oral and written tasks (a recount, a narrative essay and an argumentative essay)	Beginning of instruction; early starters = 5 mid starters = 7 late starters = 9
Jaekel et al. (2022)	7289	German Primary students learning English as L2, tested in Grade 5 (10 years)	Early starters = 245 hours Late starters = 140 hours	Paper-based reading and listening multiple-choice tests targetting picture recognition and sentence completion	Beginning of instruction; early starters = 7-8 years (Grade 1) late starters = 9-10 years (Grade 3)

Table 2: Comparison of studies conducted in instructed setting