Chapter 2

Language Aptitude at Primary School (LAPS): Research design

Carina Steiner[a], Raphael Berthele[b] & Isabelle Udry[b,c]
[a]University of Bern, Center for the Study of Language and Society [b]University of Fribourg, Institut de Plurilinguisme [c]Zurich University of Teacher Education

This chapter delineates the design of the LAPS project. We start with an outline of the research questions, followed by a description of the curricular context of foreign language learning at Swiss primary schools. Next, we will give a full description of the test instruments and how they were implemented, as well as details on the participants and procedures. Finally, data entry and scoring will be outlined.

1 Research questions

The aim of the project *Language Aptitude at Primary School* (LAPS) is to explore the extent to which skills, abilities, and socio-environmental factors contribute to successful language learning by primary school children.

We consider individual difference (ID) variables and environmental factors previously found to affect foreign language learning along four broad categories:

1. Language aptitude according to Carroll (1958): Phonetic coding ability, grammatical sensitivity, inductive ability, rote memory;
2. General cognitive abilities or general learning abilities: Intelligence, working memory, creativity, field independence;
3. Affective dispositions: L2/L3 motivation, L2/L3 self-concepts, foreign language learning anxiety, dedication, perceived teacher encouragement, perceived parental encouragement, locus of control;

Carina Steiner, Raphael Berthele & Isabelle Udry. 2021. Language Aptitude at Primary School (LAPS): Research design. In Raphael Berthele & Isabelle Udry (eds.), *Individual differences in early instructed language learning: The role of language aptitude, cognition, and motivation*, 51–70. Berlin: Language Science Press. DOI: 10.5281/zenodo.5464747

4. Environmental factors: Socio-economic status (SES), language background, region, teaching paradigm (i.e. task-based teaching and learning as prescribed by the Swiss curriculum).

The theoretical underpinnings of these variables are discussed in Chapter 1. In the LAPS project, we address the following research questions:

1. What are the underlying dimensions of the ID variables assessed in the test battery and how do these dimensions relate to L2 proficiency? (Chapter 3)
2. What is the predictive value of the ID variables and environmental factors for young learners' L2 proficiency? (Chapter 4)
3. What is the relationship between the ID variables, environmental factors, and foreign language skills?
    - Association of socio-environmental background variables with foreign language ability (Chapter 5)
    - Creativity and the task-based learning environment (Chapter 6)
    - Affective predispositions and the proximity to the target language community (Chapter 7)
4. What are the developmental patterns identified over two academic years (1.5 years of foreign language learning) in terms of:
    - The dynamics of affective dispositions (Chapter 8)
    - Competences in school language German and L2 English (Chapter 9)
    - The language analysis component of language aptitude, i.e. grammatical sensitivity and inductive ability (Chapter 10)

2 Study context

The research questions were investigated in two subprojects between 2017 and 2019. LAPS I took place in a German-speaking region close to the French-German language border. The children learnt L2 French and L3 English. It was designed as a cross-sectional study which was complemented by a second data collection to further investigate affective dispositions and L3 proficiency. LAPS I also served as a pilot study for the test battery.

LAPS II was conducted in the north-eastern German-speaking part of Switzerland. Participants learnt L2 English and L3 French. It was longitudinal with three

data collections (T1, T2, T3) over two academic years. We followed the development of L2 English proficiency, L1 school language proficiency in German, aptitude (language analysis component), and affective dispositions.

2.1 The foreign language curriculum in Switzerland

All Swiss children learn two foreign languages as part of the mandatory curriculum: English and one national language, i.e. French, Italian, or Romansh. The cantons (equivalent to provinces or districts) are free to organise their own language curricula, although, based on a constitutional article accepted by the people in 2006, the last decades were characterized by a tendency to harmonize curricula across the country. Despite this tendency, cantons are still free to choose which two languages they want children to learn and in what order. Overall, this has resulted in two different systems across the German-speaking part of Switzerland where this study took place:

a. Regions close to the French-speaking area introduce French in 3rd grade and English in 5th grade. Similarly, regions close to the Italian and Romansh speaking parts of the country choose Italian and/or Romansh as one of their languages.

b. The other regions start with English as the first foreign language in 3rd grade and introduce a national language later, e.g. French, in 5th grade.[1]

As mentioned before, in the current project, both systems are represented. LAPS I took place in region a) with L2 French and L3 English, and LAPS II in region b) with L2 English and L3 French.

Each foreign language is taught for 2 to 3 lessons a week, depending on the grade. Tables 5 and 8 indicate the total number of lessons children in the LAPS project had attended at different times of testing.

2.2 Overall goals of the Swiss foreign language curriculum

Foreign language teaching in Switzerland aims at developing functional multilingualism. All four skills are taught in the L2 and L3 from the start: Listening, reading, speaking, and writing. The curriculum also contains a domain called *Sprachen im Fokus* (languages in focus) which covers formal aspects of language

[1]The region where LAPS II was conducted is an exception: Until 2019 L2 English was introduced in 2nd grade. In 2020 the region adapted to the common national practice, i.e. L2 English classes now start in 3rd grade.

(including grammar and pronunciation), language awareness, and the use of strategies. *Kulturen im Fokus* (culture in focus) stipulates objectives of cultural knowledge and attitudes. In keeping with communicative approaches to language teaching and learning, fluency is given priority over accuracy. At the end of primary school (age 12), children should reach beginner levels in the L2 and L3. More specifically, the national standards based on the CEFRL[2] target L2 levels of A2.1 in listening, reading and speaking, and A1.2 in writing. In the L3, children are expected to reach A1.2 levels in all areas of competence (Bildungsdirektion Kanton Zürich 2017: 17).

Children are taught in a task-supported approach, i.e. a weak form of communicative task-based learning and teaching (Ellis 2017). This means that tasks are central to the lesson, but they are usually complemented with form-focused elements. As described in more detail in Chapter 6, this method is reflected in the teaching manuals which in most cantons are prescribed by the local board of education. The teaching manuals contain several units structured around a topic that is introduced via authentic input followed by meaning-focused activities. Vocabulary learning and some elements of explicit grammar are also part of the lesson plans. At the end of each unit, learners use all aspects of language they have acquired to complete a communicative task about the topic, such as writing a poem, or doing role plays.

The Swiss curriculum prescribes awareness raising elements that draw on all languages in a student's repertoire (EDK 2004, Passepartout, Arbeitsgruppe Rahmenbedingungen 2008). The aim is to enhance learning under relatively limited input conditions by developing metalinguistic awareness. Therefore, teaching manuals include specific sections on intercomprehension and language learning strategies that should encourage transfer among languages. However, these sections are not the main focus of the manuals and teachers therefore integrate them flexibly into their lessons (Bildungsdirektion Kanton Zürich 2017: 8).

3 Test battery

The ID variables were assessed in a comprehensive test battery, including psychometric tests and two student questionnaires (motivation, locus of control) and a parent questionnaire (family background information). Where possible, the constructs were measured with standardized tools. However, some tests had to be translated into German and/or adapted to the age of our participants. In the following, short descriptions of each test are given. Tables 1–4 summarize dimen-

[2]Common European Framework of Reference for Languages

sions, conditions for administration, and in which subproject the tests were used. The reliability analysis is available in the technical report at https://osf.io/hstv7/.

The test battery was trialled in LAPS I with 10 classes of 4th and 5th graders from 9 different schools. Some changes were added in consultation with the scientific advisory board of the project before LAPS II (see 3.2).

3.1 Tests for language aptitude

Team up Words! Test of grammatical sensitivity based on the MLAT-E, part 2 Matching Words (Carroll & Sapon 2010), translated into German and adapted for the target group of the present study.

Participants are instructed to identify functions of words in sentences (no explicit grammatical terms are used). After the training phase, they are presented with paired sentences. In the first sentence, one word is highlighted. The participants' task is to find the corresponding word (i.e. the word with the same grammatical function) in the second sentence.

Language Detective Test for inductive abilities based on PLAB form 4 (Pimsleur et al. 2004), translated and adapted for the target group of the present study.

Participants are presented with a list of words and short sentences in an artificial language as well as their translation in German (=language of instruction). From this input, participants have to deduce how sentences in the artificial language may be formed.

Llama-E Test for sound-symbol association and phonemic working memory (Meara et al. 2005). In the training phase, participants learn how sounds are represented graphically in an artificial language. In the test phase, they listen to bi-syllabic words in the artificial language and have to choose the correct "spelling" between two options.

Llama-D Sound recognition task (Meara et al. 2005). In the training phase, participants hear different sounds of an artificial language. In the test phase, participants listen to strings of sounds (combined set of training and novel sounds) and are asked to identify the sounds they have already heard in the training.

3.2 Tests for cognition/general learning abilities

CFT 20-R: Matrices and topological deductions (Weiß 2006)*:* Fluid intelligence was assessed with two subtests from the CFT 20-R (Weiß 2006). In the matrices subtests, geometric patterns with a missing piece are presented to the participants. They are given five choices to pick from and fill in the missing piece. The topologies deductions subtest consists of geometric pattern containing one to three dots. Participants are asked to select, from five given options, the one that replicates the conditions in the example (e.g. the one where a dot could be placed outside a triangle but inside a circle).

CFT-20-R: Number sequences (Weiß 2006)*:* In this test for crystallized intelligence, participants are presented with a sequence of numbers in a certain pattern and are asked to choose the next logical number from five options.

Corsi Blocks: The Corsi Blocks task assesses visual working memory. An increasingly long sequence of squares on the screen is lit up, and participants are asked to reproduce the correct order.

Digit Span (Forward/Backward): Number strings of increasing length are presented visually and aurally. The participants have to reproduce them either in the order of presentation (forward) or in reversed order (backward) until they reach maximum recall capacity. The Digit Span task measures verbal working memory. The forward version is a simple measure of short-term phonological memory (or phonological loop), and the Backward Digit Span is a complex task tapping into executive working memory.

Alphabet Task: A measure of automatic letter access, retrieval, and production by Berninger et al. (1992) in which participants are instructed to write down the alphabet as quickly as possible without sacrificing legibility. The Alphabet task has been associated with children's levels of composition in the L1 (Berninger et al. 1997, Graham et al. 2006).

Test of Creative Thinking (Divergent Production) (TCT-DP) (Urban & Jellen 1995)*:* Participants are presented with a sheet of paper containing a frame with 6 figural fragments (a semicircle, waveline, dot, right angle, dashed line, and a lying "u" located outside the frame). They are asked to complete the unfinished picture without a time limit. The test is scored according to 14 criteria which include boundary breaking, risk taking, introducing new elements, humour, and the ability to link elements to give meaning to the overall picture. Thus, the test goes beyond scoring quantitative

aspects of divergent thinking, such as the number of ideas. The test taps more holistically into an individual's creative potential by considering qualitative aspects of divergent thinking as well, i.e. integrating various elements meaningfully or unexpected ideas.

Group Embedded Figures Test (GEFT): An assessment of field independence (Witkin et al. 2014) in which participants need to find simple geometrical figures embedded in more complex figures.

3.3 Assessment of affective dispositions

Based on existing test instruments from Horwitz et al. (1986), Stöckli (2004), Dörnyei (2010), Heinzmann (2013), and Peyer et al. (2016), we put together a student questionnaire covering the following dimensions: Intrinsic motivation, extrinsic motivation (school/leisure), lingua franca motivation, foreign language learning anxiety, self-concepts (L2 + school language), teacher motivation, parental encouragement, dedication, and future L2 self. The questionnaire comprised a section for L2 and L3 with the same items for each language.

Locus of control was assessed with a German translation of the N-S Personality Scale by Nowicki & Strickland (1973).

3.4 Assessment of environmental factors

A parent questionnaire filled in at the beginning of the study assessed personal and linguistic background (country of origin, years of schooling, L1, family language, literacy language), SES (parents' highest level of education, n° of books, financial resources, monthly income), and school context (classes in German as a second language/heritage language and culture, French/English homework).

3.5 Tests for language proficiency

ELFE 1–6, Reading Proficiency (Lenhard & Schneider 2006) is a normed test for reading skills in the language of instruction German. Items are presented at word, sentence, and text level.

Oxford Young Learners Placement Test (Oxford English Testing 2013) was used for L2/L3 English. The test consists of two sections: Language use (vocabulary and grammar) and listening (short and extended listening exercises) and is said by the distributors to cover levels A1–B1.

C-tests were used for L2 French and L3 English. Participants need to reconstruct meaning from partly deleted words in a short text, completing the missing part of the words. C-tests measure general language proficiency which is

conceptualized as an underlying ability consisting of knowledge and skills displayed in all areas of language use (Eckes & Grotjahn 2006). The C-tests were based on topics and vocabulary covered in the curriculum. They were piloted with classes who did not participate in the LAPS project. For L2/L3 English, texts were adapted from Babaii & Shahri (2010) and Porsch & Wilden (2017) in accordance with curricular content of the target group.

3.6 Piloting and adapting the test battery

After piloting the test battery in LAPS I with 10 classes of 4th and 5th graders from nine different schools, the following changes were added in consultation with the scientific advisory board. The adapted version was used in LAPS II.

3.6.1 L2 English proficiency measure

For LAPS II, the dependent variable for L2 proficiency needed to be changed from L2 French to L2 English, due to curricular differences in the regions of LAPS I and LAPS II outlined in §2.1. This change had been anticipated and tests for L2 English had been selected at the start of the project.

For L2 English proficiency, we initially planned to use the Oxford Young Learners Placement Test (OYLPT, Oxford English Testing 2013). This is an online test assessing L2 English listening comprehension and language use (vocabulary and grammar) embedded in communicative situations. The OYLPT is supposed to cover CEFRL levels A0 to B1. The test seemed appropriate for two reasons: First, items are focused on communicative aspects of language use which is in keeping with curricular goals set for the target group. Second, because our participants are expected to reach A2 levels by the end of primary school, we assumed that the OYLPT was suitable to cover the range of proficiency levels in our sample.

We were therefore surprised to find a large group of participants reaching close to the maximum score at the first time of testing (T1). To avoid ceiling effects at T2 and T3, we decided to change the English proficiency measure. In hindsight, it would have been important to pilot the OYLPT with a sample of learners comparable to our LAPS-II-learners as part of the trial phase.

C-tests were chosen as an alternative because they have been shown to be a time-efficient and reliable measure of general language proficiency (Eckes & Grotjahn 2006). Since the Swiss curriculum fosters all communicative skills, including reading and writing, participants have acquired the skills needed to cope with C-tests.

Table 1: Description of language aptitude tests

Subdimension	Test	Conditions	Study
Grammatical sensitivity	Team up Words! (adapted from MLAT-E subtest Matching Words, Carroll & Sapon 2010)	30 items, 13min 30s	LAPS I&II
Inductive ability	Language Detective (adapted from PLAB 4, Pimsleur et al. 2004)	15 items, 10min	LAPS I&II
Sound-symbol association (phonemic working memory)	Llama-E (Meara et al. 2005)	Training: 2min., 24 sounds to learn, Test: 20 bi-syllabic words	LAPS I
Sound recognition task (phonemic discrimination)	Llama-D (Meara et al. 2005)	Training: 10 words, Test: learned words alongside 20 novel words	LAPS II

Table 2: Description of tests for affective dispositions and environmental factors

Subdimension	Test	Conditions	Study
	Affective dispositions		
Motivation Foreign language learning anxiety L2/L3 self-concepts	Student questionnaire	LAPS I: T1: 40 items (L2 French), T2: 72 items (L2 French + L3 English), LAPS II: T1: 40 items (L2 English), T2–3: 72 items (L2 English + L3 French) 4-point likert scale, no time constraints	LAPS I&II
Locus of control	N-S personality scale (Nowicki & Strickland 1973)	20 yes/no items	LAPS I&II
	Environmental factors		
SES Language background	Parental questionnaire	20 items	LAPS I&II

Table 3: Description of tests for cognition/general learning abilities

Subdimension	Test	Conditions	Study
Fluid intelligence	CFT 20-R: Matrices (Weiß 2006)	15 items, 3min	LAPS II
Fluid intelligence	CFT 20-R: Topological Deductions (Conditions) (Weiß 2006)	11 items, 3min	LAPS II
Crystallised intelligence	CFT 20-R: Number Sequences (Weiß 2006)	21 items, 12min	LAPS I
Visual working memory	Corsi Blocks	LAPS I: Start with 2 squares, 2 trials per level, 1 out of 2 trials must be correct to reach next level. LAPS II: Start with 2 squares, 3 trials per level, 1/3 trials must be correct to reach next level.	LAPS I&II
Verbal working memory	Digit Span (Forward/Backward)	LAPS I: Start with 3 digits, 2 trials per level, 1/2 trials must be correct to reach next level. LAPS II: Start with 2 digits, 3 trials per level, 1/3 trials must be correct to reach next level.	LAPS I&II
Automatic letter access, retrieval, and production	Alphabet Task (Berninger et al. 1992)	Time limit: 60s, Scoring: number of legible letters in the correct alphabetic order in the first 15s	LAPS II
Creativity (divergent thinking)	Test of creative thinking (divergent production) (TCT-DP) (Urban & Jellen 1995)	Maximum score: 72 no time constraints	LAPS I
Field independence	Group embedded figures test (GEFT) (Witkin et al. 2014)	Part 1: 7 training items, 2min, not scored; Part 2: 9 test items, 5min, Part 3: 9 test items, 5min Total score: 18	LAPS I&II

Table 4: Description of language proficiency tests

Subdimension	Test	Conditions	Study
School language German reading comprehension	ELFE 1–6, Reading Proficiency (Lenhard & Schneider 2006)	72 words, 28 sentences, 20 short texts Time limits: LAPS I 4^{th}/5^{th} grade: 3/2min_{word}, 3/2$min_{sentence}$, 7/6min_{text} LAPS II reduced limits at T2 and T3 (cf. technical report)	LAPS I&II
English listening and language use	Oxford Young Learners Placement Test (Oxford English Testing 2013)	language use: 18 items Listening: 12 items, duration: no time constraints, approx. 35min	LAPS I (T2)&II (T1)
General foreign language proficiency	C-tests French	4 independent texts, 20 gaps per text. Maximum score: 80 5min per text (20min)	LAPS I
General foreign language proficiency	C-tests English	5 independent texts, 20 gaps per text. Maximum score: 100 4min per text (20min)	LAPS II (T2, T3)

We modelled the C-tests on a version developed by Porsch & Wilden (2017) for a similar target group of young learners in Germany. We also consulted C-tests for teenage learners of English by Babaii & Shahri (2010) to be able to capture higher levels of competence, i.e. to avoid ceiling effects. To make sure that the texts would be appropriate in terms of vocabulary knowledge, we consulted the English manuals used in the LAPS II region to identify content areas. We adapted the texts to include topics the children were likely to be familiar with. The C-tests were piloted with three classes (1 × 4^{th} grade, 2 × 5^{th} grades, 2 × 6^{th} grades) who did not participate in LAPS II.

3.6.2 Other subdimensions

Modifications to the test battery were also made for intelligence, phonetic coding ability, working memory, and creativity.

In LAPS I, we used a measure of crystallized intelligence (number sequencing) to account for cognitive abilities unrelated to language. This was substituted with a test of fluid intelligence (matrices). Fluid intelligence was judged to be a more accurate assessment of general learning abilities, as it is independent of academic knowledge, such as reflected in the number sequencing test. We chose two subtests from a culture fair test (CFT 20-R, Weiß 2006) with language-free and descriptive test items.

Originally, we assessed the aptitude subcomponent of phonetic coding ability with the LLAMA-E (Meara et al. 2005). The LLAMA-E is a measure of sound-symbol association and phonemic working memory. After discussions with our panel of experts, the sound-symbol aspect of this test was deemed too closely related to literacy skills, rather than the phonemic part of language aptitude which we intended to target. In order to have a more robust indication of the phonemic aptitude component, we therefore opted for the LLAMA-D subtest (Meara et al. 2005) which measures phonemic discrimination and phonemic memory.

We also decided to strengthen the working memory (WM) measure by complementing the forward digit span for verbal WM with a backward version, which some authors argue is also a measure of the central executive component (for an overview see e.g. St Clair-Thompson & Allen 2013, Hilbert et al. 2014).

The Berninger-Graham Alphabet Task (Berninger et al. 1992) was added as a speed test for automatic letter access, retrieval, and production. The Alphabet Task is easy to administer and has been found to be predictive of children's levels of composition in the L1 (Berninger et al. 1997, Graham et al. 2006). The test was chosen with regard to our aim to explore robust predictors for L2 proficiency. If the Alphabet Task turned out to be among them, it would be a convenient option for teachers wishing to assess their students' L2 potential. To our knowledge, this possibility has not been explored previously.

We tried to accommodate these changes without adding to test taking time. We therefore decided to omit the TCT-DP for creative thinking from the test battery. This choice seemed justified, as creativity in connection with the task-based learning environment did not yield strong associations with learning outcomes in LAPS I (see Chapter 6).

4 Design

This section details participant characteristics, recruitment and procedures adopted in the LAPS I and LAPS II subprojects. In terms of the total number of participants, we draw attention to the fact that Tables 5 and 8 refer to the entire sample. The numbers reported in subsequent chapters may vary, e.g. due to the exclusion of students with L1 English or French.

The test battery was administered by members of the LAPS team and/or trained research assistants. All instructions and time limits were recorded and played via speakers to have maximum control over the elicitation process, i.e. to create situations that were as similar as possible.

4.1 Recruitment

Participants for LAPS I and LAPS II were recruited via school administrators. Teachers decided voluntarily if they wanted to participate with their class. Written consent was obtained from the pupils' parents.

4.2 LAPS I

In spring 2017, the test battery presented in Tables 2.1 to 2.4 was administered to 174 primary school pupils from 10 different classes from nine different schools (T1). In spring 2018, a second data collection (T2) took place in order to investigate students' L3 (English) proficiency as well as French and English learning motivation (see Berthele & Udry 2019 for an analysis of the skills in both FLs). Nine out of 10 classes from T1 participated in the second data collection. Table 5 presents an overview of the participants of LAPS I.

Note that the variable for multilingualism is binary and based on the parent questionnaire administered at T1. Children being classified as multilingual met at least one of the following criteria:

- L1 other than German
- Family language currently used: Other than German or German plus other
- Literacy skills in a home language other than German
- Child has received or is receiving tuition for non-native German speakers (referred to in the Swiss system as *Deutsch als Zweitsprache, DaZ*)
- Child has received or is receiving tuition in the home language (referred to in Switzerland as *Heimatliche Sprache und Kultur, HSK*)

The number of L2 and L3 lessons is an estimated average of the total tuition received at the time of testing. The estimate is based on an evaluation of the cantons' current timetables which was mandated by the government and conducted by Bucher & Zemp (2019). According to the authors, the L2 is taught to Swiss primary school children with a total of 390 45-minute lessons. These lessons are distributed over 4 academic years (comprising 39 weeks each), usually with three weekly lessons in grades 3 and 4, and two weekly lessons in grades 5 and 6.

Depending on the canton, the number of L3 lessons taught at primary school ranges from 152 to 234.[3] The total varies because of different timetables in grades 5 and 6, the period of L3 instruction: Some cantons opt for three weekly lessons per grade, while others offer only two weekly lessons per grade.

Table 5: Participants LAPS I

Time	Grade	*N*	Multiling.	Age		Mean *n* lessons	
				Mean	Range	L2 French	L3 English
T1	4	57	11	10.3	9.4–11.0	172	0
T2	5	55	10	11.3	10.5–11.6	248	58
T1	5	117	27	11.5	10.6–13.2	248	58
T2	6	103	23	12.6	10.5–14.2	324	134
T1		174	38	11.1	9.4–13.2		
T2		158	33	12.1	10.5–14.2		

4.3 LAPS I procedure T1

Data collection took place between March and April 2017. Tests were administered by two or three assistants or LAPS-researchers, depending on the number of students per class. The administration of the entire test battery took approximately 3 hours and was divided into two sessions in order to prevent fatigue scheduled within a week (see Table 6). The test sequence was organized to allow for alternation between more and less cognitively demanding tasks. For practical reasons the order of the tasks could not be varied. It is possible that there are order effects of the tasks. However, if so, they arguably affect the participants in a similar way. Since we are interested in differences among the students, we deem the impact of such order effects on the overall results to be modest.

[3] In 2019, Appenzell Innerrhoden was the only canton to introduce L3 teaching at secondary school.

Table 6: Procedure LAPS I, T1

Session 1 (≈100min)	Slot 1	Introduction; Language Detective; GEFT
	Slot 2	Team up Words!; ELFE 1–6
Session 2 (≈90min)	Slot 3	C-test French; TCT-DP; CFT 20-R (numerical sequences)
	Slot 4	LLAMA-E; Corsi Blocks; Digit Span; Motivation Questionnaire

4.4 LAPS I procedure T2

Data collection took place a year later in April and May 2018. We administered measures of L3 English proficiency and motivation in one session that lasted approximately 75 minutes. This session was conducted by two assistants or LAPS researchers, applying standardized instructions and procedures. Language tests and questionnaires were alternated (see Table 7).

Table 7: Procedure LAPS I T2 – Spring 2018

Test session[a]	Introduction; Questionnaire L2 English; Oxford Young Learners Placement Test; Questionnaire L3 French

[a](Short breaks between tasks)

4.5 LAPS II

After LAPS I, minor changes discussed in section 3 were made to the test battery. In autumn 2017 (T1), the adapted version was administered to primary school pupils from 32 different classes (13 4^{th} graders, 15 5^{th} graders and 4 mixed grade classes). As in LAPS I, instructions and procedures were standardized and introduced to research assistants in a training session.

In spring 2018 (T2) and 2019 (T3) five measures were re-administered to the same participants to monitor longitudinal development: L2 English proficiency, English/French motivation questionnaire, language of instruction German, grammatical sensitivity, and inductive ability. A total of 637 pupils participated in LAPS II. Table 8 provides participant details. Note that the same criteria apply to the variable *multilingual* as in LAPS I.

Table 8: Participants LAPS II. T1: Autumn 2017; T2: Spring 2018; T3: Spring 2019

Time	Grade	*N*	Multiling.	Age		Mean *n* lessons	
				Mean	Range	L3 French	L2 English
4th graders at T1							
T1	4	289	171	10.0	9.1–11.6		181
T2	4	274	161	10.6	9.7–12.2		259
T3	5	260	150	11.6	10.7–13.2	66	346
5th graders at T1							
T1	5	326	157	11.0	9.2–12.4	14	294
T2	5	304	143	11.6	9.8–13.0	66	346
T3	6	306	146	12.6	10.8–14.0	146	426
Total participants							
T1	615	328	10.5	9.1–12.4			
T2	578	304	11.1	9.7–13.0			
T3	566	296	12.1	10.7–14.0			

4.6 LAPS II procedure T1

The administration of the entire test battery took approximately 3.5 hours. Similar to LAPS I, it was divided into two sessions (see Table 9). Nine assistants were recruited and trained for test administration.

4.7 LAPS II procedure T2 and T3

At T2 and T3, selected measures from the test battery were repeated to longitudinally track their development: L2 English proficiency, language of instruction German, motivation, grammatical sensitivity, and inductive ability. They were administered in one session of approximately 90 minutes divided into two slots (Table 10). As only paper and pencil group tests were administered at T2 and T3, fewer assistants were needed for data collection. Four research assistants at T2 and two research assistants at T3 were recruited and trained analogous to T1.

Table 9: Procedure LAPS II T1: Autumn 2017

Session 1 (≈120min)	Slot 1	Introduction; Language Detective; GEFT
	Slot 2	ELFE 1–6; CFT 20-R
	Slot 3	Alphabet Task; Team up Words!
Session 2 (≈90min)	Group A 2 A	LLAMA-D; Corsi blocks; Digit Span (f/b); Motivation Questionnaire English
	Group B 1A+ teacher	Oxford Young Learners Placement Test (OYLPT); Motivation Questionnaire English

Table 10: Procedure LAPS II T2 and T3: Spring 2018 & 2019

Test session 1–2TL	Slot 1	Introduction; Language Detective Team up Words!; English Motivation Questionnaire
	Slot 2	ELFE 1–6; English Proficiency (C-test); French Motivation Questionnaire

5 Scoring and data entry

In LAPS I and LAPS II, computer-administered tests were scored automatically. Paper and pencil tasks were scored by members of the LAPS team and assistants who participated in the data collection. Subsequently, all data were entered and stored for analysis. To minimise the chances of mishaps in data entry, a special platform was created restricting the data format of the input and displaying error messages when impossible data was entered (cf. Vanhove 2018).

Test scoring and data entry procedures were defined in a manual and the research assistants were trained accordingly. The most important aspects of the process will be summarised in the following paragraphs. Full details on handling missing values and scoring the tests are given in the technical report (https://osf.io/d9gnh/).

For the English proficiency tests administered at T2 and T3, two different scoring types were applied:

- Spelling errors are penalized: This constitutes the standard scoring type.
- Spelling errors are ignored: This accounts for the communicative language learning approach practised in Swiss EFL-classrooms which doesn't focus strongly on spelling.

Due to the acceptance of phonetically correct spellings, the second scoring type opens the door to variable criteria as to what represents an acceptable response and what not. When trialling the C-tests, lists of accepted and unaccepted spellings were defined. At T2 and T3, a randomly selected subset was scored by two independent raters and the guidelines for spelling variants were supplemented accordingly.

A subsequent analysis reveals that the two scoring types are strongly correlated ($r_{T2} = 0.98, r_{T3} = 0.98$, cf. technical report, Figure 16.1 and Figure 17.1). For any statistical analysis we used the first scoring type, i.e. where spelling errors are penalized.

References

Babaii, Esmat & Somaiyeh Shahri. 2010. Psychometric rivalry: The c-test and the cloze test interacting with test takers' characteristics. In Rüdiger Grotjahn (ed.), *Der C-Test: Beiträge aus der aktuellen Forschung/ The C-Test: Contributions from Current Research*, 41–56. Frankfurt am Main: Lang.

Berninger, Virginia W., Katherine B. Vaughan, Robert D. Abbott, Sylvia P. Abbott, Laura Woodruff Rogan, Allison Brooks, Elizabeth Reed & Steve Graham. 1997. Treatment of handwriting problems in beginning writers: Transfer from handwriting to composition. *Journal of Educational Psychology* 89(4). 652–666.

Berninger, Virginia W., Cheryl Yates, Ana Cartwright, Judith Rutberg, Elizabeth Remy & Robert Abbott. 1992. Lower-level developmental skills in beginning writing. *Reading and Writing* 4(3). 257–280. DOI: 10.1007/BF01027151.

Berthele, Raphael & Isabelle Udry. 2019. Multilingual boost vs. cognitive abilities: Testing two theories of multilingual language learning in a primary school context. *International Journal of Multilingualism*. 1–20. DOI: 10.1080/14790718.2019.1632315.

Bildungsdirektion Kanton Zürich. 2017. *Lehrplan 21: Sprachen*. Zürich: Lehrmittelverlag Zürich.

Bucher, Monika & Benedict Zemp. 2019. *Stundentafeln zum Lehrplan 21: Auswertung 2017/2018/2019*. Luzern: BKZ Geschäftsstelle Luzern.

Carroll, John B. 1958. A factor analysis of two foreign language aptitude batteries. *The Journal of General Psychology* 58(1). 3–19. DOI: 10.1080/00221309.1958.9710168.

Carroll, John B. & Stanley M. Sapon. 2010. *Modern language aptitude test: Elementary manual*. Rockville: Second Language Testing Foundation.

Dörnyei, Zoltán. 2010. *Questionnaires in second language research*. 2nd edn. New York: Routledge.

Eckes, Thomas & Rüdiger Grotjahn. 2006. A closer look at the construct validity of c-tests. *Language Testing* 23(3). 290–325. DOI: 10.1191/0265532206lt330oa.

EDK. 2004. *Sprachenunterricht in der obligatorischen Schule: Strategie der EDK und Arbeitsplan für die gesamtschweizerische Koordination*.

Ellis, Rod. 2017. Task-based language teaching. In Shawn Loewen & Masatoshi Sato (eds.), *The Routledge handbook of instructed second language acquisition*, 108–125. New York: Routledge.

Graham, Steve, Miriam Struck, Julie Santoro & Virginia W. Berninger. 2006. Dimensions of good and poor handwriting legibility in first and second graders: Motor programs, visual–spatial arrangement, and letter formation parameter setting. *Developmental Neuropsychology* 29(1). 43–60. DOI: 10.1207/s15326942dn2901_4.

Heinzmann, Sybille. 2013. *Young language learners' motivation and attitudes: Longitudinal, comparative and explanatory perspectives*. London: Bloomsbury.

Hilbert, Sven, Tristan T. Nakagawa, Patricia Puci, Alexandra Zech & Markus Bühner. 2014. The digit span backwards task. *European Journal of Psychological Assessment* 31(3). 174–180. DOI: 10.1027/1015-5759/a000223.

Horwitz, Elaine, Michael B. Horwitz & Joann Cope. 1986. Foreign language classroom anxiety. *The Modern Language Journal* 70(2). 125–132.

Lenhard, Wolfgang & Wolfgang Schneider. 2006. *ELFE 1–6: Ein Leseverständnistest für Erst- bis Sechstklässler*. Göttingen: Hogrefe.

Meara, Paul M., James Milton & Nuria Lorenzo-Dus. 2005. *Llama language aptitude tests: The manual*. Swansea.

Nowicki, Stephen & Bonnie R. Strickland. 1973. A locus of control scale for children. *Journal of Consulting and Clinical Psychology* 40(1). 148–154.

Oxford English Testing. 2013. *Oxford young learners placement test*. https://www.oxfordenglishtesting.com/ (5 May, 2017).

Passepartout, Arbeitsgruppe Rahmenbedingungen. 2008. *Passepartout: Didaktische Grundsätze*. http://www.passepartout-sprachen.ch/informationen-fuer/lehrpersonen-primarstufe/worum-geht-es/ (5 July, 2017).

Peyer, Elisabeth, Mirjam Andexlinger & Karolina Kofler. 2016. *Projekt Fremdsprachenevaluation BKZ: Schlussbericht zu den Befragungen der Schülerinnen und Schüler*. Freiburg: Institut für Mehrsprachigkeit, Universität Freiburg–PH Freiburg.

Pimsleur, Paul, Daniel J. Reed & Charles W. Stansfield. 2004. *Pimsleur language aptitude battery PLAB*. Rockville: Second Language Testing.

Porsch, Raphaela & Eva Wilden. 2017. The development of a curriculum-based c-test for young EFL learners. In J. Enver & E. Lindgren (eds.), *Researching the complexity of early language learning in instructed context* (Multilingual Matters), 289–304. Clevedon: Multilingual Matters.

St Clair-Thompson, Helen L. & Richard J. Allen. 2013. Are forward and backward recall the same? A dual-task study of digit recall. *Memory & Cognition* 41(4). 519–532. DOI: 10.3758/s13421-012-0277-2.

Stöckli, Georg. 2004. *Motivation im Fremdsprachenunterricht: Eine theoriegeleitete empirische Untersuchung in 5. und 6. Primarschulklassen mit Unterricht in Englisch und Französisch* (Pädagogik bei Sauerländer). Oberentfelden: Sauerländer Verlag.

Urban, Klaus K. & Hans G. Jellen. 1995. *Test zum Schöpferischen Denken – Zeichnerisch*. Frankfurt am Main: Pearson PsychCorp.

Vanhove, Jan. 2018. *A data entry form with sanity checks*. https://janhove.github.io/design/2018/07/06/data-entry-failsafes (31 July, 2020).

Weiß, Rudolf H. 2006. *Manual Grundintelligenztest Skala 2 – Revision (CFT 20-R)*. Göttingen: Hogrefe.

Witkin, Herman A., Philip K. Oltman, Evelyn Raskin, Stephen A. Karp & Jack Demick. 2014. *Group embedded figures test manual*. Menlo Park: Mind Garden.