INQUIRIES IN LANGUAGE LEARNING
Forschungen zu Psycholinguistik und Fremdsprachendidaktik
Edited by / Herausgegeben von Christiane Bongartz / Jutta Rymarczyk
30

Stefanie Frisch / Jutta Rymarczyk (eds.)

Current Research into Young Foreign Language Learners' Literacy Skills

PETER LANG

The role of teaching reading, spelling, and writing in two languages at primary school has become more important over the last few years. Current research therefore needs to answer more specific questions than ever before and, e.g., explore the challenges of developing literacy skills in a foreign language, appropriate teaching approaches and innovative assessment procedures. This volume features contributions on theoretical and methodological aspects, teachers' diagnostic skills and issues in educational policy. Researchers from Germany, Great Britain, Portugal, and Switzerland share their findings with the objective of improving foreign language teacher education at university level and primary school classrooms of English, French, and Spanish as a foreign language.

Stefanie Frisch is Professor of English as a foreign language at the University of Wuppertal. Her research interests focus on teaching young learners, including bilingual education.

Jutta Rymarczyk is Professor of English as a foreign language at Heidelberg University of Education. Her current interests include research into early biliteracy.

www.peterlang.com

Current Research into Young Foreign Language Learners‘ Literacy Skills

INQUIRIES IN LANGUAGE LEARNING

FORSCHUNGEN ZU PSYCHOLINGUISTIK UND FREMDSPRACHENDIDAKTIK

Edited by
Christiane Bongartz and Jutta Rymarczyk

VOLUME 30

Zu Qualitätssicherung und Peer Review der vorliegenden Publikation

Die Qualität der in dieser Reihe erscheinenden Arbeiten wird vor der Publikation durch eine Herausgeberin der Reihe geprüft.

Notes on the quality assurance and peer review of this publication

Prior to publication, the quality of the work published in this series is reviewed by one of the editors of the series.

Stefanie Frisch / Jutta Rymarczyk (eds.)

Current Research into Young Foreign Language Learners' Literacy Skills

PETER LANG

Bibliographic Information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available online at http://dnb.d-nb.de.

Library of Congress Cataloging-in-Publication Data
A CIP catalog record for this book has been applied for at the Library of Congress.

The publication of this volume has been funded by Heidelberg University of Education and by the Ministry of Science, Research and the Arts in Baden-Wuerttemberg.

ISSN 1868-386X
ISBN 978-3-631-81711-7 (Print)
E-ISBN 978-3-631-86817-1 (E-PDF)
E-ISBN 978-3-631-86818-8 (EPUB)
DOI 10.3726/b19101

This publication has been peer reviewed.

www.peterlang.com

Preface

Language Learning is a field which bridges the gap between the research conducted within Psycholinguistics and the applied research within Foreign Language Didactics. For a long time, these two fields were regarded as separate disciplines, and the emphasis lay on their differences. However, just as there has been a gradual convergence between the concepts of *language acquisition* and *language learning*, over the past few years Psycholinguistics and Foreign Language Didactics have also been moving closer together. While Psycholinguistics is taking a growing interest in the classroom context in which language learning takes place, Foreign Language Didactics have fully embraced empirical research which sheds light on the linguistic phenomena found in the interactions within the classroom.

The series *Inquiries in Language Learning (Forschungen zu Psycholinguistik und Fremdsprachendidaktik)* aims to reflect this development. Since the areas of intersection between these two research fields have a high level of interdisciplinarity, the contributions to this series are relevant in many different ways for educators and researchers who are concerned with language learning. On the one hand, good foreign language or second language teaching requires teachers whose methodological and pedagogical decisions are based on a sound knowledge of language acquisition theory. Furthermore, foreign language textbooks should have a solid empirical foundation. On the other hand, the interpretation of linguistic data requires familiarity with the types of classroom activities and rituals that shape the various learning processes. After all, psycholinguistic research design must attend to the technicalities of classroom teaching and learning in order to obtain authentic results.

In this series we hope to contribute to the cross-disciplinary efforts in our research fields, bringing together psycholinguistic principles and classroom-based developments, thus reconciling theories and methods with research and practice.

Christiane Bongartz
Jutta Rymarczyk

Vorwort

Sprachenlernen/ *Language Learning* ist das Bindeglied, das die naturwissenschaftliche Forschung der Psycholinguistik und die anwendungsorientierte Forschung der Fremdsprachendidaktik zusammenführt. Lange Zeit wurden die Disziplinen getrennt voneinander behandelt und die Betonung lag auf den disparaten Anteilen der beiden Gebiete. Vergleichbar zur Annäherung der Begriffe „Spracherwerb" und „Sprachenlernen" (*language acquisition* und *language learning*) ist jedoch seit einigen Jahren eine Annäherung der Psycholinguistik und der Fremdsprachendidaktik zu beobachten. Während die Psycholinguistik den schulischen Kontext des Spracherwerbs stärker beachtet, ist aus der Fremdsprachendidaktik die empirische Forschung nicht mehr wegzudenken, die linguistische Phänomene der Interaktion im Klassenzimmer beleuchtet.

Mit der Reihe „*Inquiries in Language Learning*. Forschungen zu Psycholinguistik und Fremdsprachendidaktik" wollen wir dieser Entwicklung Rechnung tragen. Da die Schnittstelle der beiden Forschungsgebiete, die durch die Reihe bedient wird, naturgemäß eine hohe Interdisziplinarität aufweist, strahlt ihre Relevanz in unterschiedliche Richtungen aus: Einerseits braucht guter Fremdsprachenunterricht Lehrkräfte, deren methodisch-didaktische Entscheidungen auf detaillierter Kenntnis spracherwerbstheoretischer Aspekte beruhen. Das Schreiben von Lehrbüchern für den Fremdsprachenunterricht muss auf einer soliden empirischen Basis geschehen. Andererseits bedarf die Interpretation psycholinguistischer Daten der Vertrautheit mit Unterrichtsabläufen und den Ritualen, die Vermittlungsprozesse prägen. Das Entwerfen eines psycholinguistischen Forschungsdesigns muss unterrichtstechnische Aspekte einbeziehen, um letztlich authentische Ergebnisse abbilden zu können.

Mit der Gesamtschau unserer Arbeitsbereiche hoffen wir dem Ineinandergreifen und den Verschränkungen von psycholinguistischen Grundlagen und fachdidaktischen Weiterentwicklungen, von Theorien und Methoden sowie von Forschung und Praxis gerecht werden zu können.

Christiane Bongartz
Jutta Rymarczyk

Table of Contents

This work contains additional information as an attachment. The appendix can be downloaded from our website. Please go to the page 137 of this publication for the activation code.

Stefanie Frisch and Jutta Rymarczyk

Introduction

Most of the chapters in this edited volume started as papers presented in two strands of the third international conference on Early Language Learning, hosted in 2018 by the Vigdís Finnbogadóttir Institute of Foreign Languages and the School of Education at the University of Iceland in Reykjavik, Iceland, in association with the AILA Research Network on Early Language Learning. One of these two strands, organised by Stefanie Frisch, dealt with reading and writing in the primary EFL (English as a Foreign Language) class in general, while the other strand, coordinated by Jutta Rymarczyk, explored questions related to assessing learners' reading and writing skills. The idea of bringing together the participants in a joint publication had the aim not only of discussing issues surrounding multilingualism and young language learners but actually also of pointing out the advances made in the course of this discussion. Clear foci have emerged in the discussion of early language learning, with early biliteracy as probably one of the most important and most evolved.

About 10 years ago it was being questioned whether the phrase "momentous shift", used by Diehr and Rymarczyk (2012) in order to describe the then current developments in teaching English as a Foreign Language in German primary schools, had been justified. In their use of the phrase "momentous shift", the two authors were referring to the shift from playful activities, which had been meant to allow young learners a first encounter with foreign languages, towards more cognitive yet age-appropriate approaches to foreign language learning (FLL). These approaches included, among other elements, early reading, spelling, and writing, as well as the fostering of language awareness. While at the time there were quite a few skeptical voices, the situation today seems to be a very different one.

The extension of the Common European Framework of Reference (CEFR) to integrate the level Pre-A1 / A1, i.e. descriptors for the primary age group (ages 7–10), might be seen as a milestone in the general acceptance of a more exacting approach to young learners' foreign language classes. In 2018, the CEFR was complemented with collations of new

descriptors for young learners which also comprised the skills "Written Reception", "Written Interaction", and "Written Production" (Szabo 2018). In other words, reading and writing in a systematic framework has become fully accepted as an important element of primary foreign language teaching.

The main reason for the general adoption of literacy in foreign language classes at primary level is probably its role in the transition from primary to secondary school level. A survey on how practitioners, researchers, teacher educators, and school administrators involved in early foreign language learning view a range of such issues as the transition from primary to secondary school, early literacy, assessment etc., revealed broad acceptance (more than 75 % of the participants) of a high aspiration level in early writing. This strong showing was accompanied by the desire to be allowed to grade the students' free writing at the end of primary school. It was argued that while there is undoubtedly the primacy of oracy, an exclusive focus on oral skills would result in too wide a gap between the two school levels (Hempel, Kötter & Rymarczyk 2017: 46).

At the same time, the part of the scientific community that held a secondary school perspective on early foreign language learning claimed that students at the beginning of Grade 5 showed a lack of accuracy and literacy skills. In fact, secondary school practitioners had obviously started to assign a diminished role to oracy at the very moment early reading and writing had been given greater attention (Kolb & Legutke 2016: 10). However, it is important to realize that the young learners' achievements (e.g. listening skills, vocabulary knowledge) should determine the evaluation of primary school foreign language classes, and that primary level methodology should integrate the skills – with an on-going acceptance of the primacy of oracy for primary level. There needs to be a mutual understanding between primary and secondary school teachers of the characteristic didactical and methodological issues at the respective school forms. As research could prove, collaboration between the two groups of teachers can lead to deeper insights into the respective learning processes as far as oracy but also early literacy is concerned which is beneficial for the beginning learners (cf. the project PEAK1, Kolb & Legutke 2019). Consequently, it does not come as a surprise that young researchers currently in the field claim that "[w]ithout contesting the primacy of oral

skills, the question is no longer if reading and writing should be included in early foreign language learning, but how" (Reckermann 2021).

The fact that early FLL is accorded little importance at a political level holds true not only for Germany; in his survey report "Language teaching in primary and secondary schools in England" for 2020, Collen remarks as one of his "headline findings for 2020": "[P]rimary languages are embedded in policy, but not in practice" (Collen 2020: 3).

Indeed, there are some uncomfortable parallels between England and Germany as far as early foreign language classes are concerned: "[…] almost 40 % of schools state that, in practice, pupils do not always receive language teaching according to the time allocated each week throughout the year" (Collen 2020: 6). In numerous German schools it is foreign language classes which are rededicated as soon as some organizational class matters have to be discussed. It also seems to be a standard procedure, in both England and Germany, to withdraw students from the foreign language classroom in order to support them in the language of schooling, especially in the respective reading and spelling skills, or in Mathematics (ibid., p. 12); (Rymarczyk, oral communication). These approaches, however, should be viewed in a very critical light, as the relegation / demotion of foreign language classes aggravates general social inequity. Foreign language learning is, after all, considered a basic skill in today's society, allowing as it does for participation, equal opportunities and empowerment for interaction in a globalized world (cf. Rymarczyk 2021). As early as 2002, the European Council emphasized the importance of improving "the mastery of basic skills, in particular by teaching at least two foreign languages from a very early age" (European Council 2019). And today it is stressed that "literacy competence and multilingual competence are defined among the eight key competences in the Council Recommendation on key competences for lifelong learning" (ibid.).

With these language policy issues in mind, the situation primary schools face in terms of untrained foreign language teachers seems inexplicable. Yet, again, in both Germany and England the picture is deplorable. In Germany today at least half of the English teachers at primary level are not trained for foreign language teaching (the BIG-Study reported 50 % in Börner et al. 2016). Their competence profile as outlined by Rymarczyk (2018: 160) shows shortcomings in several areas:

- no professionally developed foreign language knowledge
- no foreign language skills which have been developed with regard to teaching this foreign language
- no subject matter content knowledge (Linguistics, Literature, Culture)
- pedagogical knowledge and skills but no pedagogical content knowledge

One question that demands address is whether the existing pedagogical knowledge and skills are sufficient to meet the requirements of a professional foreign language classroom. Even if the teachers' dispositions allow them to add competency components like affect and motivation, compensation for deficits in competence cannot be expected, as psychologists hold the competence dimensions to be of a multiplicative rather than an additive character (ibid.).

Another issue which needs to be considered in this connection is that most teachers deployed in the primary foreign language classroom without appropriate training are not in their situation out of choice. While complying with the schools' needs, they show distanced attitudes towards the roles imposed upon them. The situation in England elucidates this point: "Where the highest language qualification of school staff is a GCSE, the figure for those who have not availed of subject specific CPD [Continuing Professional Development, S.F./J.R.] increases to 86 %" (Collen 2020: 7).

This attitude of reluctance towards one's teaching and the deficits untrained teachers cannot easily redress is manifested in the results of the educational monitoring *IQB-Bildungstrend 2015*. It was shown that the achievements of students taught by untrained teachers were lower than those of students whose teachers possessed a teaching degree in English as a foreign language (Hoffmann & Richter 2016: 504). These results coincide with former insights about learner perception and teacher cognition. They are known as the factors that govern classrooms processes and practices (Kumaravadivelu 2006: 165).

This volume brings together current research on reading, spelling, and writing from England, Germany, Portugal and Switzerland. In all the texts of this volume it can be seen that reading, spelling, and writing in a foreign language place specific challenges on learners. In their first language(s) children have already learnt that strings of sounds (words) can be translated

into written graphemes and conversely that strings of graphemes can be decoded into words. Foreign language learners rely on their knowledge in their first language(s) when they start reading, spelling, and writing in the foreign language. On the one hand, this can be a support because French, German, Portuguese, Spanish, and English all have alphabetic writing systems. On the other hand, it is known that each orthography follows specific rules which differ from one alphabetic orthography to another. Concurrently with grapho-phonological knowledge, learners also need to develop vocabulary, grammatical, and discourse knowledge in order to read and write short sentences and simple texts.

The challenges outlined above constitute the matter for consideration throughout the whole of this publication. The chapters of the volume are divided into two parts: I – "Assessing Reading, Spelling, and Writing" and II – "Fostering Reading, Spelling, and Writing".

Part I: Assessing reading, spelling, and writing

Assessment is one of the key activities of foreign language teachers. Jennifer Hammond, e.g., argues that "on-going diagnostic assessment of language and literacy strengths and weaknesses [is important, S.F./J.R.] so that individual learning programmes can be developed to target identified needs and facilitate accelerated rates of learning which are appropriate for age, academic needs and interests of children" (2011: 38).

University students aiming to become teachers acquire few insights into this topic, and opportunities for developing diagnostic competences remain scarce. It is therefore not suprising that unexperienced teachers feel burdened as they are forced into dilettantism (cf. Diehr/ Frisch 2008: 19).

This volume contains five chapters which provide current insights from research projects into this important topic.

Jutta Rymarczyk's paper "*Teachers' diagnostic skills in feedback on German primary school students' first attempts to spell in English*" continues the discussion above on the need to train teachers thoroughly to enable them to teach responsibly. She focuses on revealing evidence that the invented spelling of EFL young beginning writers which is based on phonographic interferences from German is not yet taken sufficiently into

account in most teachers' evaluations. Regarding these mistakes as an asset rather than a deficit can inform teachers about their students' phonological awareness and grapheme-phoneme principles already acquired. Thus, developmental aspects in their young students' literacy acquisition can be diagnosed in a much more comprehensive way, distinguishing advanced errors from possible indications of spelling weakness or even dyslexia.

The chapter "*Developing and assessing reading comprehension in primary learners of English as a Foreign Language*" by Stefanie Frisch, Carsten Breul, Bärbel Diehr, Claudia Kastens, and Annette Becker, is informed by the EULE study which is part of a research initiative at the University of Wuppertal (Germany) to improve initial teacher education at university by developing innovative seminars in which students are required to link their knowledge acquired in different classes (Linguistics, TEFL, and Educational Science) for solving specific problems (in this case assessing young EFL learners' reading skills). The authors present a reading comprehension test which is based on a theoretical model of reading. The test contains items related to a specific picture book. The test items can be used as a model for developing new tests based on other reading texts.

In their study "*Assessing young language learners' receptive skills: Should we ask the questions in the language of schooling?*" Malgorzata Barras, Katharina Karges, and Peter Lenz shed light on the question whether test validity is reduced if the test questions and/or the test instructions are formulated in the foreign language. The study was conducted in the Swiss context with primary school learners (grade 6) learning French as a Foreign Language. The authors identify the risk that questions formulated in the foreign language might be partly or entirely misunderstood by learners, which then leads to a limitation of a valid interpretation of the learners' test results.

The chapter "*Conceptualising and measuring writing in English as a Foreign Language at primary school*" by Ruth Trüb and Stefan D. Keller is based on the authors' research "An Empirical Study of EFL Writing at Primary School" conducted in the Canton of Aargau, Switzerland. The authors describe the challenges of learning to write in English as a foreign language. In their research project, scales for assessing writing are developed and their suitability is evaluated with the help of two writing tasks for learners in year 6. In their findings, the authors highlight the need

for integrating assessment criteria with a focus on meaning and combine formal and communicative scales.

Sandie Mourão, Maria Alfredo Moreira, Carolyn Leslie, and Estela Monteiro tackle the debate over "*Battling against a traditional assessment culture: The case of early English learning in Portugal*". Here, a systematic analysis of current assessment practice is conducted in which official guidelines for assessment, the teachers' perspective, and assessment instruments all come in for critical reflection. The authors highlight the importance of choosing appropriate tests for assessing the different skills of young EFL learners. They identify a mismatch between the general aim to foster oral skills in the primary EFL classroom and the widely used traditional assessment instruments which focus on written skills (pen-and-paper tests). This chapter goes on to offer explanations for the mismatch, together with many valuable recommendations on how to address the problem.

Part II: Fostering reading, spelling, and writing

There exists a noticeable lack of challenging tasks which engage primary foreign language learners cognitively (see e.g. discussions in Frisch 2013; Brunsmeier 2016; Dreßler et al. 2016; Kolb/Legutke 2019). The second part of this volume is intended to furnish readers with insights into innovative approaches towards teaching reading, spelling, and writing.

In the first chapter, Gee Macrory reflects on "'*Commas in the air': Young children's experiences of learning the orthographies of French and Spanish as a Foreign Language*". By means of learner and teacher interviews, she investigates the extent to which English primary school learners are aware of differences between the English and the Spanish/French orthographies and in how far their knowledge is related to teaching practice. Macrory detects a discrepancy between the learners' needs and current teaching practice. Whereas learners are aware of differences between the English and the Spanish/French orthographies, and try to discover rules, teachers focus more on teaching oral skills and avoid teaching literacy systematically, presupposing a detrimental effect on learners' motivation to learn foreign languages, together with a perception that including awareness-raising activities takes up too much time.

The second chapter in part II, authored by Annika Kolb, seeks to gain more insight into how early reading might be fostered by using digital resources. She asks: "*Story Apps – New ways in teaching reading in Primary EFL?*" and is interested in the potential for fostering independent reading. In a qualitative study with learners aged 8–11 in the German context, Kolb is able to identify three features of story apps that might influence reading development in a positive manner: providing comprehension support, sustaining reading motivation, and supporting the development of reading strategies. Kolb also critically discusses the quality of story apps and puts emphasis on exercising due care when selecting these.

Julia Reckermann and Karoline Wirbatz aim to uncover factors which explain why CLIL-learners have better EFL reading competences than non-CLIL learners in their chapter "*EFL reading in CLIL and Non-CLIL primary schools: A comparison of classroom reading activities, learners' preferences and actual reading comprehension competences*". The data from learner questionnaires and teacher questionnaires reveals that CLIL learners as well as children taught traditionally are highly motivated to read in the foreign language. Reckermann and Wirbatz recommend the inclusion of more reading activities at text level in both regular and CLIL classes to fully exploit the learners' potential.

In her chapter "*Scaffolding creative writing in the primary EFL classroom: Exploring the role of picture dictionaries and composition guidelines in the creation of 'Elfchen' poems*", Karen Glaser presents an analysis of writing samples composed in two fourth-grade classes. She investigates the degree to which different types of scaffolding (model texts, composition criteria, and word provision) have an effect on the adherence to the composition guidelines, on the lexical make-up of a poem, and on the children's spelling proficiency. Her research provides specific and evidence-based guidance for fostering creative writing in young learners.

Taken together, the chapters which comprise this volume confirm the challenges of learning to read and write in a foreign language. At the same time, it is hoped that researchers, teachers, teacher trainers and educational experts will feel encouraged to further develop current teaching practice and to bring into even sharper focus the topics treated here.

Bibliography

Alderson, J. Charles (2005). *Diagnosing Foreign Language Proficiency: The Interface Between Learning and Assessment*. London, New York: Continuum.

Börner, Otfried/ Böttger, Heiner/ Müller, Tanja/ Kierepka, Adelheid/ Kronisch, Inge/ Legutke, Michael/ Lohmann, Christa & Schlüter, Norbert (2016). Der Lernstand im Englischunterricht am Ende von Klasse 4 – Erste Ergebnisse der BIG-Studie. In: Böttger, Heiner & Schlüter, Norbert (eds.) (2016), *Fortschritte im frühen Fremdsprachenlernen: Tagungsband zur 4. FFF-Konferenz*. Braunschweig: Westermann, 8–44.

Brunsmeier, Sonja (2016). *Interkulturelle Kommunikative Kompetenz im Englischunterricht der Grundschule. Grundlagen, Erfahrungen, Perspektiven*. Tübingen: Narr.

Collen, Ian (2020). *Language Trend 2020. Language Teaching in Primary and Secondary Schools in England. Survey Report*. British Council. [Online: https://www.britishcouncil.org/sites/default/files/language_trends_2020_0.pdf. 24.08.2021].

Diehr, Bärbel & Frisch, Stefanie (2008). *Mark their words. Sprechleistungen im Englischunterricht der Grundschule fördern und beurteilen*. Braunschweig: Westermann.

Diehr, Bärbel & Rymarczyk, Jutta (2012). Current Developments in Teaching English as a Foreign Language in German Primary Schools. In: *Anglistik*, Volume 23, Issue 1, 13–23.

European Council (2019). Council Recommendation of 22 May 2019 on a Comprehensive Approach to the Teaching and Learning of Languages. [Online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019H0605%2802%29#document1. 24.08.2021].

Frisch, Stefanie (2013). *Lesen im Englischunterricht der Grundschule*. Tübingen: Gunter Narr.

Hammond, Jennifer (2011). Working with children who speak English as an additional language: an Australian perspective on what primary teachers need to know. In: Ellis, Sue/ McCartney, Elspeth (eds.), *Applied Linguistics and Primary School Teaching*. Cambridge: CUP, 32–43.

Hempel, Margit/ Kötter, Markus & Rymarczyk, Jutta (2017). *Fremdsprachenunterricht in der Grundschule in den Bundesländern*

Deutschlands: Eine Bestandsaufnahme des Status quo und seiner gewünschten Weiterentwicklung. Frankfurt am Main: Peter Lang. [Open Access: https://duepublico.uni-duisburg-essen.de/servlets/DerivateServlet/Derivate-47187/Hempel_et_al_Fremdsprachenunterricht_in_der_Grundschule.pdf. 24.08.2021].

Hoffmann, Lars & Richter, Dirk (2016). „Aspekte der Aus- und Fortbildung von Deutsch- und Englischlehrkräften im Ländervergleich". In: Stanat, Petra/ Böhme, Katrin/ Schipolowski, Stefan/ Haag, Nicole (eds.), *IQB Bildungstrend 2015. Sprachliche Kompetenzen am Ende der 9. Jahrgangsstufe im zweiten Ländervergleich*. Münster: Waxmann, 481–507.

Kolb, Annika & Legutke, Michael (2016). Chancen und Herausforderungen des Englischunterrichts am Übergang von Klasse 4 nach Klasse 5. In: Dreßler, Constanze/ Kolb, Annika/ Kollmann, Sabine/ Legutke, Michael (eds.) (2016). *Herausforderung Übergang: Kontinuität im Englischunterricht: Handreichung für die Praxis in den Klassen 4 und 5*. Braunschweig: Diesterweg, 9–13.

Kolb, Annika & Legutke, Michael K. (eds.) (2019). *Englisch ab Klasse 1 – Grundlage für kontinuierliches Fremdsprachenlernen*. Giessener Beiträge zur Fremdsprachendidaktik. Tübingen: Gunter Narr.

Kumaravadivelu, B. (2006). *Understanding Language Teaching: From Method to Postmethod*. Cambridge: Cambridge University Press.

Loder Büchel, Laura (2016). Introduction. In: *Babylonia*, Themenheft „Evaluation". 2/2016, 6–9.

Reckermann, Julia (2021). Reading and writing in the primary school EFL classroom. Announcement of presentation in the lecture series "English in primary education: Concepts, Research, Practice" organized by Theresa Summer, Otto-Friedrich-Universität Bamberg, Germany, summer term 2021. [Online: https://www.uni-bamberg.de/eng-didaktik/ringvorlesung. 24.08.2021].

Rymarczyk, Jutta (2018). Rolle und Professionalität der Fremdsprachenlehrpersonen. In: Eva Burwitz-Melzer/ Claudia Riemer/ Lars Schmelter (eds.), *Rolle und Professionalität von Fremdsprachenlehrpersonen. Arbeitspapiere der 38. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts*. Tübingen: Narr, 152–163.

Rymarczyk, Jutta (2021). Nachhaltigkeit im Fremdsprachenunterricht. Desiderat, Selbstverständlichkeit oder modische Worthülse? In: Eva Burwitz-Melzer/ Claudia Riemer/ Lars Schmelter (eds.), *Entwicklung von Nachhaltigkeit beim Lehren und Lernen von Fremd- und Zweitsprachen. Arbeitspapiere der 41. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts*. Tübingen: Narr, 152–163.

Szabo, Tunde (2018). *Collated Representative Samples of Descriptors of Language Competences Developed for Young Learners*. Eurocentres Consultancy for the Council of Europe. [Online: https://rm.coe.int/1680697fca. 24.08.2021].

Part I: Assessing reading, spelling, and writing

Jutta Rymarczyk

Teachers' diagnostic skills in feedback on German primary school students' first attempts to spell in English

Abstract Even though early literacy teaching seems to have found its place in primary school curricula for English as a foreign language (EFL), little is known about the ways how teachers correct and evaluate their students' first spelling attempts. Thus, the empirical study, whose piloting is presented in this paper, tries to shed some light on the diagnostic abilities of primary school teachers, both trained and untrained for the early EFL classroom, and pursues the final goal to establish sensitive and constructive corrections of spelling mistakes at primary level. As the teachers' corrections and evaluations of an authentic text written by a 3rd-grader show, both groups of teachers only assign mistakes to the categories 'right' and 'wrong'. In order to avoid misjudgements, however, this deficit perspective needs to be complemented by an achievement perspective, taking into account what students already learnt about spelling, in both their EFL and their German classrooms. A close analysis of the mistakes made in the sample text and their categoriation into overgeneralizations, interferences and arbitrary mistakes shows what is needed to assess students' texts in a way which allows for precise diagnoses and is appreciative at the same time.

1. Corrective feedback on spelling in early foreign language learning as a subject neglected by educational policy

For many years, early foreign language learning in Germany has been the source of a controversial discussion. Soon after the inception of English as a foreign language (EFL) in Grade 1 in two German federal states in 2004, a number of researchers asked for a more comprehensive approach which included a greater focus on speaking rather than just on listening, and the integration of grammatical aspects as well as literacy skills from the very beginning (Diehr & Rymarczyk 2012). As the results of empirical studies supported this demand, the shift from a playful to a more age-appropriate, analytical approach took place and in 2016, this

eventually led to the preponement of reading and writing from Grade 3 to Grade 1. Nevertheless, as early as 2011, there had been a first attempt to reduce the number of foreign language classes in favour of German and Mathematics – subjects in which primary school students had performed rather poorly (Baumert 2011). At that time, the objections put forward by proponents of an early start to foreign language teaching were successful, i.e. English and French remained in the curriculum of Grade 1. However, when a different ministry decided to postpone the beginning of foreign language teaching to Grade 3 in 2017, the arguments put forward were ignored which reduced foreign language teaching at primary level from four to two years only.

Surprisingly, Baden-Wurttemberg, as one of the two pioneer states to start foreign languages in Grade 1 in 2004, was the first state to implement this reduced foreign language programme at primary schools (MKJS BW 2020). Plans are also in place in North Rhine-Westphalia, another German state, to follow this latter example in 2021/22 (MSB NRW 2020) although, again, numerous experts in the field of foreign language learning have raised their voices against the decision (cf. Bartosch, Frisch, Kötter & Reckermann 2020).

With only four federal states left in 2021 offering foreign language classes right from Grade 1 on (Brandenburg, Mecklenburg-Vorpommern, Rheinland-Pfalz, and Schleswig-Holstein), the relevance of early foreign language learning as such is still being fiercely discussed by the German general public. At the same time, there is a consensus among experts not only about the necessity to maintain the early start but also about the need to pursue early literacy skills (cf. Hempel, Kötter & Rymarczyk 2017: 43 ff.).

Surely, returning to a late start of foreign language classes in Grade 3 entails some changes in the methodology of teaching literacy like a more rigorous integration of phonics exercises, e.g., as students are already familiar with the basics of reading and writing, or spelling, respectively. These methodological changes due to the postponement to Grade 3, however, do not influence the opinions about the need to assess early literacy skills in the second half of primary school. When asked in a survey about assessment at the end of Grade 3, almost 80 % of the responding experts (n = 38, mainly researchers, university lecturers) said it was a necessary

component and that orthography, for example, needed more attention: "Gladly greater value on orthography, to reduce the shock upon entering secondary school"[1] (Hempel, Kötter & Rymarczyk 2017: 64). However, it was also pointed out by some respondents that deviations from the orthographic norm[2] should not be seen in too negative a light or have negative consequences for the learners, up to the point where the dropping of orthographic assessment was suggested: "To encourage children to write is important and motivating, punishment of mistakes, however, demotivating. Given the difficulties of English orthography, spelling should not be a part of assessment by any means"[3] (ibid., p. 65).

According to the survey conducted by Hempel, Kötter and Rymarczyk (2017), in the 16 federal states of Germany, three forms of assessment are used at primary level: short evaluative texts (Wortgutachten), formative assessment reports (Lernentwicklungsbericht), and grades (Ziffernnoten), with the latter assumed to be the most fortright and hence, perhaps, the least pedagogical variety. Interestingly, 50 % of the respondents in the survey stated that they would want to implement exactly this variety of assessment, at the end of Grade 3 in their federal state. Furthermore, 75 % of the experts surveyed stated that they would also want marks to refer to written performance. As far as the kinds of written texts are concerned which the experts regarded as appropriate at this level of foreign language learning, not only labelling tasks and copying texts were named. More than 75 % of the respondents claimed that the young learners should also have

1 Translated by the author. Original wording: "Gerne auch etwas mehr Wert auf Orthographie, damit der Schock zur Sek I nicht zu groß ist".

2 In this paper, language errors and mistakes are not distinguished. Brown's (1994) classical distinction according to which a mistake occurs when the learner is failing to use an already known system while an error indicates the lack of a learner's grammatical knowledge, cannot be employed here. Young learners, who are the focus of our interest, are still in the process of learning vocabulary in general and how to spell these new words. Consequently, their deviations from the written norm cannot be judged as either errors or mistakes without knowing the individual learners and the input they received.

3 Translated by the author. Original wording: "Kinder zum Schreiben zu ermuntern, ist wichtig und motivierend, die Ahndung von Fehlern dagegen demotivierend. Im Bewusstsein um die Schwierigkeiten der englischen Orthographie sollte die Rechtschreibung in keinem Fall in die Bewertung einfließen."

gained some experience in free text production by the end of Grade 4, i.e. before they move on to secondary school. Grammar exercises, dictations and vocabulary tests, however, are not seen favourably, and are even forbidden in some of the states (Hempel, Kötter & Rymarczyk 2017: 43 ff.). Actually, the topic of assessing literacy acquisition seemed to be the topic which the experts who participated in the survey were most concerned about. It was this part of the survey in which the most remarks were found that called for support as to how teachers were to proceed. And it was only in this context that the request for national standards was made (ibid.).

The uncertainty with regard to assessment in early foreign language learning which became obvious in the survey is also reflected in earlier publications which go beyond the German context. Drew, Oostdam and van Toorenburg (2007), e.g., who refer to Norway and the Netherlands, ascribe the need for information on feedback to the fact that teachers in the primary EFL classroom are not specifically trained as foreign language teachers. The same line of argument can be found in Nikolov (2001) who sketches the situation in Hungary as one where there are not enough well-qualified teachers for early foreign language education. According to Nikolov, uncertainty in assessment among foreign language primary school teachers often leads to inappropriate tests, and the resulting grades tend to demotivate students easily, especially in the case of young, less-able learners.

It becomes clear that primary school teachers – contrary to popular and educational policy makers' beliefs – need to be trained for foreign language teaching and that this training needs to include a particular focus on assessment as well as, more specifically, on feedback on spelling if early literacy is to be supported. After all, teachers providing feedback to early attempts to spelling have to be able to walk the tightrope. In other words, they, on the one hand, need to avoid too much of a focus on meaning as this leads to a neglect of accuracy and does not support the students' further development of orthographic skills. On the other hand, teachers have to consider the young learners' limited ability to follow metalinguistic explanations and they have to strive to keep them motivated which can be jeopardized if corrective feedback is not perceived by each learner as being useful.

A review of empirical research studies to ascertain which stance on orthographic assessment is most suitable for the primary foreign language classroom, yields only relatively little results. The three foci of early foreign language learning, corrective feedback and feedback on spelling taken together have hardly been researched. Ghandi and Maghsoudi (2014), e.g., critique the scarcity of research into orthographic assessment for the secondary school sector: "Though a great deal has been revealed with regard to various applications of corrective feedback in educational settings, there still seems [sic!] a paucity of research in the domain of corrective feedback and its effects on spelling errors" (ibid., p. 55). Schoormann & Schlak's metastudy on corrective feedback in second and foreign language education (2012) shows a similar result. While in the framework of the cognitive-interactionist as well as the socio-cultural approach to second language acquisition research a plethora of empirical studies has tried to identify the most effective feedback strategies (cf. ibid.), the results of these studies cannot be fully exploited for our purpose, i.e. insights into corrective feedback on spelling appropriate for primary school level. First, they mainly took oral corrective feedback into account and second, they did not concentrate exclusively on young learners. However, one result which the author of this article believes can definitely be transferred to our research interest is the insight that corrective feedback needs to be custom-made to be effective (ibid.). Traditional approaches to corrective feedback mainly focused on external factors. Its effectiveness was believed to be influenced by first, the characteristics of the feedback to be applied (its length, form, complexity etc.) and second, the linguistic target structure in question. As Schoormann & Schlak convincingly point out, this focus on external factors, however, needs to be complemented by internal factors to be taken into account (ibid.). The learners' attitudes towards the language and towards corrective feedback, their prior knowledge about foreign language learning, their foreign language level, their age, and factors such as foreign language learning aptitude, motivation, and language anxiety need to be considered (ibid., p. 176). Yet, the most crucial point in the context of early foreign language learning seems to be the social dynamics of the participants, i.e. the matching point between the teacher and their attitudes, experience and corrections routines on the one hand, and the

learner with their sensitivities and characteristics enumerated above, on the other hand.

Turning to the scarce literature connected to corrective feedback on young learners' attempts to spell, Nikolov (2017), for example, stresses the educational implications for teachers in this context by stating that: "[Teachers of young language learners] should apply and benefit from innovative approaches to assessment including […] diagnostic assessment" (ibid., p. 249) and that they need "techniques to tap into what children can and cannot do […]" (ibid.). The necessity to be able to pay attention to language errors and mistakes mentioned by Nikolov is also clearly pointed out by other researchers like Pfingsthorn: "Language errors can also serve as a highly informative resource, to learners, teachers, and researchers, as they reflect both the current stage of learners and the ability to deal with communication problems" (Pfingsthorn 2013: 2).

While this is undoubtedly true, the question remains whether the insight into the value of language errors causes researchers and teachers alike to take an achievement perspective or a deficit perspective on assessment in general and how this is converted into concrete, practical steps in the classroom. Especially in the context of teaching beginning learners, there seems to be a strong tendency to value the students' efforts: "When learners manage to express the intended meaning despite lacking resources, their errors can also be perceived as signs of success" (ibid., p. 2), as assessment is meant to encourage young learners to further develop their spelling skills (cf. e.g. Nikolov 2017: 250) and by no means to demotivate them. The North Rhine-Westphalian EFL-curriculum for primary schools referring to the end of Grade 4, e.g., asks teachers to accept incorrectly spelled words that are still recognisable as correct (MSW NRW 2008: 82). While this seeks to be student-oriented, it is of no concrete help to teachers as the phrase "display short words from their oral vocabulary 'phonetically' more or less correctly in writing"[4] leaves it to the teacher's own judgement to decide what is meant by 'more or less correctly'. Interestingly, in the new curriculum which is going to be valid for students from school year 2022/23 onwards, there is an indication of a stance towards orthographic

4 Translated by the author. Original wording: "geben kurze Wörter aus ihrem mündlichen Wortschatz 'phonetisch' einigermaßen korrekt schriftlich wieder".

accuracy. There we find the wording: "[Students] use basic rules of English orthography in the production of their own texts (lower case of nouns, capitalization of the pronoun "I"[5]) (QUA-LiS NRW 2021: 14).

The tendency towards an increased emphasis of orthographic accuracy is regarded as positive by the author of this article. It is important to see that a pedagogical attitude that is very cautious might be quite well-meaning, but, at the same time, counterproductive in the effort to support young learners in their literacy acquisition, and particularly in spelling. Ferris (2002), for example, underlines that while indirect feedback that elicits the correct forms is beneficial for more experienced writers, young learners depend on direct feedback on their spelling mistakes as they lack the linguistic background for self-correction. Similarly, Rymarczyk (2011) emphasizes the need to fight fossilization of invented spelling by providing input tailored to the students' needs. These needs are, among other factors, determined by the time constraints of the primary foreign language classroom. The contact time for primary students learning a foreign language in an instructional setting in Germany generally comprises two 45-minute sessions a week. This is far less than what is allowed for first language literacy acquisition and even for reading and writing in a second language learning environment. Given this time restriction, natural language acquisition mechanisms hardly come into play and thus need to be accompanied or even replaced by direct instruction, i.e. feedback (also cf. Kahn-Horwitz 2015: 149).

Our research interest subsequently focuses on teachers' diagnostic skills in the evaluation of primary school students' first attempts to write in English. Only if the teachers' linguistic, analytical, and diagnostic skills are well developed, are they able to design input and feedback in a way which is tailored to the young learners' needs and conducive to their literacy development.

5 Translated by the author. Original wording: " wenden elementare Regeln der Rechtschreibung des Englischen bei der Produktion eigener Texte an (Kleinschreibung von Nomen, Großschreibung des Pronomens 'I'").

2. An empirical study on primary school teachers' diagnostic skills

2.1 *Research question*

In order to find out whether teachers' diagnostic assessment is based on a deficit perspective ("geared toward weaknesses so that problem areas can be further practiced in a focused fashion", Nikolov 2017: 251) or on an achievement perspective ("errors can also be perceived as signs of success", Pfingsthorn 2013: 2) the following research question was decided on:

Research Question

To what degree do subject knowledge, pedagogical content knowledge (*fachdidaktische Kompetenz bzw. Wissen*) and further individual factors (specific academic education and experience, perceived analytical complexity, attitudes based on affect) influence the quality of teachers' diagnostic judgments on their young EFL learners' spelling skills?

This question is quite complex because of two reasons: First, because connections between a relatively high number of factors and the quality of the teachers' diagnostic judgements are to be found. Second, and this seems to be the more decisive point in this respect, the factors enumerated need to cover two groups of teachers: those trained for the job and those who have not gone through teacher training as EFL-teachers. Unfortunately, the latter group outnumbers teachers with a full-fledged teaching qualification. While in 2016 the number of non-trained or just partially trained EFL primary school teachers was believed to be higher than 50 % (Börner, Böttger, Müller, Kierepka, Kronisch, Legutke, Lohmann & Schlüter 2016, 18), this figure was estimated at 70–80 %[6] in 2018.

Of the factors mentioned in the research question, "subject knowledge", "pedagogical content knowledge" (*fachdidaktische Kompetenz bzw. Wissen*), and "perceived analytical complexity" primarily refer to trained teachers. The last factor ("perceived analytical complexity") aims at the learner mistakes which the teachers have to assess. The restriction by the word "perceived" was included because even among trained teachers we cannot be sure whether error treatment was taught in any of

6 Informal conversation among professors of foreign language didactics at a conference in 2018.

their university courses. Reading and writing in early foreign language learning is still a fairly recent topic, and error treatment just a subtopic of the larger field. A search in the schedules of lectures of universities of education in Baden-Wurttemberg showed that hardly any seminars had been offered on early literacy acquisition and spelling in the period 2018–2019 (Jaskorsky 2019).

"Perceived analytical complexity" and the remaining factors in the research question ("specific academic education and experience" and "attitudes based on affect") are listed for the large group of non-trained teachers. The latter refers to the teachers' attitudes towards their EFL-classes. Very often, non-trained teachers feel rather insecure as they lack the subject knowledge, pedagogical content knowledge, and foreign language competence to teach English. This might even lead to what can be perceived as a general refusal to follow the syllabus of the EFL-classroom. In its extreme form, this refusal means that the teachers do not teach English at all even though they are obligated to do so by their timetable. Others teach English in German or have the children carry out activities characterized by extremely low demands[7]. The differentiation in the individual factors is to acknowledge and observe the degree of heterogeneity among current EFL primary school teachers we face at the moment.

In the piloting of our project, we worked with 23 primary school teachers in order to obtain some insight into our research interest which comprises the questions listed in the following subsection. Twelve of these teachers were not trained to be EFL primary school teachers whereas the eleven teachers who hold the respective diploma all work as teacher trainers in the second phase of the German teacher education system.

2.2 *Research interest*

1. Do teachers categorize mistakes or do they only distinguish between right and wrong?
2. How do teachers deal with phonographic (and other) interferences from German? Are they seen as an asset which shows the learner's phonemic awareness?

7 Personal observation and anecdotal information.

3. Do teachers take on an achievement or a deficit perspective in the evaluation of the mistakes?
4. Which attitudes towards teaching English at primary schools can be determined from the teachers' answers?

Question No. 2 is the most crucial question here as it is presumably phonographic interferences that occur most often. This dominance is due to the fact that the children learn how to read and write in their German classroom which includes about five lessons per week. In contrast to this, the time dedicated to reading and writing in English is almost non-existent at the moment. This means that the children use their knowledge of German phoneme-grapheme correspondences to write in English. Numerous orthographic mistakes are the almost inevitable consequence. This is due to the fact that a positive transfer of phoneme-grapheme correspondences from German (as a language with a relatively shallow orthography and a regular system of phoneme-grapheme correspondences) to English (as a language with a deep orthography and a highly irregular system of phoneme-grapheme correspondences) is hardly possible (cf. Frisch (2013: 55 ff) for a comparison of German and English orthography).

When it comes to the resulting interferences, teachers might not see any learning gains in them but they might just regard them as misspelled words. This might be different with overgeneralizations as here, linguistic material from the target language is used. Yet, since this is merely a presumption, we need to see how teachers actually deal with different kinds of mistakes.

2.3 *Data collection*

In order to investigate how primary EFL teachers handle mistakes in their students' spelling and which factors might influence this process, we used teacher corrections of a text written by a 3rd-grader (cf. Fig. 1), teacher comments on the student's spelling, and a questionnaire to gather the teachers' bio data, their ways of spelling instruction, and their attitudes towards teaching English at primary schools.

The remaining part of this chapter primarily focusses on the teachers' corrections and feedback only (research interest, questions 1–3). In order

to do so, however, first, the student text as an instrument of data collection is introduced, and subsequently the teacher corrections and comments are presented and discussed. The teachers (n = 23) had been asked to correct the text and also to comment on it in the way they usually do. In case they had not corrected and/or commented on spelling before, they were requested to do it in a way they saw fit.

The short text which we asked the teachers to correct contains several kinds of mistakes which we assigned to five categories: "overgeneralization", "orthographic interference", "phonographic interference", "lexical interference", and "other". This categorization is based on well-established distinctions like the one employed by Dulay and Burt (1974), who, in terms of negative transfer, distinguish between developmental errors (similar to L1 acquisition, here: overgeneralization), interference errors (reflecting the L1 structure, here at the orthographic, phonographic, and lexical level), and unique errors (neither developmental nor interference, classified here under "other"). The graphic below shows the text the teachers were asked to correct and comment on.

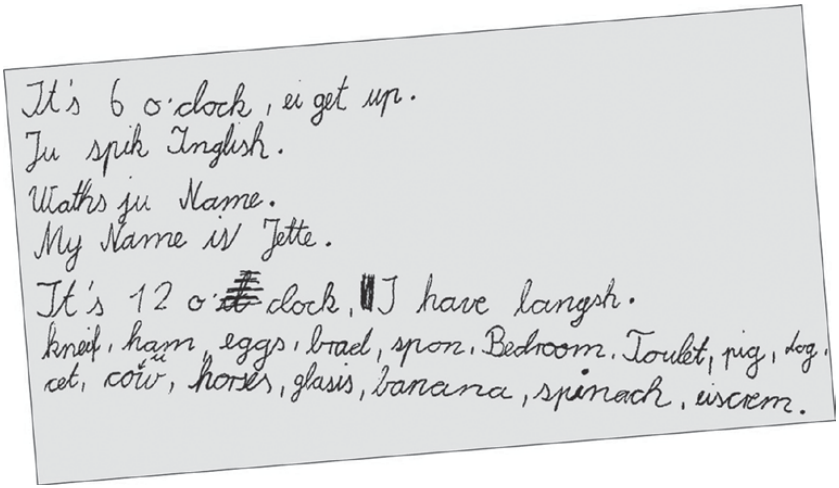


Fig. 1: Learner text, Grade 3

The phenomenon of overgeneralization is an intralingual one leading to errors which "arise when the learner creates a deviant structure on the basis of other structures in the target language" (Ellis 1994: 59). Overgeneralization can be regarded as a learning strategy based on the use of hypothesis formation, is also found in first language acquisition, and occurs at various linguistic levels (phonology, morphology, grammar, etc.). The instances of orthographic overgeneralizations found in the text employed in this research are **langsh*, **couw*, and **eiscrem* (cf. Fig. 2, underlined).

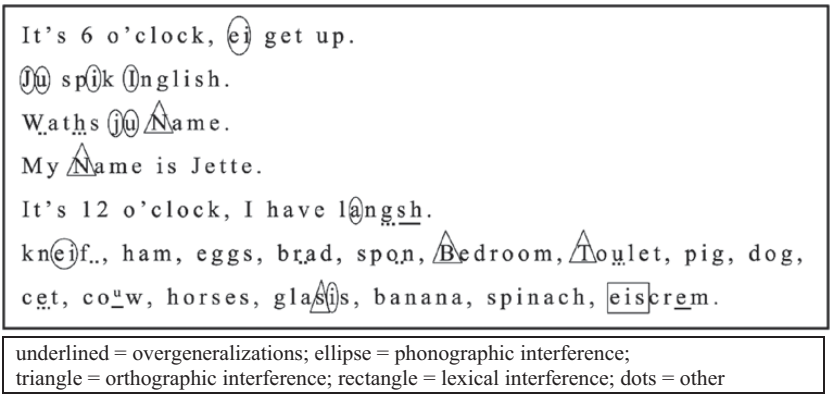


Fig. 2: Categorization of mistakes

To our minds, overgeneralizations are the most positive kind of mistake as they are clearly based on the learner knowledge about the target language. The child, who produced this text, spelled, for example, *cow* with the letter sequence <ou> instead of with just <o>. The reason for this spelling might be the child's knowledge of familiar words like *mouth*, *mouse*, *house* etc. which feature the letter sequence <ou> for the diphthong /au/.

Three further categories of mistakes which occur in the text belong to the interlingual phenomenon of language contact called interference. Acccording to Mackey (1962), "Interference is the use of features belonging to one language while speaking or writing another" (cited in Müller, Kupisch, Schmitz & Cantone 2006: 16).

The instances of interferences which appear, include, for example, *eiscrem as a lexical interference because here the whole German word *Eis* has found its way into English, just with a lower case letter instead of a capital letter (cf. Fig. 2, rectangle). Furthermore, there are four instances of orthographical interferences since, first and foremost, the young writer has not yet fully mastered the rule that English nouns (with the exception of proper nouns) are spelled with lower case letters: *Name (2x), *Bedroom, *Toulet. The fourth orthographical interference, *glasis, is due to the use of a single consonant (<s>) where English asks for a double consonant (<ss>) (cf. Fig. 2, triangle). The third kind of interference, which can be found in this text, is phonographic interference: *ei, *Ju (2x: J u), *spik, *Inglish, *langsh, *kneif, *glasis (cf. Fig. 2, ellipsis). This is based on the writer's use of phoneme-grapheme correspondences that belong to her dominant language, German (also cf. below).

Interferences of various kinds are also regarded as positive in our context of early literacy acquisition as they show some knowledge of the very nature of phoneme-grapheme-correspondences and phonemic awareness in general. This means that even if a learner makes a high number of mistakes, this learner is surely not dyslexic as long as these mistakes are interferences. In the case of interferences, learners simply choose structural features from the wrong language. In contrast to this, mistakes of the category "other" (the fourth category of mistakes in the text, here appearing eight times) might give cause for some concern.

The distribution of the mistakes made across the various categories clearly shows that phonographic interferences are the largest group of mistakes (actually 10 out of a total of 27 mistakes) in the given text.

The frequency of interferences is a well-known phenomenon. Analyses by Maas and Mehlem (2003: 509–511) and Weth (2008: 197–206) prove that both primary and secondary school students transfer their knowledge of forms pertaining to the written word to the linguistic structures of a further language which is just known orally to them. In the context of primary school students, Butler (2017: 361) cautions that as "young learners' L1 is not fully developed, the nature of transfer from their L1 to L2/FL may be very different from adults" and that their "unique characteristics […] require special consideration in assessment" (2017: 372). The most important characteristic feature in our context is undoubtedly the

fact that children have not yet fully grasped that the phoneme-grapheme correspondences of German and English differ and that because of this, their L1-knowledge of certain letter realizations of sounds cannot be applied when writing English. Hence, in the case of phonographic interferences, the children's invented spelling of English words is not based on structures belonging to the target language, but it is invented spelling of English based on German phoneme-grapheme correspondences. The following paragraph details our perspective on this kind of interference in texts produced by young learners of a foreign language.

When learners want to write an unknown word in a foreign language with an alphabetic writing system, they use the alphabetic strategy, i.e. they assign letters to the sounds they recognized in the lexemes in their input. The mistakes in the text employed in this study occurred because of the wrongly chosen German phoneme-grapheme correspondences, but the young writer obviously knows how to link sounds to letters in general which proves that she is not dyslexic. On the contrary, the high number of phonographic interferences shows that the child who produced the text possesses a high level of phonemic awareness as she heard the sounds of the individual lexemes correctly. Mishearing the phonemes would not have allowed her to choose, for example, the letters <ei> to spell "I" in line 1 of her text. It is to be noted that the letter sequence <ei> represents the diphthong /ai/ in the German phoneme-grapheme correspondences, so, from this perspective, the girl's attempt to write the word "I" is absolutely correct. Consequently, a highly developed level of phonemic awareness can be attested to her, if not an advanced level of literacy development – in general.

2.4 *Data analysis*

Of the 23 teachers who participated in the piloting six did not correct the text but only commented on the student's spelling. While some of these respondents did not explain their decision, others of this group wrote that they would keep the student's text as it is but correct a transcript and hand both text versions back to the student. This corrected copy, however, was not always sent back to us, so there was no insight into these teachers' feedback practices. Those teachers who did correct the text, also varied

considerably in the way in which they gave their feedback. There was both direct and indirect feedback[8] to be found, as well as comprehensive and selective error correction[9]. The identification of error types in general varied to a very high degree while direct and uncoded feedback prevailed by far. The following detailed description shows the great heterogeneity among the teachers' feedback practices.

– Six texts remained completely untouched. Among these 26,09 %, three teachers planned to return a corrected transcript with the original 'untouched' text; one wanted to provide her feedback just orally; one said she was too inexperienced to work on the text; one gave an example of how she would correct a mistake in her comment in order to explain the researcher the way in which she would provide her feedback.
– Four texts (17,39 %) showed examples of indirect feedback: The mistakes were highlighted by means of dots and crosses. In two texts, whole words were underlined. The remaining 13 texts (56,52 %) in which direct feedback was provided, show different ways in which the correct spelling of the words is presented. Eight teachers wrote the correct spellings into the text, five provided them in their comment.
– Only three teachers (13,04 %) chose selective feedback; the majority chose comprehensive feedback, partly by providing a corrected transcript of the whole text (cf. Fig. 1). Two teachers added to their comprehensive

8 There are six main different types of corrective feedback. Direct (or explicit) feedback means that the teacher provides the exact correct form for the learner or at least a metalinguistic explanation of a rule accompanied by some examples of the rule in use (Bitchener & Knoch, 2008, p. 411). Indirect feedback, in contrast to this, only indicates that there is a mistake. It can be used as coded or uncoded feedback. As in uncoded indirect feedback mistakes are only underlined or circled, students are left to diagnose and correct the error themselves. Coded feedback specifies the exact location of a mistake as well as the type of mistake with a code from a standardized set of symbols (Bitchener, Young & Cameron, 2005: 193). Direct feedback which provides the exact correct form and which is uncoded seems to be the type of feedback which asks for the least learner investment.

9 Whereas in selective corrective feedback only a certain number of mistakes are considered, in comprehensive feedback all mistakes made are corrected.

feedback that they might/would have asked the girl to correct some selected words on her own, for example, by using a dictionary.

– Coded feedback, which mentions the category of a mistake, was hardly used at all; it occurred in only three corrections (1,304 %). Two corrections displayed the use of the category "lexical mistake – (W)"[10], and another highlighted wrong capitalization in nouns, employing downwards pointing arrows ((⇩)).

The figure below (Fig. 3) offers an example of feedback as provided by the teachers participating in this research. It displays direct, uncoded and comprehensive corrective feedback in order to complement the overview above as closely as possible. The combination of these three kinds of feedback was chosen by the most participants and will be discussed in the following to allow for some concrete insight into current feedback practices.

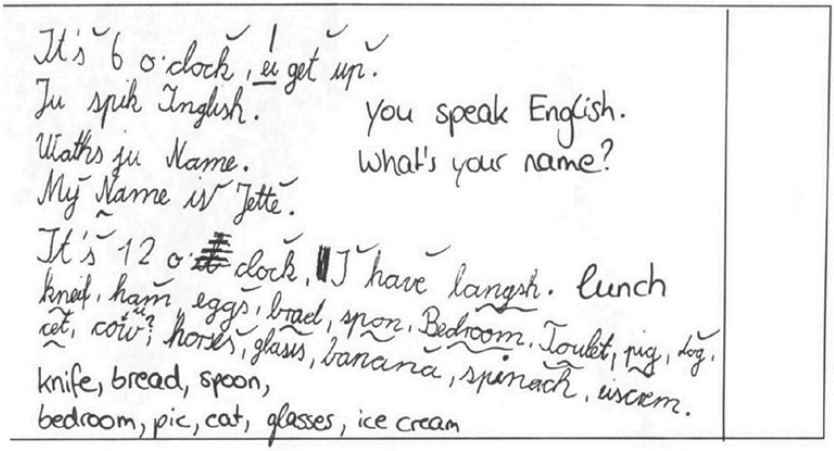


Fig. 3: Text with feedback (direct, uncoded, comprehensive)

10 In both cases, however, this category was used for the spelling "Name" where it supposedly referred to the alleged German word while orthographic interference (wrong capitalization) was more likely the case due to several instances of wrongly capitalized nouns.

In this example of the text corrections (cf. Fig. 3.), words which are written correctly are mostly ticked off whereas mistakes are underlined with straight or wavy lines. The mistakes are mainly corrected directly, i.e. the correct form is provided, here either above the word, at the end of the same line, or below the text. A number of problems which arise if there is no direct correction but only the low lines, could be noted: First, it may not be clear to the student where exactly they find the mistake. Second, the number of mistakes is not indicated. Some examples of this latter problem are the corrections of the wrongly spelled words **Bedroom* and **Toulet*. Even though in **Bedroom* the only mistake is the initial capital letter, most of the word is underlined. Without the direct feedback below the text, the student could not possibly know where exactly the spelling mistake occurred. In contrast to this, the connected lines under **Toulet,* which features two mistakes, are positioned more or less under the letters meant to represent the diphthong; thus the faulty capitalization is neglected. In addition to this, here no direct feedback given (cf. Fig. 3.).

Additionally, the comments which accompany the corrected text and are meant to address the student show a considerable variety:

- Five comments (21,74 %) were exclusively positive, i.e. there was no indication at all that there were any mistakes in the text.
- Eighteen teachers (78,06 %) referred to categories of mistakes. Among those, 16 teachers (69,57 %) explained that English and German phoneme-grapheme correspondences differ and six (26,09 %) mentioned the rules for capitalization in English.
- Five teachers (21,74 %) provided very few spelling corrections in their comments.

In addition to this, the generally motivating comments which praised the students' work included some recommendations, i.e. advice for the student to consult a dictionary, that the direct corrections should be written out by the student, or that the corrected spelling should be practiced with index cards. Furthermore, some teachers also stressed that they understood what the student had written and/or that students at her grade level do not need to be able to spell correctly. In a similar vein, information about incorrect spelling was sometimes not addressed to the student herself but in a general way using the neutral German pronoun "man" whose

lack of direct personal reference matches passive constructions in English as can be seen in the following comment: "Unfortunately, there are also many words which are written in a way very different from how they are pronounced. These words need to be learnt by heart"[11]. One might assume that this impersonal way of mentioning the mistakes was chosen in order not to give rise to any negative feelings in the student for not having been able to spell the words correctly. It seems that the teachers are careful not to demotivate the student, and if praise is not possible, choose very considerate forms of articulating their criticism.

Following up the feedback shown in Figure 3, the comment written by the same teacher is to serve as one of the examples which are discussed in the remaining part of this paragraph. These comments allow some concrete insight into the teachers' feedback in text form. The teacher's enthusiastic comment which goes with the corrective feedback illustrated in Figure 3 reads: "You already have a great, large vocabulary in English and are able to express yourself very well. Super!"[12] and is accompanied by a large smiley. While this is certainly a motivating remark, it does not help the learner to learn the spelling rules of English as it does not refer to the mistakes made. In this example, it becomes obvious that Ferris (2002) is right in his above mentioned conviction that young learners are in need of direct feedback in order to be able to pursue correct orthographic forms. Indirect feedback and the attempt to elicit the correct forms from the students are indeed doomed to failure due to a lack of input which includes the correct form in question unless the students are allowed to consult a dictionary (for selected items).

Another example of feedback to the text used in this research shows a completely different approach: The teacher did not correct the text at all. She explained since writing was not a part of English lessons at her school, she would not correct any texts. Her feedback (which she would provide orally), however, contains more than just praise for the good parts:

---

11 Translated by the author. Original wording: "Es gibt leider auch ganz viele Wörter, die man anders schreibt als man sie ausspricht. Diese Wörter muss man dann leider auch auswendig lernen."

12 Translated by the author. Original wording: "Du hast schon einen tollen großen Wortschatz in Englisch und kannst dich sehr gut ausdrücken. Super!"

> You are already able to write many English words and sentences. You remembered that most words in English are written with lower case letters.
> You wrote many English words right.
> With some, you wrote like the words are spoken. [Here I would give some examples (Ju → You, Inglish → English…) and that it is not that easy in English after all. As a whole, I would give the student positive feedback.][13]

Along with her praise for the girl's achievements, first, she repeats a spelling rule (lower case letters for nouns) and, second, she reminds the child that the phonographic approach to spelling does not work for English. In order to illustrate her explanations, she would provide "some" examples (including direct feedback) which corresponds to Lee's suggestion (2003) that comprehensive corrective feedback should be avoided in favour of selective feedback. Comprehensive feedback might be too time-consuming for teachers and is too easily frustrating for learners (Lee 2003: 218).

2.5 *Interpretation of the results*

Considering that these are the data of only 23 teachers, the diversity represented means that there is hardly any homogeneity in the corrective feedback of EFL teachers at primary school level. The teachers seem to be mainly, if not exclusively, guided by subjective theories which corresponds to the finding mentioned above that the topics of early literacy acquisition, spelling, orthography assessment and corrective feedback can hardly be found in the schedules of lectures of universities of education in Baden-Wurttemberg (cf. Jaskorsky 2019). One might even argue that the result of this research, i.e. that the participating teachers were mainly guided by subjective theories, might actually be a general tendency. After all, EFL course books for primary school level usually do not include any guidelines for teachers on how to correct students' spelling mistakes either.

---

13 Translated by the author. Original wording:

> "Du kannst schon viele Wörter und Sätze im Englischen schreiben. Du hast Dir gemerkt, dass die meisten Wörter im Englischen klein geschrieben werden. Viele englische Wörter hast Du richtig geschrieben.
> Bei manchen hast Du so geschrieben, wie man die Wörter spricht. [Hier würde ich einige Beispiele machen (Ju → You, Inglish → English…) und erklären, dass das im Englischen auch nicht ganz einfach ist. Insgesamt würde ich der Schülerin eine positive Rückmeldung geben.]"

Final support for the claim that teachers' mostly rely on their subjective theories in the corrective feedback to their students' early spelling attempts can be seen in the request for assistance made in the context of the survey by Hempel, Kötter and Rymarczyk (2017) mentioned above. Respondents in the survey had asked to support teachers in the field of early literacy and especially with respect to orthography assessment. So, interim results indicate that, currently, it appears to be necessary to enable teachers to turn to a systematic form of corrective feedback on spelling. This falls in line with the claim made by Hasselgreen, Kaledaite, Maldonado Martín and Pižorn (2011: 19):

> Corrective feedback on errors can be given, and the writing can be shown to and discussed with the pupil and his/her parents, and kept to compare with earlier or later performances. However, without some systematic way of carrying out this assessment, it may have little formative value and can yield imprecise summative information.

With reference to our research interest, the first question presented above („Do teachers categorize mistakes or do they only distinguish between right and wrong?"), can be negated almost completely. With the exception of denoting vocabulary gaps as lexical mistakes and highlighting wrong capitalization, the teachers did not assign any mistakes to different categories. However, they did not just distinguish between right and wrong either. A fair number of teachers (16 teachers or 69,57 %) explained in their comments that the phoneme-grapheme correspondence of English is different to the German one, i.e. referred to phonographic interferences. While an explanation like this is considered direct feedback (cf. Footnote 7 above), it bears some similarity to coded feedback which assigns mistakes to categories.

In answer to Question 2 ("How do teachers deal with interferences from German? Are they seen as an asset which shows the learner's phonemic awareness?"), it can be stated that phonographic interferences seem to be the focal point of interest to the teachers. This prevalence is certainly due to the predominance of this kind of mistake in the text for which feedback had to be given: More than two thirds of the mistakes in the student's text are in this category. However, the fact that this category is also the one which 69,57 % of the teachers did not only correct but actually explained to the student, also shows the significance teachers assign to phonographic interferences in young learners' literacy acquisition.

Whether the teachers see interferences as an asset which shows the learner's phonemic awareness cannot be seen in the vast majority of the data. First and foremost, interferences seem to be treated in the same way as other mistakes. However, the phenomenon of phonographic interference seems to be regarded as so significant that it needs to be explained to the student. Only one teacher seemed to conclude from the mistakes that the student's pronunciation was good: "Certainly you can speak English really well"[14].

As for Question 3 ("Do teachers take an achievement or a deficit perspective in the evaluation of the mistakes?"), with the exception for one, all the teachers praised the skills which the student had already developed. The achievement perspective seems definitely to prevail: "I am proud of you!" – "You did a great job! You have already written very many words in exactly the way in which the words are written in the English language." – "Great that you can already write so many words and sentences!"[15]. This leads us to the hypothesis that praise seems to refer exclusively to the words spelled correctly. The application of knowledge of German orthography transferred to English is not mentioned as a positive strategy in any of the comments. The answers to the questionnaire do not tell us either whether teachers view interferences as assets or deficits. This fact needs to be regarded as a crucial point with regard to the teachers' diagnostic skills and assessment. The teachers might actually not be able to distinguish advanced learners (those whose mistakes are interferences or even overgeneralizations) from learners at risk at being dyslexic (whose mistakes are random, i.e. not based on any interlingual or intralingual transfer of knowledge).

Our last question, Question 4 ("Which attitudes towards English at primary schools can be derived from the teachers' answers?") is only indirectly answered. The answers in the questionnaire tell us that negative

14 Translated by the author. Original wording: "Sicher kannst Du Englisch ganz toll sprechen."

15 Translated by the author. Original wording: "Ich bin stolz auf Dich!" – "Das hast Du prima gemacht! Ganz viele Wörter hast Du auch schon genau so geschrieben, wie die Wörter in der englischen Sprache geschrieben werden." – "Toll, dass Du schon so viele Wörter und Sätze schreiben kannst."

attitudes towards Early Foreign Language Learning seem to dominate among untrained teachers: Half of these teachers stated in their answers that they teach much less than they are supposed to. They teach only five instead of 50 minutes or 45 instead of 90 minutes. They use the time saved for other subjects or to talk about class affairs. However, it is interesting that, in spite of this, the majority of this group opts for starting EFL classes in Grade 1, and for starting the teaching of spelling in Grade 3.

Two teachers expressed their insecurity about how to give feedback and hence did not provide any or only indirect feedback, respectively. They explain their insecurity as being due to a lack of experience with "free" writing and the (alleged) restriction of spelling to Grades 3 and 4. Of these two teachers, one is trained and one is not trained to teach EFL.

Our main research question, which among other things targeted the difference between trained and untrained EFL teachers, cannot yet be answered. Even though the number of trained teachers in this piloting was almost equivalent to the number of those who were not trained for EFL, the number of participants in total was still too small to show any differences between the two groups.

Additionally, the decision made by some teachers not to correct the text but to state that the student would be provided with a transcript, might be both the well-founded pedagogical goal not to flaw the value of the student's work or the avoidance of the face-threatening risk to make mistakes oneself.

Similarly, the fact that only very few teachers opted for selective corrective feedback might be due to supposed social desirability rather than to the genuine conviction that all mistakes in a student's text of the kind presented should actually be pointed out and/or corrected. After all, the teachers might have felt obligated to prove that they could spot all the mistakes in the student's text when they were asked to correct the text.

Last but not least, since grades were not considered in the evaluations or rather, in the comments, we cannot say for sure whether the teachers considered interferences – or overgeneralizations for that matter – in the same way as the other mistakes. Yet, there is a tendency towards equating all mistakes as shown in two cases where teachers counted the numbers of mistakes in the margins.

3. Conclusion

In conclusion, despite the lack of a final answer to the main research question, valuable insights were gained by this piloting. Five conclusions can be drawn on the basis of the data:

1. Teachers hardly differentiate between (i.e. code) students' spelling mistakes.
2. Praise seems to refer exclusively to the words spelled correctly.
3. The results do not provide any insight on whether teachers view interferences as assets or deficits.
4. Direct, uncoded correction is held to be the main kind of feedback which helps students improve their spelling.
5. Negative attitudes towards early EFL or personal insecurity in teaching EFL seem to obscure the recognition of the beneficial effects of corrections and feedback on spelling.

As a consequence, it is our challenge now to find a research design which allows us to find out whether teachers view overgeneralizations and especially interferences as assets or deficits.

Eventually, it is our aim to design a module for teacher training courses which qualifies teachers to deal with mistakes in a professional way, i.e. to categorize them and to evaluate them differently on the basis of these categories. Maybe a greater awareness of the nature of mistakes is needed among teachers in order to have them realize that their students' mistakes are more informative when they are not just contrasted with the correct spelling. It is our aim to qualify teachers to see whether a child just needs "regular" support or whether they need special help as they might be dyslexic to some degree.

These goals go in line with what has been said about the effective use of assessment for learning, i.e. formative assessment (Wiliam 2018, Teaching & Learning Team, Cambridge International 2019): There needs to be "fidelity of implementation" (ibid., no page), which means that students gain the opportunity to implement the feedback in the future. For the teachers "fidelity of implementation" means that they align their teaching with the students' true developmental level. Only then, can the teachers' corrections be expected to make a positive impact on achievement (ibid., no page).

The last aim and challenge which should be mentioned here is the search for a way to motivate teachers to create a transparent connection between their corrections and their evaluation of a student's text. After all, it seems of utter importance to inform the young learners and also their parents that there are "good mistakes" and what this means for the students on their path to becoming successful readers and writers.

Bibliography

Bartosch, Roman/ Frisch, Stefanie/ Kötter, Markus/ Reckermann, Julia (2020), *Stellungnahme von Vertreter*innen der Fremdsprachenforschung zur Rückverlegung des Englischunterrichts in Klasse 3 in NRW zum Schuljahr 2021/2022*. [Online: https://www.fmks.eu/files/fmks/download/Stellungnahme%20zum%20Masterplan%20Grundschule%20NRW_Nov2020_fmks.pdf. 22.05.2021].

Baumert, Jürgen (2011), *Expertenrat "Herkunft und Bildungserfolg". Empfehlungen für Bildungspolitische Weichenstellungen in der Perspektive auf das Jahr 2020 (BW2020)*. Stuttgart.

Bitchener, John/ Knoch, Ute (2008), The value of written corrective feedback for migrant and international students. *Language Teaching Research*, 12(3), 409–431.

Bitchener, John/ Young, Stewart/ Cameron, Denise (2005), The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*, 14(3), 191–205.

Blömeke, Sigrid/ Gustafsson, Jan-Eric/ Shavelson, Richard J. (2013), Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift für Psychologie,* 223(1), 3–13.

Börner, Otfried/ Böttger, Heiner/ Müller, Tanja/ Kierepka, Adelheid/ Kronisch, Inge/ Legutke, Michael/ Lohmann, Christa/ Schlüter, Norbert (2016), Der Lernstand im Englischunterricht am Ende von Klasse 4 – Erste Ergebnisse der BIG-Studie. In: Böttger, Heiner & Schlüter, Norbert (eds.) (2016), *Fortschritte im frühen Fremdsprachenlernen: Tagungsband zur 4. FFF-Konferenz*. Braunschweig: Westermann, 8–44.

Brown, Douglas (1994), *Principles of Language Learning and Teaching*. New Jersey: Prentice Hall Regents.

Butler, Yuko G. (2017), Assessing young learners. In: Tsagari, Dina & Banerjee, Jayanti (eds.) (2017), *Handbook of Second Language Assessment* (1st edn.). Boston: Mouton de Gruyter, 359–375.

Diehr, Bärbel & Rymarczyk, Jutta (2012), Current developments in teaching English as a foreign language in German primary schools. *Anglistik* 23: 1, 13–23.

Drew, Ion/ Oostdam, Ron/ van Toorenburg, Han (2007), Teachers' experiences and perceptions of primary EFL in Norway and the Netherlands: A comparative study. *European Journal of Teacher Education* 30: 3, 319–341.

Dulay, Heidi/ Burt, Marina K. (1974), Errors and strategies in child second language acquisition. *TESOL Quarterly* 8: 2, 129–136.

Ellis, Rod (1994), *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

Ferris, Dana R. (2002), *Treatment of Error in Second Language Student Writing*. Ann Arbor: The University of Michigan Press.

Frisch, Stefanie (2013), *Lesen im Englischunterricht der Grundschule*. Tübingen: Narr Francke Attempto.

Ghandi, Maryam/ Maghsoudi, Mojtaba (2014), The effect of direct and indirect corrective feedback on Iranian EFL Learners' spelling errors. *English Language Teaching* 7: 8, 53–61.

Hasselgreen, Angela/ Kaledaite, Violeta/ Maldonado Martín, Natàlia/ Pižorn, Karmen (2011), *Assessment of Young Learner literacy linked to the Common European Framework of Reference for Languages*. Graz: European Centre for Modern Languages, Council of Europe Publishing.

Hempel, Margit/ Kötter, Markus & Rymarczyk, Jutta (2017), *Fremdsprachenunterricht in der Grundschule in den Bundesländern Deutschlands: Eine Bestandsaufnahme des Status quo und seiner gewünschten Weiterentwicklung*. Frankfurt am Main, Bern, Wien: Peter Lang. [Online: https://duepublico2.uni-due.de/servlets/MCRFileNodeServlet/duepublico_derivate_00047187/Hempel_et_al_Fremdsprachenunterricht_in_der_Grundschule.pdf. 02.05.2021].

Jaskorsky, Esther (2019), *The Assessment of Spelling in the EFL Classroom*. (Unpublished thesis, first state exam for the qualification

and authorization to teach at primary school). Heidelberg School of Education, Germany.

Kahn-Horwitz, Janina (2015), Organizing the mess in my mind: EFL teachers' perceptions and knowledge of English orthography. *Reading and Writing* 28: 5, 611–631.

Lee, Icy (2003), L2 teachers' perspectives, practices and problems regarding error feedback. *Assessing Writing* 8: 3, 216–237.

Maas, Utz/ Mehlem, Ulrich (2003), *Schriftkulturelle Ressourcen und Barrieren bei marokkanischen Kindern in Deutschland: Abschlußbericht des von der Stiftung Volkswagenwerk 1999–2002 am Institut für Migrationsforschung und Interkulturelle Studien (IMIS) der Universität Osnabrück geförderten Forschungsprojekts*. Osnabrück: IMIS Universität Osnabrück.

(MKJS BW) Ministerium für Kultus, Jugend und Sport des Landes Baden-Württemberg (eds.) (2020), *Bildungsplan 2016: Grundschule*. [Online: http://www.bildungsplaene-bw.de/,Lde/LS/BP2016BW/ALLG/GS/E34. 02.05.2021].

(MSB NRW) Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen (eds.) (2020), *MASTERPLAN GRUNDSCHULE: Qualität stärken-Lehrkräfte unterstützen*. [Online: https://www.schulministerium.nrw/system/files/media/document/file/Masterplan%20Grundschule.pdf. 01.05.2021].

(MSW NRW) Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (eds.) (2008), *Richtlinien und Lehrpläne für die Grundschule in Nordrhein-Westfalen: Deutsch, Sachunterricht, Mathematik, Englisch, Musik, Kunst, Sport, Evangelische Religionslehre, Katholische Religionslehre* (1st edn.). Frechen: Ritterbach.

Müller, Natascha/ Kupisch, Tanja/ Schmitz, Katrin & Cantone, Katja (2006), *Einführung in die Mehrsprachigkeitsforschung: Deutsch, Französisch, Italienisch*. Tübingen: Narr.

Nikolov, Marianne (2017), Students' and teachers' feedback on diagnostic tests for young EFL learners: Implications for classrooms. In: García Mayo, María del Pilar (ed.) (2017), *Learning Foreign Languages in Primary School: Research Insights*. Bristol, Blue Ridge Summit: Multilingual Matters, 249–266.

Nikolov, Marianne (2001), A study of unsuccessful language learners. In: Dörnyei, Zoltán & Schmidt, Richard (eds.) (2001), *Motivation and*

Second Language Acquisition. Honolulu: University of Hawaii Press, 149–170.

Pfingsthorn, Joanna (2013), *Variability in Learner Errors as a Reflection of the CLT Paradigm Shift*. Frankfurt am Main: Peter Lang GmbH.

QUA-LiS NRW (2021), Lehrplan für die Primarstufe in Nordrhein-Westfalen Fach Englisch. Auszug aus Heft 2012 der Schriftenreihe „Schule in NRW", Sammelband: Lehrpläne Primarstufe, RdErl. d. Ministeriums für Schule und Bildung v. 01.07.2021 – 526-6.08.01.13-150096 [Online: https://www.schulentwicklung.nrw.de/lehrplaene/lehrplan/284/ps_lp_e_einzeldatei_2021_08_02.pdf. 25.08.2021].

Rymarczyk, Jutta (2011), „Lautes Lesen = mangelhaft / Leises Lesen = sehr gut" – Diskrepanzen in den Leseleistungen von Zweitklässlern im Fremdsprachenunterricht Englisch. In: Kötter, Markus & Rymarczyk, Jutta (eds.). *Fremdsprachenunterricht in der Grundschule: Forschungsergebnisse und Vorschläge zu seiner weiteren Entwicklung*. Frankfurt a. M. et al.: Peter Lang, 49–67.

Rymarczyk, Jutta (2018), Rolle und Professionalität der Fremdsprachenlehrpersonen. In: Burwitz-Melzer, Eva/ Riemer, Claudia & Schmelter, Lars (eds.) (2018), *Rolle und Professionalität von Fremdsprachenlehrpersonen: Arbeitspapiere der 38. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts*. Tübingen: Narr Francke Attempto, 152–163.

Schoormann, Matthias & Schlak, Torsten (2012), Sollte korrektives Feedback „maßgeschneidert" werden? Zur Berücksichtigung kontextueller und individueller Faktoren bei der mündlichen Fehlerkorrektur im Zweit-/Fremdsprachenunterricht. *Zeitschrift für interkulturellen Fremdsprachenunterricht* 17: 2, 172–190.

Teaching & Learning Team, Cambridge International (2019), *Assessment for Learning*. [Online: https://www.cambridgeinternational.org/Images/271179-assessment-for-learning.pdf. 02.05.2021].

Weth, Constanze (2008), *Mehrsprachige Schriftpraktiken in Frankreich: Eine ethnographische und linguistische Untersuchung zum Umgang mehrsprachiger Grundschüler mit Schrift*. Stuttgart: Ibidem-Verlag.

Wiliam, Dylan (2018), *Embedded Formative Assessment* (2nd edn.). Bloomington: Solution Tree Press.

Stefanie Frisch, Carsten Breul, Bärbel Diehr,
Claudia Kastens, and Annette Becker

Developing and assessing reading comprehension in primary learners of English as a foreign language

Abstract Teaching English as a foreign language has become a standard component of primary education worldwide. Although expectations for reading skills are tentatively mentioned in many syllabi, there are no research-based standards defining the reading competences of primary school leavers. The present paper reports on a study investigating reading comprehension among 4th-grade English as a Foreign Language (EFL) learners (N = 115) in primary school in Germany. The paper begins by sketching the role and function of reading in the German primary school. A model of reading is then expounded that is intended to support teachers in developing reading comprehension in primary school classes. The model furthermore serves as the basis for a new test. Designing and piloting the test is an integral part of a project aimed at improving future EFL teachers' competences in fostering reading skills. This project and the test items are described, followed by a presentation of the data analysis and the results, which are cautiously interpreted as evidence suggesting that primary school EFL readers can be expected to cope with unfamiliar texts, even where these may be rather challenging in terms of lexis and syntax.

1. The role and the function of reading in the German primary school EFL class

With the spread of English as a *lingua franca,* the expansion of teaching English as a foreign language (TEFL) has led to a worldwide increase in the numbers of young children learning English in kindergarten, pre-school, and primary school. While early foreign language learning had, until the end of the 20th century, constituted an optional feature of the German educational system, by 2005 TEFL had attained compulsory status in all primary schools nationwide. Owing to the degree of independence in the legislation and administration of educational and cultural affairs in all 16 federal states, schools in some states, like e.g. Rhineland Palatinate,

provide EFL classes from Year 1 onwards to learners aged 5–10, whereas schools in other states, such as Bavaria, offer EFL classes from Year 3 onwards to learners aged 7–10, or even to the age of 12 where primary schooling comprises 6 years, as in Berlin, instead of the standard 4 years.

Although research into TEFL in primary education has received wider attention over the last two decades, there are still several areas that have not received close scrutiny (e.g. the role of grammar, the role of written English, and the potential of content integrated language learning (CLIL)). As a result, viable primary school leaving standards for EFL have not yet been developed and practitioners are fairly unsure about the competences which learners at the end of primary schooling can be expected to have developed. This is particularly true of reading and writing skills. The teaching schemes outlined in early curricula and handbooks, as well as in empirical studies in early foreign language learning, were characterised by an almost exclusive orientation towards oral skills. A shift of emphasis both in research and in practice became noticeable approximately a decade ago, when expected outcomes for reading and writing entered the first syllabi of German federal states (e.g. BSB HH 2010; MBFJ RP 2004; MSW NRW 2008; TH KM 2001) and when the first assessment studies (BIG-Kreis 2015; Paulick/Groot-Wilken 2009) indicated that primary school learners can reasonably be expected to read in English.

Researchers, teachers, and politicians have begun to understand that reading fulfils important functions in primary school EFL learning (for an overview of studies cf. e.g. Diehr/Rymarczyk 2010 and Frisch 2013). Once learners become literate in their first language, they urge their teachers to also use written forms of the foreign language. Those who continue to be taught without exposure to written English have been observed to write down English words and sentences at their own initiative (Diehr 2010) and to use invented spelling (Rymarczyk/ Musall 2010).

For more than one and a half decades it has been generally accepted that instruction in EFL reading has a positive influence on children's overall language awareness (Reichart-Wallrabenstein 2004). Further positive impact by learning how to read in English as a foreign language has been discussed for the learners' mnemonic abilities (Duscha 2007) as well as for their word recognition processes (Frisch 2013). In the course of primary schooling, the written language has developed from an object of learning to the most

important medium of learning, so written English is used by young learners also to expand their general knowledge of facts about the world they live in. With digital media being increasingly accessed by children and youngsters, literacy in the foreign language opens up new ways of learning and boosts independent learning inside and outside school. In addition, being able to read in English gains admission for young learners to a fast-growing body of children's literature and young adult fiction (Matz/ Stieger 2015) that can be enlisted for use as classroom reading and individual reading for pleasure. Amongst young learners, especially, the sense of achievement they feel after completing a whole book in English has often been noticed (e.g. Diehr 2010) and empirically illustrated (Diehr/ Frisch 2012; Kolb 2012; Reckermann 2018).

In order to put the beneficial functions of reading into practice, primary school teachers are required to draw on both accepted theory and successful practices. Young learners should be given space for implicit learning, reading for pleasure, and discovering patterns; but they also need tasks and explicit teaching to become aware of specific features of written English (Diehr 2010; Frisch 2013; Jöckel 2015; Mayer/ Schick 2013). So far, reading competences for EFL learners in primary school have been only cautiously, or rather vaguely, described in the official syllabi. At the end of primary school, 4th-graders in North Rhine-Westphalia, for instance, are expected to be able to understand written sentences from their set book or workbook, from invitations and postcards, and to understand short texts, including authentic texts written for native speaker readers (MSW NRW 2008: 78). However, training sessions with experienced teachers have repeatedly demonstrated that practitioners remain unsure as to how to interpret the notion "short and authentic texts", what books to choose for their learners and how to support them to cope with material which may suit their cognitive development but which, being written for native speakers, poses too great a challenge language-wise.

When the University of Wuppertal hosted the first EFL reading competition for 328 primary school learners in Germany in 2009 (Diehr/ Frisch 2010, 2012[1]), teachers from 10 neighbouring schools participated in a

1 Cf. also https://www.anglistik.uni-wuppertal.de/de/fachdidaktik/evw/start-seite.html.

12-month in-house training programme before nominating their pupils prior to the finals of the reading competition. One of the most intensely debated questions revolved around the choice of reading material. Due to the lack of empirical evidence of what can count as appropriate material for the different age groups in primary school, teachers reported that they resorted mainly to picture books with little text written for children aged 2–4 because they felt that English aimed at toddlers would help them avoid the risk of overtaxing their German EFL learners aged 9–10. They also admitted that they had been unable to bridge the gap between their learners' intellectual demands and their foreign language competences[2]. The teachers' intuitive comments have been confirmed by the questionnaire study conducted among the 328 Year 4 pupils participating in the reading competition (Diehr/ Frisch 2010, 2012): 76 % state that they are interested in the reading material provided in their English language class, 60 % would even like to read more, and 72 % say that they would prefer to read different books as the ones offered are often "babyish" and "boring" (Diehr/ Frisch 2012: 16).

This discrepancy between the cognitive demands and the language demands of children's literature used in German primary school EFL classes, as well as the shortcomings in teacher education with a focus on literacy, together form the points of departure for a research initiative at the University of Wuppertal. The project, called EULE[3], is an integral part of a larger scheme to improve initial teacher education at university by rendering courses more coherent. The scheme is financed by the Federal Ministry of Education and Research (BMBF – *Bundesministerium für Bildung und Forschung*) and proposes to sustain excellence in teacher

2 One of the children's books discussed and analysed prior to the reading competition was *I'm not invited?* (Bluthenthal 2003) for which the reading comprehension test presented in this paper has been developed. In 2009, the teachers participating in the in-house training rejected this particular book, even though they considered its story, theme, and plot age-appropriate and motivating. However, they felt that the text was too long as well as lexically and syntactically too challenging.

3 EULE: *Entwicklung von Unterrichtskonzepten zum Lesen lernen im Englischunterricht der Grundschule*. German for 'Developing teaching programmes to foster literacy in the primary school EFL class'.

education for schools of the future[4]. EULE aims at developing knowledge, competences, and skills among future EFL teachers that will enable them to foster skills in reading and writing in primary school learners (for more details see Section 4 of this paper).

To evaluate the effectiveness of the teaching programme and material designed and used by the research team in cooperation with the student teachers involved in the project, a reading comprehension test is required in order to assess with accuracy the reading competences of the young learners participating in the scheme. Thus, the research aim of EULE is to investigate to what extent German EFL learners are able to read and understand an unfamiliar text which contains familiar vocabulary.

2. A model of reading comprehension

In testing reading comprehension, researchers are required to ensure that the test used meets the standard criteria, with validity, objectivity, reliability, and feasibility being the most significant ones. It seems that all existing tests that have been employed in previous studies on reading in EFL at primary school level in Germany pay scant attention to content and construct validity, or, at least, lack a sufficiently intelligible and transparent description thereof (see Section 3). To address the central requirement of validity and to ensure that the reading comprehension test designed for the EULE project does in fact measure what it claims to, namely reading comprehension in EFL, content and construct validity are targeted by designing a test consisting of the constituent components of the skill that is commonly designated as reading. The test is based on a theoretical foundation in the form of a model of reading that explains what reading is, what its sub-components are, and how these interact.

Addressing the peculiarities of reading in EFL, as distinct from the learners' first language (L1), Diehr and Frisch (2010) propose a model of reading which combines the principles of the Dual Route Theory (Coltheart 2005) with those of the mental model of reading (McNamara/ Miller/ Bransford 1991) and which incorporates core elements of the model used for L1 testing in IGLU (*Internationale Grundschul-Lese-Untersuchung*,

4 Kohärenz in der Lehrerbildung (KoLBi) (Förderkennzeichen 01JA1507)

the German equivalent to *Progress in International Reading Literacy Study* (PIRLS)) (Bos et al. 2009).

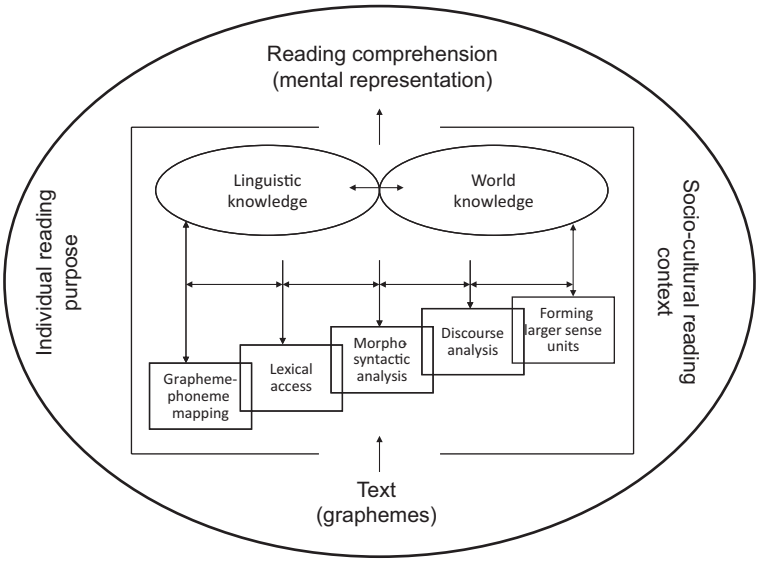


Fig. 1: Model of reading (adapted from Diehr/ Frisch 2010: 27)[5].

Reading is generally understood as the cognitive ability to construct meaning from written language. The model (cf. Fig. 1) emphasises the notion that the reading process occurs in a sociocultural context and is driven by the individual's interests and purposes, both influencing the success and the depth of the actively constructed mental representation of the text read. The mental picture, or representation, that emerges as the product of the reading process is seen as the reader's idea of a text's core meaning including – for narrative texts – events and situations. This process is initiated by visual stimuli, usually black lines forming different shapes against a white background. Most children on entering primary

5 This model can only visualize the complex phenomenon of the interaction between the model components in a rather simple representation (suggested by the bi-directional arrows). As for word recognition (here specified as 'lexical access'), for instance, this interaction has been elaborated in detail by the Cohort Model (Marslen-Wilson 1989).

education already have a vague idea that those lines and dots stand for letters. In the process of acquiring literacy they come to identify specific shapes as specific graphemes and match them, or specific sequences of them, to the phonemes they represent. Due to what is regarded as the deep, or rather irregular, orthography of the English language (cf. Seymour/Erskine 2003), German learners of EFL who simultaneously acquire literacy mainly in German, a language with a comparatively shallow, i.e. regular, orthography (cf. ibid.), are faced with the unfamiliar challenge of matching letters to sounds on the basis of an irregular grapheme-phoneme correspondence. Thus, a valid reading comprehension test for EFL ought to include items testing precisely this basic matching skill (see Section 4.3). Word recognition skills from grapheme-phoneme mapping to forming large sense units, and particularly lexical access (cf. Fig. 1), are targeted by almost all the existing EFL reading tests for young learners (see Section 3). Although children learning a foreign language may draw on their world knowledge, i.e. the conceptual store developed while acquiring their first language (L1), their ability to associate orthographic forms with concepts in their second language (L2) will be considerably weaker than in L1, thus making lexical access in L2 a crucial skill (cf. Pavlenko 2009: 147); it constitutes a necessary, yet insufficient basis in itself for comprehension at sentence level. Readers' competence in lexical access includes the ability to integrate semantic and morpho-syntactic information, for instance to distinguish between 'play' as noun and as verb, to identify markers of inflection and to apply knowledge of word order. Thus, lexical access merges with and often coincides with the morpho-syntactic analysis on phrases consisting of several words strung together. In order fully to understand a given text, readers also need to grasp its discursive and generic features above sentence level, i.e. to perform a discourse analysis drawing on their knowledge about how different types of texts (e.g. narrative, argumentative, or exploratory) serve different communicative purposes (e.g. entertaining, persuading, or informing). Finally, to be able to construct a mental representation of the entire text at hand, readers form larger sense units concerning the main ideas or main events and outcomes.

The reading process is characterised by all these components in constant interaction with each other. Depending on their interests, abilities, and tasks, readers sometimes focus more on data driven, so-called bottom-up

processing, as in grapheme-phoneme matching, at other times they engage more in concept driven, so-called top-down processing, as in inferencing a fictional character's motives from her verbal utterances and drawing on experiences not directly related to the text but to their general world knowledge. In EFL lessons, the components as well as their interaction are sometimes dealt with separately by focusing on selected skills, e.g. spelling with one children's book and identifying the text type with another. The main goal of using children's literature in primary school, however, lies in holistically encouraging young learners to read in a foreign language, in boosting their sense of achievement, and in gradually increasing their language awareness.

Unlike the teacher, researchers use the model of reading to assess the variety of different sub-skills and therefore have to ensure that in a reading comprehension test the components mentioned above are represented by appropriate items which allow them to form a multifaceted and more nuanced picture of young learners' reading competences. The test designed for the EULE project is based on the four components of phoneme-grapheme mapping, lexical access, morpho-syntactic analysis, and forming larger sense units. The component of discourse analysis was omitted from the test since discursive features are rarely targeted explicitly in primary school English classes and, besides, the learners were only expected to read a narrative text. The four remaining components are then summarized in the two dimensions of linguistic abilities and textual abilities, i.e. reading comprehension at the level of words and single sentences, and at the level of meaning making above sentence level progressing towards a mental representation of the entire text.

3. Tests available for assessing young EFL learners' reading comprehension

For our research we analysed three different reading tests for young L2 learners: A1 Movers (Cambridge Assessment English 2018); EVENING test (Paulick/ Groot-Wilken 2009); Reading Test for Polish Young Learners (PYL) (Szpotowicz/ Campfield 2016), and three reading tests for native speakers: Primary School Assessment Kit (PSAK) (Ireland: Little et al. 2006); Tests of Reading Comprehension (TORCH) (Australia: Mossenson

et al. 2003); Ein Leseverständnistest für Erst- bis Sechstklässler (ELFE) (Germany: Lenhard/ Schneider 2006). We focused on four aspects in our analysis: the theoretical conceptualisation of reading comprehension, the purpose of the reading test, the choice of texts, and the choice of task types. In the test manuals – where available –, a definition of reading comprehension is missing. Therefore, the informed reader has to infer which sub-skills are actually assessed. Of the above, the ELFE reading test stands alone as a standardised test instrument which reports values for validity and reliability.

The purpose of these tests is to identify the learners' reading level in order to provide individual feedback for the learners, teachers, and parents (A1 Movers, ELFE, PSAK, TORCH) or to evaluate the quality of a teaching programme based on the learners' outcome (EVENING, PYL). The individual results can be referred to reading levels which are defined either by descriptors set out in the Common European Framework of Reference for Languages (Level A1, 2001), proficiency benchmarks (e.g. PSAK, ELFE), or by national documents such as syllabi (e.g. MSW NRW 2008).

Our analysis of the six different reading tests for young learners also reveals that lexical access skills and sentence comprehension are targeted by most of them, namely five. Lower level skills such as the ability to match auditorily presented words with their orthographic forms and the ability to apply a morphological analysis are not assessed. Only three tests (ELFE, PYL, TORCH) assess the ability to form larger sense units with the help of gap filling activities and questions with multiple-choice answers. The ELFE reading test also uses reading fluency as an indicator of reading comprehension.

For EULE, a different testing approach is necessary as the aim is to gain insights into the learners' ability to understand the content of a specific story. The results are used to discuss the appropriateness of this reading text for EFL 4th-graders. The research team therefore developed a new reading comprehension test with some task types taken over from previous tests (like e.g. picture-word-matching tasks for assessing lexical access) but with new task types for assessing additional components of reading comprehension (see Section 4.2).

4. Development of a reading comprehension test in the context of the project EULE

In what follows, the choice of the book, the development and conducting of the reading test and the results are presented in turn.

4.1 *Choice, adaptation, and analysis of the text*

The reading comprehension test presented in this paper is based on the illustrated children's book *I'm not invited?* (Bluthenthal 2003). The plot revolves around a young girl named Minnie who believes that she has not been invited to her friend Charles's birthday party. The story has a simple structure moving from Tuesday to Saturday, covering a week that leaves the protagonist extremely upset, until she learns that the party in question is not Charles's but in fact his sister's, which comes as a huge relief to her after all. It can be easily related to the typical topics of the primary EFL classroom: friendship and loyalty, school, and free time activities. The images support the text but leave enough gaps for the reader to have to depend on the verbal narrative for a full understanding of the story. Content-wise there are compelling arguments for using this book in Year 4. However, EFL learners cannot be expected to read the whole story on their own, as the following linguistic analysis suggests.

The analysis is conducted on a version of *I'm not invited?* that is shortened but not otherwise simplified, in the part that covers the Tuesday to Friday episodes of the narrative[6]. The part of the narrative covering the Saturday episode has been left unchanged. The results of the analysis were taken as a basis for developing teaching materials and support for introducing and practicing vocabulary items and sentence structures that need an explicit focus.

The shortened version of *I'm not invited?* comprises 843 grammatical word tokens and 297 lemmas[7]. This count of lemmas includes the

6 This part, in the shortened version, was used by teachers to prepare their 4th-graders for the reading comprehension test. The unmodified part that covers the Saturday episode of the narrative is the text that the pupils were to read during the test. See also Section 4.3 below.

7 (a) For the collection of these data, the following tools were used: Microsoft Excel, TextAnalysisOnline (http://textanalysisonline.com; last access 2021/07/

following lexical distinctions – and thus goes some way in the direction of a lexeme count:

- *all* as a determiner vs. predeterminer ("Minnie didn't see Charles at school all day" vs. "all the kids divided up into teams");
- *his* as a pronominal determiner vs. pronoun ("… when she heard his voice" vs. "She knew his was in June, like hers");
- *like* as a verb vs. preposition ("I thought you'd like that" vs. "She knew his was in June, like hers");
- *long* as an adjective vs. adverb ("into the long, metal tunnel" vs. "All week long");
- *missing* as an adjective vs. verb ("Minnie worried about her missing invitation" vs. "I'm not the only one missing the party");
- *smile* as a verb vs. noun ("Minnie smiled big" vs. "Minnie's smile");
- *what* as a pronoun vs. determiner ("What are you doing here" vs. "What time is the party");
- *that* as a subordinator vs. determiner vs. pronoun ("I heard that you're …" vs. "That afternoon, she raced home to check the mailbox" vs. "I thought you'd like that");
- *there* as a locative adverb vs. the item that is commonly considered to perform the syntactic function of subject in so-called existential *there*-sentences ("while you're out there" vs. "Now there's something to be sad about").

The lemma count also distinguishes between *'s* as possessive marker vs. third person singular marker ("Minnie's ears" vs. "It's not his birthday on Saturday"). Note that there are occurrences of the lexeme *a* realised by *a* and *an*, of the lexeme *would* realised by *'d* and *would*, of the lexeme *will* realised by *'ll* and *will*, of the lexeme *not* realised by *not* and and *n't*, and of the lexeme *be* realised, among others, by *'m*, *'re,* and *'s*. These realisations of a lexeme are not distinguished for the lemma count, i.e. they

28) and AntConc (http://www.laurenceanthony.net/software.html; last access 2021/07/28).

(b) Expressions such as *I'm*, *she's* or *Minnie's* count as one orthographic word, but two grammatical words.

are subsumed under the same lemma, respectively. Also subsumed under the same lemma are different lexemes associated with *to*.

The cases of homonymy and of variable orthographic realisation of the same lexeme or of the same grammatical word just pointed out raise some potential challenges for 4th-grade EFL learners with German as their first language. The same holds for the fact that the text contains a number of lexemes that these learners are unlikely to know without specific preparation (such as, for example, *cancel*, *garlic*, *moan*, *soggy*, *squeal*, *whimper*). In addition, the text shows the following lexical characteristics that may pose difficulties:

- Phrasal and prepositional verbs: *perk up*; *start on*; *worry about*; *slump down*; *tuck in*; *lean in*. These may pose difficulties as the learner has to grasp that the particle or preposition is required by the verb in order to express a meaning that may be, or has to be, expressed in a structurally different or lexically unexpected way in German (e.g. "to tuck her in" ≈ 'um sie zuzudecken'; "Minnie's mother leaned in close" ≈ 'Minnies Mutter kuschelte sich an sie').
- Orthographically separated compounds: *bus stop*; *overdue notice*; *science class*; *garlic bread*. The beginning reader has to understand that, in English, words separated by a space may form a complex word whose components are as closely connected as the components of a German compound, which usually are not separated by a space (e.g. *bus stop* vs. *Bushaltestelle*).
- Idiomatic and/or strongly collocating expressions: *never mind*; *cat's got your smile*; *now there's something to*; *long face*; *no way*; *'d rather*. Here again, the reader has to grasp these expressions as units which may correspond to a (part of a) structurally different or lexically unexpected construction. Even a case like *long face*, where the idiomatic meaning is metaphor- and metonymy-based and where German has a meaning equivalent that is structurally and lexically equivalent, too (*langes Gesicht*), may not be completely unproblematic. Young learners may not be familiar with this German expression, and even if they are, they cannot be assumed to suspect that English is analogous to German in this particular case.

As for the verb-related grammatical categories of tense, aspect and modality, the non-modal, non-progressive past (e.g. "'Charles!' she squealed when she heard his voice") is the most frequent combination by far. The past tense is the tense used by the narrator, and it occurs in the quoted passages of character speech or thought (character discourse) as well. The non-modal, non-progressive present ("I know how you feel") also often appears in character discourse. There are occasional occurrences of the non-modal, progressive present and past. These are: "Charles is having a party"; "you're … [g]rowing worms in science class"; "some of us are playing football"; "What are you doing here?" (present; all in character discourse); "if her invitation was waiting"; "the phone was ringing"; "what games they were playing" (past; all in narrator discourse). The text also contains quite a range of modal constructions, including those serving to express future time: "I'll call you back"; "It will come tomorrow"; "you'd like that"; "he could invite six guests"; "I don't have to be invited to everything"; "her mother would protest"; "Then something else must have happened at school today"; "it might be Charles"; "some dirt could do you good"; "And will you please try to lose that long face"; "I'll try"; "I had to escape"; "I'd rather be here any day"; "you were … [g]oing to be at my sister's party". Except for the clause involving *would protest*, all of the foregoing come from direct speech or thought. The occurrence of *have happened* in the clause "Then something else must have happened at school today" is the only one involving a perfect, here non-finite.

As concerns clause-internal constructions that may pose difficulties for 4th-grade EFL learners with German as their first language, the following ones ought to be pointed out.

- Non-finite, i.e. infinitive as well as present and past participle clauses (only the non-finite verb that projects the non-finite clause is underlined in the following list): "She ran all the way home from the bus stop to see if her invitation was waiting"; "'Hmm,' Minnie said, staring into the long, metal tunnel"; "my list of things to bring to school"; "She set the table for supper, counting the days until Saturday"; "Minnie's father looked around, puzzled"; "She forced herself to eat breakfast, knowing her mother would protest if she didn't"; "But as she sat watching her

cereal turn soggy in the milk, …"; "something to be sad about". The participle clauses in particular commonly lack a structurally similar German counterpart that is also appropriate in terms of register (that is, it may, for example, be possible to translate the clause "counting the days until Saturday" from the text by 'die Tage bis Samstag zählend', but it would not be appropriate to do so if the translation were to maintain the colloquial register of the original as well).

- Passive and middle constructions (ignoring the question whether and how to distinguish between a passive construction and a predicative adjectival one with an adjective derived from a past participle): "I'm not invited"; "Maybe the party was cancelled"; "I don't have to be invited to everything"; "all the kids divided up into teams". The last-mentioned instance of a middle construction is the only occurrence of that construction type in the text[8].
- Existential and/or presentational constructions: "But there was nothing in the mailbox for her"; "Now there's something to be sad about"; "there was a rustling in the bushes"; "It's not his birthday on Saturday"; "It's not that"; "It's what didn't happen"; "It's not my party".
- Ellipsis: As long as the particular type of ellipsis functions similarly in English and German, it ought not to raise any specific difficulties, as, presumably, in cases like the following: "'What time is the party?' Kathleen said to Charles as they whizzed home on bicycles after school Tuesday. 'One o'clock, Saturday.' Charles replied"; "'Me too!' she said."[9] Potentially more problematic are elliptical constructions of the following kinds: "'Are you too?' he asked"; "Charles meant to invite me, but forgot to"; "She forced herself to eat breakfast, knowing her mother would protest if she didn't".

8 The reason for mentioning passives and middles together here is that they are semantically similar in that their subjects carry the same semantic role that a direct object carries in a corresponding active or non-middle clause (e.g. *the party was cancelled ~ somebody cancelled the party*; *the kids divided up into teams ~ somebody divided the kids up into teams*).

9 However, the last example raises the difficulty that the pronoun appears in the objective (accusative) case, while an equivalent German utterance would require it to be in the subjective (nominative) case (*Ich auch!*). This genuine difference between English and German is triggered by ellipsis here, but is not specific to it.

There is one instance of polarity-focusing *do*-support in a non-negated declarative clause in the text ("Minnie did receive a note from Charles") and one instance of a serial verb construction ("Minnie kept thinking").

As far as clause types are concerned, in addition to what has been said above in connection with non-finite constructions, we may point out that the large majority of finite main clauses are declarative. But imperatives (" 'Think *positive*,' she told herself"; "Never mind") and interrogatives of the *wh*- and polarity types do occur, too ("What time is the party?"; "What are you doing here?"; "Why aren't you at your party?"; "Are you too?"; "Do you have one?"; "Did it go to the wrong house?"; "And will you please try to lose that long face while you're out there?"). We also find examples of sentences that are syntactically declarative but functionally – and intonationally if imagined as spoken utterances – questions: "I'm not invited?" (the title of the book); "Cat's got your smile again tonight, Min?"; " 'All week long I thought that you were …' - 'Going to be at my sister's party?' said Charles". In the domain of (finite) subordinate clauses, the reader comes across complement and adverbial clauses of various kinds, such as, for example: "She ran all the way home from the bus stop to see if her invitation was waiting"; " 'Charles!' she squealed when she heard his voice"; "I thought you'd like that"; "Charles meant to invite me, but forgot to, and then forgot that he forgot"; " 'And that's terrible,' she moaned, 'because how do you ask someone if he meant to invite you?' "; "Minnie hoped it might be Charles, wanting to know why she wasn't at the party". The last three examples represent the upper limit of complexity for clause combining in our adapted version of *I'm not invited?* – very probably quite demanding for 4th-grade EFL learners who are beginning to read texts in English.

4.2 *Development of items for assessing reading comprehension*

The development of the test items is based on the reading model proposed by Diehr/ Frisch (2010), adapted as described in Section 2, targeting abilities in the model components 'grapheme-phoneme relation', 'lexical access', 'morpho-syntactic analysis', and 'forming larger sense units'. All items were developed on the basis of the Saturday episode. We consider abilities associated with the first three components to be linguistic abilities

and those associated with the fourth component to be textual abilities. Since correct solutions of the items targeting linguistic abilities do not actually presuppose familiarity with the text, coping may have been grounded in the pupils' ability to activate previously (i.e. before their reading of the text) established grapho-phonological, lexical, and grammatical competences, or in their ability to generate and activate competences like this during their engagement with the text.

Phoneme-grapheme mapping was examined using listening comprehension tasks in which the learners are required to match the oral form of a word with its written form. Six monosyllabic words were read to the pupils, who were expected to identify their written form from a group of four English words including three distractors all similar in syllable structure but differing in at least one phoneme and the corresponding graphemes. For example, for 'some':

(1) swam / song / same / some

Lexical access was tested at both, the orthographic and conceptual level. For the orthographic level, the pupils were expected to identify the correct spelling of six words. Distractors were pseudo-words in invented spelling close to potential L2 errors. For example:

(2) Minnie rode her bik, bike, beik, baik to Charles's house.

For the conceptual level, the pupils were shown six pictures and asked to identify the matching written words from the story. Distractors were either orthographically and phonetically similar, as in (3), or semantically related, as in (4).

(3) 

☐ face

☐ foot

☐ fan

☐ food

(4) 

☐ chair
☐ table
☐ bed
☐ sofa

Morpho-syntactic competence was tested by six items aiming at the recognition of inflected verbs and auxiliaries as well as of their appropriate syntactic integration, as in (5) and (6).

(5) Minnie hoped, hope, hoping, helped it might be Charles.
(6) What is, are, has, am you doing here?

This competence was also assessed by the ability to identify the correct German translation of an English sentence. The distracting German sentences differed in at least two words from the correct German sentence which the readers have to process in order to find the correct translation (7).

(7) In the kitchen, the phone was ringing.
- ☐ In der Küche stand das Telefon.
- ☐ In der Küche war ein Telefon.
- ☐ In der Küche klingelte ein Telefon.
- ☐ In der Küche klingelte das Telefon.

The pupils' competence to **form larger sense units** (textual abilities) was assessed both in English and in German. The learners were required to choose a correct expression in an English sentence for which information from different sentences needs to be integrated and by identifying the correct answer to a question in German. It is often claimed that learners are not able to show their receptive competences if the questions are formulated in English (see Karges/Malgorzata/Lenz in this volume). For this reason, the items at text level consisted of sentences from the story in English, as in (8), and of German questions and answers, as in (9). The formulation of the German questions and multiple-choice answers ensured that considering the pictures does not guide test takers to a correct solution.

(8) "It's not my party", said Charles, Minnie, Minnie's mother, Charles's sister.

(9) Warum ist Minnie bis Samstagmorgen traurig?
- ☐ Sie hat kein gutes Ergebnis im Vokabeltest.
- ☐ Sie wurde nicht zum Fußballspielen eingeladen.
- ☐ Sie isst nicht gerne Nudeln.
- ☐ Sie denkt, dass sie nicht zu Charles' Party eingeladen wurde.

4.3 *Administration of the test*

During their regular English lessons at school, the pupils read the first part of this story ending with the Friday episode. All English teachers had been previously provided with suggestions for a scaffolded lesson series (eight units) along with corresponding teaching materials developed within the context of the EULE project, such as picture cards, word cards and various creative tasks. On the day of the reading test, the pupils were given the last episode (Saturday) in the original leading up to the conclusion of the story. The episode consists of 237 words and eight pictures. The text was read silently by each learner on their own (10 minutes), forming the basis of the EULE reading comprehension test. The test was administered by trained testers. Based on a standardised test manual, task instructions were read aloud, example questions completed and further explanations given. Then, each learner had enough time (30 minutes) to work on the test. Our sample includes 115 pupils, of which 61 (53 %) were girls. Eight children had been diagnosed with special needs but did not receive (or need) any special treatment or material during the lessons.

4.4 *Methods and results*

As pointed out above, we developed a set of items to assess four of the components of the model of reading comprehension described in Section 2. In order to determine whether our test items meet more than mere face and content validity, we applied different methods commonly used to test the psychometric properties of a test (Moosbrugger/ Kelava 2012). In a first step we determined item difficulty values. Only items exhibiting difficulty values between .20 and .80 were used for further analysis. Out of the initial 40 items (29 targeting linguistic abilities, 11 targeting textual abilities) seven items proved too easy, indicated by difficulty values above .80, and three items were too difficult, indicated by difficulty values below .20.

With the remaining 30 items, we conducted exploratory factor analyses to test for construct validity. Neither exploratory factor analyses conducted on the whole set of items (divergent validity), nor on single model components (convergent validity), yielded acceptable results. We were not able to extract the a-priori postulated model components from our data. However, we were able to identify two global dimensions of reading competences matching linguistic abilities (represented by the components 'phoneme-grapheme mapping', 'lexical access', 'morpho-syntactic analysis') on the one hand and textual abilities (represented by the component 'forming larger sense units') on the other.

As an indicator of internal consistency (scale reliability) we calculated Cronbach's α for both dimensions (see Tab. 1). 16 of the remaining 30 items had to be eliminated due to item-total correlations below .20. The final scale representing linguistic abilities included nine items while textual abilities could be assessed by seven items. For both resulting scales internal consistencies are > .60, and can be considered acceptable, taking the number of items per scale into account (Cortina 1993).

Tab. 1: Internal consistencies and descriptives for both abilities

	Cronbach's α	*M*	*sd*
Linguistic abilities	.64	.51	.25
Textual abilities	.62	.61	.27

Note: *M* = mean frequency correct solutions, *sd* = standard deviation of mean

As seen in Table 1, on average pupils answered 51 % of the items assessing linguistic abilities correctly; performance for textual abilities was better, resulting in an average of 61 % of correct answers. The variation between pupils' abilities was rather high (see also Fig. 2). One child did not solve any of the items assessing linguistic abilities correctly, two children did not solve any items assessing textual abilities correctly, seven children solved all of the items assessing linguistic abilities correctly and 16 children solved all items assessing textual abilities correctly. Eight children solved only one of the items assessing linguistic abilities correctly and ten children solved only one item assessing textual abilities correctly. The interquartile range is higher for textual abilities than for linguistic abilities. The median for

linguistic abilities was .44, and that for textual abilities .71. This means that 50 % of the participants were able to solve 44 % or more of the items assessing linguistic abilities, whereas 50 % of the participants were able to solve 71 % or more of the items assessing textual abilities. For the linguistic abilities we found no effect for the pupils with special needs, but for the textual abilities these pupils performed significantly lower (M = .63, sd = .27 for the group without special needs; M = .35, sd = .28 for the group with special needs; $F(1,112) = 3.08$, p = .01).

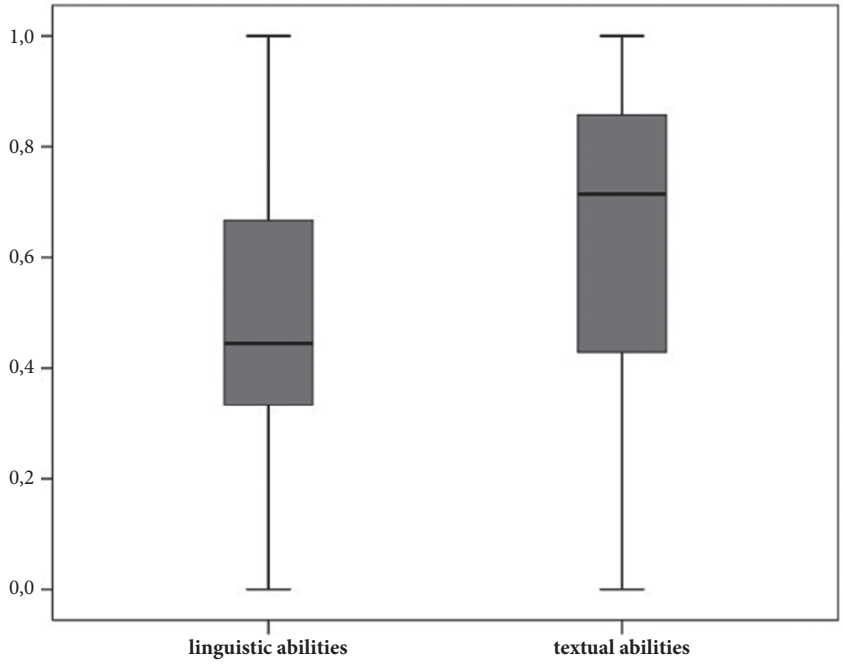


Fig. 2: Boxplot for both dimensions of reading abilities.

As expected, girls performed better than boys (cf., e.g., Hartig/ Jude 2008; BIG-Kreis 2015: 14; Paulick/ Groot-Wilken 2009: 190) within both dimensions, but these differences only reached significance for the textual abilities (see Tab. 2).

Tab. 2: Multivariate analysis (F-Test) of variance for sex differences

	Boys		Girls			
	M	*sd*	*M*	*sd*	$F_{(1,105)}$	eta^2
Linguistic abilities	.47	.26	.54	.26	2.41	.02
Textual abilities	.52	.26	.69	.23	11.00*	.09

Note: *M* = mean frequency correct solutions, *sd* = standard deviation of mean. *$p < .05$

The relationship between performance in the linguistic abilities dimension and in the textual abilities dimension is moderate, $r = .45$: Low as well as high linguistic abilities occur in conjunction with both, low as well as high textual abilities. Pupils, who solved 67 % or more of the items assessing textual abilities correctly, generated from 22 % to 100 % correct solutions for linguistic abilities.

Interestingly, the correlation between the performance in each of the two dimensions is stronger for the girls ($r = .53$) and considerably lower for the boys ($r = .35$). Not only do the boys show lower competence levels on both abilities, but their level of textual abilities is apparently less dependent on their linguistic abilities (see Fig. 3). They display more cases of low linguistic abilities (< 50 % of correct solutions), irrespective of how they perform on textual abilities (see lower and upper left field of the scatter plot). By contrast, girls tend to picture a more coherent distribution, accounting for most cases within the lower left and upper right fields of the scatter plot (see. Fig. 4).

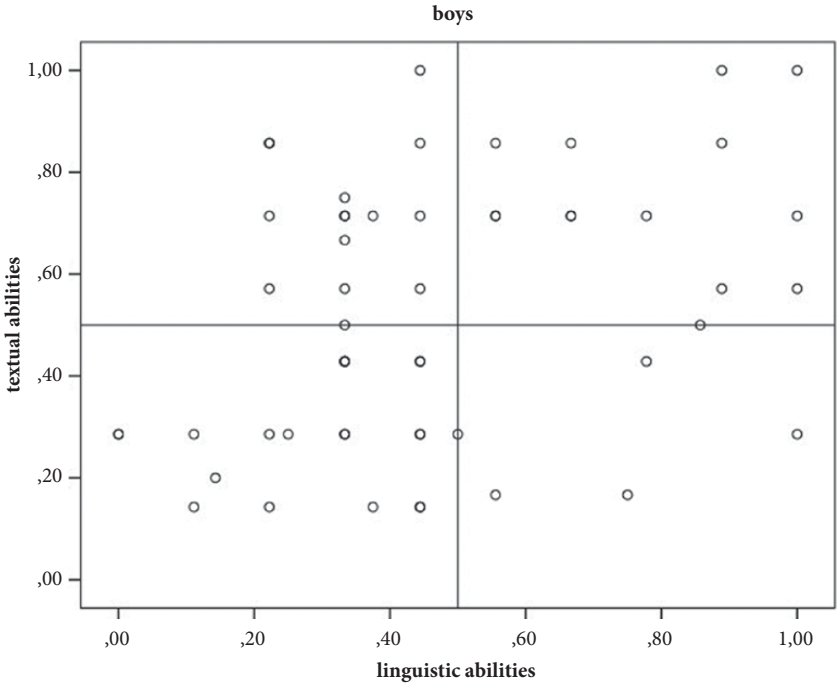


Fig. 3: Scatter plot for both dimensions of reading abilities (displayed as percentage of correct solutions for both dimensions) for boys. (One dot might represent more than one case, indicated by thicker lines.)

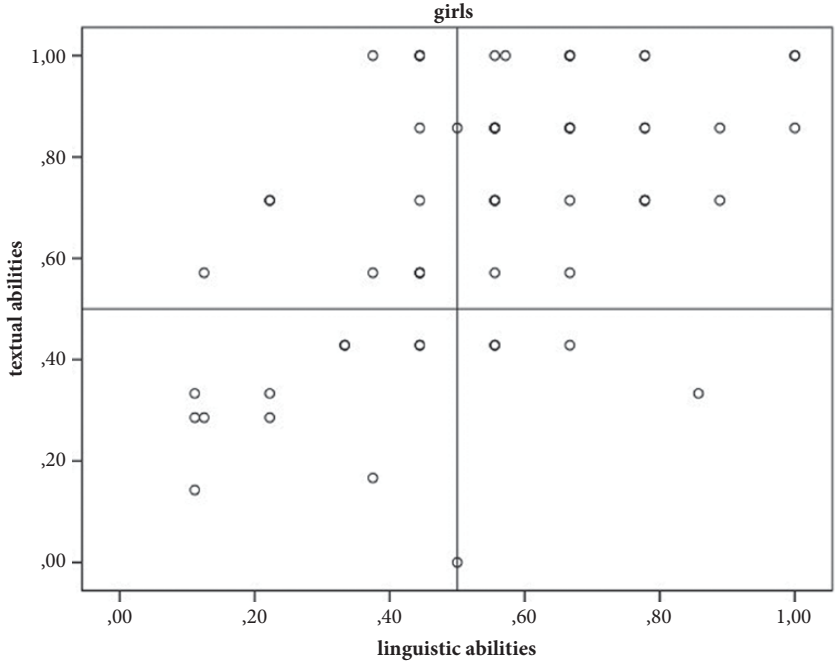


Fig. 4: Scatter plot for both dimensions of reading abilities (displayed as percentage of correct solutions for both dimensions) for girls. (One dot might represent more than one case, indicated by thicker lines.)

For 79 of 114 children[10] a stronger performance (≥ 50 % correct solutions) with higher order reading skills, i.e. textual abilities, tends to match a stronger performance with lower order reading skills, i.e. linguistic abilities. A weaker performance (< 50 % correct solutions) with higher order reading skills, i.e. textual abilities, tends to match a weaker performance with lower order reading skills, i.e. linguistic abilities. Fewer children report low scores for the textual abilities and high scores for the linguistic abilities (lower right field of the scatter plots; $n = 8$) as well as high scores for textual abilities and low scores for linguistic abilities (upper left field of the scatter plots; $n = 27$). This means that for the majority of learners a correlation of textual and linguistic abilities could be found.

10 For one child no values could be generated for textual abilities, due to non-respondence of most of the equivalent items. This case was therefore deleted from the respective analyses.

5. Discussion and interpretation of the results

With the EULE reading comprehension test, our aim was to provide an instrument to assess pupils' reading comprehension within a natural setting. What was to be assessed, more specifically, was the learning outcome of a series of actual teaching lessons in Year 4 EFL classes, preparing the learners for the challenging assignment of reading on their own an unfamiliar extract from an original L2 children's book which is, content-wise, appropriate to their age. In contrast to prior tests developed to assess EFL reading skills, we developed a test based on a model that makes explicit the sub-competences involved in EFL reading. Thus, it is not only learners' reading comprehension skills that are put to the test but also the model itself.

Despite our emphasis on the need for theory-based test development, we were not fully able to empirically confirm the theoretical assumptions of our test. Specifically, we were not able to distinctively assess each of the four components of the model that we focussed on: grapheme-phoneme relation; lexical access; morpho-syntactic analysis; forming larger sense units. However, we succeeded in finding test theoretical support for the assumption that competences associated with the first three model components constitute a rather coherent dimension, termed *linguistic abilities*, distinct from a *textual abilities* dimension concerned with the competence of forming larger sense units. The question why it was not possible to find test theoretical support for the distinctive existence of each of the model components within the linguistic abilities dimension, cannot be conclusively discussed in the present context. We must restrict ourselves to mentioning that this may be due, in principle, to the small sample size and/or sample composition, flawed item development or flawed theoretical assumptions. One aspect of the problem certainly lies in the difficulty of identifying distinct components of an integrated competence (reading comprehension) that constantly interact in the process (the reading process) in which this competence is employed. It may thus be the case that our results match the Gestalt principle that the whole is greater than the sum of its parts. Since previous EFL reading comprehension tests lack information on test validity and reliability, it is not possible to address this problem by comparing them with the EULE test.

The readers in our sample showed better results in textual abilities than in linguistic abilities. This can be explained by considering the preparation during the teaching unit. The learners received no explicit support in mapping graphemes to phonemes and detecting regularities in the English orthography or differences between the German and English orthography, nor did the teacher raise the learners' awareness of grammatical phenomena. The recognition of, among other phenomena, correctly inflected verbs and auxiliaries as well as of their appropriate syntactic integration is therefore challenging for young EFL learners. The results described above raise the question of whether the implementation of language awareness activities might have increased the learners' general comprehension of the text given the requisite role of linguistic abilities in order to fully comprehend a text. More research in this field is needed (see, e.g., Mindt 2006 for arguments in favour of an age-appropriate approach to raising language awareness).

It was not surprising that the female readers showed better results than the male readers in the EULE test. Girls have been repeatedly reported as showing more advanced abilities in foreign language related abilities, especially reading skills, than boys in former large-scale studies (e.g. Hartig/ Jude 2008; BIG-Kreis 2015: 14; Paulick/ Groot-Wilken 2009: 190). The reasons for these sex differences have been widely discussed. On the one hand, studies can be found which reveal correlations between test characteristics and sex preferences (e.g. Lafontaine/ Monseur 2009). "The gap in reading proficiency between males and females is larger for open-ended than for multiple-choice items" (ibid.: 76). As the EULE test is based on multiple-choice items, it is rather unlikely that the format of the test items caused the difference between the two sexes. On the other hand, studies highlight that focusing on the test results is not enough to account for learners' abilities. It is recommended that classroom interaction together with teachers' beliefs and values be taken into consideration for interpreting test results (e.g. Hartig/ Jude 2008: 206). Teachers tend to choose texts and methods that girls are more interested in than boys. Although the protagonist in the chosen story is a girl, we did not assume that boys would be less interested. But as we did not conduct any classroom observation or interviews with the learners, we cannot rule out the

possibility that girls were more motivated during the teaching unit than boys and that this might also have had an effect on their test performance.

The question of whether the children's book *I'm not invited?* is accessible for EFL learners in Year 4 (age 10–11) can be answered positively. In the sample, we were able to find three pupils who answered all items correctly, 36 participants (31 %) solved at least 70 % of the items which required linguistic abilities, and 58 participants (50 %) solved at least 70 % of the items which required textual abilities. If we assume that answering at least 70 % of all items in a reading test is an indicator of a satisfactory general achievement (see e.g. Paulick/ Groot-Wilken 2009; BIG 2015), we can conclude that half of the learners reached this aim and did not fail to read and understand the story. This is an astonishing finding as most curricula expect young learners to read and understand only short and simple texts (cf. Council of Europe 2001: 69). It is necessary to conduct more systematic linguistic analyses of age-appropriate children's books and to develop and evaluate innovative teaching units which provide support for reading comprehension in order to test further the hypothesis that young learners can understand texts that are more complex than is currently considered appropriate for them.

The high variation between learners' abilities highlights the importance of integrating methods of differentiation in the EFL classroom. Also worthy of consideration is whether it might be necessary to include graded readers in the primary EFL class – texts with the same content but differing in length and complexity. In this case again, a systematic linguistic analysis of English texts can make the complexity of texts visible, on the basis of which teachers could adapt the reading material to the learners' abilities and develop appropriate support material.

Bibliography

[BSB HH] Freie und Hansestadt Hamburg. Behörde für Schule und Berufsbildung (ed.) (2011), *Bildungsplan Grundschule: Englisch*. [Online: https://www.hamburg.de/contentblob/2481798/5162c9013d960305f5a2aa350cc9a84d/data/englisch-gs.pdf. 13.08.2021].

BIG-Kreis (ed.) (2015), *Der Lernstand im Englischunterricht am Ende von Klasse 4: Ergebnisse der BIG-Studie*. München: Domino Verlag.

Bluthental, Diana C. (2003), *I'm not Invited?* New York, London: Simon & Schuster.

Bos, Wilfried/ Hornberg Sabine/ Arnold, Karl-Heinz/ Faust, Gabriele/ Fried, Lilian/ Lankes, Eva-Maria Schwippert, Knut & Valtin, Renate (eds.) (2009). *IGLU 2006: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*. Münster: Waxmann.

Cambridge Assessment English (ed.) (2018), *A1 Movers. Cambridge English Qualifications.* [Online: https://www.cambridgeenglish.org/exams-and-tests/movers. 13.08.2021].

Coltheart, Max (2005), Modeling Reading: The Dual-Route Approach. In: Snowling, Margaret & Hulme, Charles (eds.) (2005), *The Science of Reading: A Handbook*. Malden: Blackwell Publishing, 6–23.

Cortina, Jose M. (1993), What is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology* 78, 98–104.

Council of Europe (eds.) (2001), *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Switzerland, Europe: Cambridge University Press.

Diehr, Bärbel (2010), Research into Reading in the Primary School: A Fresh Look at the Use of Written English with Young Learners of English as a Foreign Language. In: Diehr, Bärbel & Rymarczyk, Jutta (eds.) (2010), 51–68.

Diehr, Bärbel & Frisch, Stefanie (2010a), A Roadmap to Reading: Bewusstmachende Verfahren im Umgang mit der englischen Schriftsprache. *Grundschule* 42: 9, 26–28.

Diehr, Bärbel & Frisch, Stefanie (2010b), Lesen sie doch? – Fragen an die LiPs-Studie (Lesen im Englischunterricht auf der Primarstufe). *IMIS-Beiträge* 37, 143–163.

Diehr, Bärbel & Frisch, Stefanie (2012), Lesemotivation in der Fremdsprache – Erkenntnisse aus der Begleitforschung zum 1. Vorlesewettbewerb Englisch für Schülerinnen und Schüler der Klasse 4. In: Böttger, Heiner & Schlüter, Norbert (eds.) (2012), *3. FFF Konferenz. Fortschritte im frühen Fremdsprachenlernen: Ausgewählte Tagungsbeiträge Eichstätt 2011*. München: Domino Verlag, 10–19.

Diehr, Bärbel & Rymarczyk, Jutta (eds.) (2010), *Researching Literacy in a Foreign Language among Primary School Learners – Forschung zum*

Schrifterwerb in der Fremdsprache bei Grundschülern. Frankfurt am Main: Peter Lang Verlag.

Duscha, Michael (2007), *The Influence of Script on Foreign Language Learning in the Primary School: Represented by the Example of English Learning in Lower Saxony*. Braunschweig: Englisches Seminar.

Frisch, Stefanie (2013), *Lesen im Englischunterricht der Grundschule: Eine Vergleichsstudie zur Wirksamkeit zweier Lehrverfahren*. Tübingen: Narr.

Hartig, Johannes & Jude, Nina (2008), Sprachkompetenzen von Mädchen und Jungen. In: DESI-Konsortium (eds.) (2008), *Unterricht und Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie. Weinheim: Beltz,* 202–208.

Jöckel, Alicia (2015), Reading and Writing: Anbahnen der englischen Schrift. *Grundschulmagazin Englisch* 13: 4, 31–34.

Kolb, Annika (2012), Extensive Reading of Picturebooks in Primary EFL. In: Bland, Janice & Lütge, Christiane (eds.) (2012), *Children's Literature in Second Language Education*. London: Bloomsbury, 33–44.

Lafontaine, Dominique & Monseur, Christian (2009), Gender Gap in Comparative Studies of Reading Comprehension: To What Extent Do the Test Characteristics Make a Difference? *European Educational Research Journal* 8: 1, 69–79.

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

Little, David/ Lazenby Simpson, Barbara & Catibusic, Bronagh F. (2006), *Primary School Assessment Kit*. [Online: https://www.ncca.ie/media/2172/primary-school-assessment-kit.pdf. 14.07.2021].

Marslen-Wilson, William D. (1989), Access and Integration: Projecting Sound onto Meaning. In: Marslen-Wilson, William. D. (ed.) (1989), *Lexical Representation and Process*. Cambridge, MA: MIT Press, 3–24.

Matz, Frauke & Stieger, Anne (2015), Teaching Young Adult Fiction. In: Delanoy, Werner/ Eisenmann, Maria & Matz, Frauke (eds.) (2015), *Learning with Literature in the EFL Classroom*. Frankfurt am Main: Peter Lang, 121–140.

McNamara, Timothy/ Miller Diana L. & Bransford, John D. (1991), Mental Models and Reading Comprehension. In: Barr, Rebecca/ Kamil, Michael L. & Pearson, P. David (eds.) (1991), *Handbook of Reading*

Research: Volume 2. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc, 490–511.

Mayer, Andreas & Schick, Kim (2013), Chair, Fair, Pair! Kinder mit LRS unterstützend unterrichten. *Grundschulmagazin Englisch* 11: 5, 9–12.

Mindt, Dieter (2006), Von der Imitation zur bewussten Verwendung von Sprachmitteln: Ein neues Unterrichtsmodell. In: Schlüter, Norbert (ed.) (2006), *Fortschritte im frühen Fremdsprachenlernen: ausgewählte Tagungsbeiträge Weingarten 2004*. Berlin: Cornelsen, 68–74.

Moosbrugger, Helfried & Kelava, Augustin (2012), *Testtheorie und Fragebogenkonstruktion* (2nd edn.). Berlin, Heidelberg: Springer Berlin Heidelberg.

Mossenson, Leila/ Stephanou, Andrew/ Forster, Margaret/ Masters, Geofferey/ McGregor, Margaret/ Anderson, Prue & Hill, Peter (2006), *Tests of Reading Comprehension (TORCH)*. Melbourne: Acer Press.

[MBFJ RP] Ministerium für Bildung, Frauen und Jugend Rheinland-Pfalz (ed.) (2004), *Rahmenplan Grundschule: Teilrahmenplan Fremdsprachen*. [Online: https://grundschule.bildung-rp.de/fileadmin/user_upload/grundschule.bildung-rp.de/Downloads/Rahmenplan/Rapla-FS-Druckfassung.pdf. 13.08.2021].

[MSW NRW] Ministerium für Schule und Weiterbildung Nordrhein-Westfalen (ed.) (2008), *Richtlinien und Lehrpläne für die Grundschulen in Nordrhein-Westfalen: Deutsch, Sachunterricht, Mathematik, Englisch, Musik, Kunst, Sport, Evangelische Religionslehre, Katholische Religionslehre*. Frechen: Ritterbach.

Paulick, Christian & Groot-Wilken, Bernd (2009), Rezeptive Fähigkeiten und Fertigkeiten am Ende der 4. Klasse unter besonderer Berücksichtigung der sprachlichen Schülerbiografien. In: Engel, Gaby/ Groot-Wilken, Bernd & Thürmann, Eike (eds.) (2009), *Englisch in der Primarstufe – Chancen und Herausforderungen*. Berlin: Cornelsen, 179–196.

Pavlenko, Aneta (2009), Conceptual Representation in the Bilingual Lexicon and Second Language Vocabulary Learning. In: Pavlenko, Aneta (ed.), *The Bilingual Mental Lexicon: Interdisciplinary Approaches*. Bristol: Multilingual Matters, 125–159.

Reckermann, Julia (2018), *Reading Authentic English Picture Books in the Primary School EFL Classroom*. Berlin: Peter Lang.

Reichart-Wallrabenstein, Maike (2004), *Kinder und Schrift im Englischunterricht der Grundschule: Eine theorie- und empiriegeleitete Studie zur Diskussion um die Integration von Schriftlichkeit*. Berlin: Dissertation.de.

Rymarczyk, Jutta & Musall, Annika (2010), Reading Skills of First Graders who Learn to Read and Write in German and Englisch. In: Diehr, Bärbel & Rymarczyk, Jutta (eds.) (2010), 69–88.

Seymour, Philip & Erskine, Jane (2003), Foundation Literacy Acquisition in European Orthographies. *British Journal of Psychology* 94: 2, 143–174.

Szpotowicz, Magdalena & Campfield, Dorota E. (2016), Developing and Piloting Proficiency Tests for Polish Young Learners. In: Nikolov, Marianne (ed.) (2016), *Assessing Young Learners of English: Global and Local Perspectives*. Berlin: Springer, 109–137.

Katharina Karges, Malgorzata Barras, and Peter Lenz

Assessing young language learners' receptive skills: Should we ask the questions in the language of schooling?

Abstract This article discusses empirical evidence concerning the following questions from assessment research and development: Should items in a foreign language (FL) reading test be presented in the language of schooling, or rather, in the target language? Which language would lead to less construct-irrelevant variance? This issue was explored as part of a research project involving young learners studying French as a compulsory subject. In-depth qualitative information was gathered by means of stimulated recall interviews with individual students. Further evidence was collected through (a) a questionnaire completed by over 900 students and (b) a qualitative analysis of items used in an assessment comprising parallel items in both, the language of schooling and the target language. The combination of the results from these various sources allows for an evidence-based recommendation in favor of the use of the language of schooling in foreign language assessments for young FL learners.

1. Introduction[1]

In many European countries, children are now learning a foreign language in primary school. These young learners differ from older students in various respects, making it necessary to adapt both instruction and assessment. For instance, when assessing language skills, teachers and test developers have to take into account that young learners have less experience in reading instructions and completing tasks, have shorter attention spans and limited world knowledge (Bailey, Heritage & Butler 2013; Hasselgreen 2005). They do, however, usually share the official language

1 The authors would like to thank the many students and teachers who participated in our studies for their enthusiasm and patience. We also thank Gabriela Lüthi and Patrick Karges for their valuable input and help during the redaction of this article and the editors of this volume for their helpful comments and their overall support.

of schooling (LS), even in contexts where many students speak different languages at home. This makes it possible to phrase instructions, questions about a text and answer options in this language, which may reduce the cognitive load induced by the assessment itself. Yet, according to some, often practitioners, switching languages during an assessment may confuse the test takers. This leads to the essential question that this article attempts to answer based on empirical evidence: What use of languages can actually help improve the quality of language assessments for young learners?

2. Literature review

In the context discussed in this article, i.e. foreign language reading assessment for educational monitoring purposes[2], the test scores should allow a credible statement of whether or not a population of students has the ability to read in the foreign language at the level prescribed by the national educational standards (Messick 1990, Kane 2006). Since the educational standards themselves do not provide sufficient basis for an operationalizable construct, we needed to complement them with appropriate additional sources[3]. Therefore, our test construct is a conceptualization of foreign language reading ability based on the standards document

2 The institution which coordinates the Swiss educational monitoring defines the term as follows: "Educational monitoring is the systematic acquisition and compilation of information about an educational system and its environment. It forms the basis for educational planning and policy decisions, accountability and public debate." (SKBF 2019).

3 In Switzerland, as in many other European countries, the foreign-language curricula are based on the Common European Framework of Reference for Languages (CEFR; Council of Europe 2001). Thus, the minimal standard for reading in a foreign language at the end of primary school (Grade 6, age 12) is defined with respect to CEFR level A1.2. By the end of lower secondary school (Grade 9, age 15), students are supposed to reach CEFR level A2.2 (EDK 2011). For the actual item development, the user (not constructor) oriented CEFR descriptors (Alderson et al. 2006; Council of Europe 2001: 37 f.) were complemented by aspects of a reading test construct based on Khalifa and Weir's reading model (2009: 43 ff.). A major feature of this model are the various types of reading which are initiated by the goal setter depending on the purpose of the reading activity (ibid., based on Urquhart & Weir 1998).

(EDK 2011), the CEFR (Council of Europe 2001) and, among others, work by Khalifa & Weir (2009: 40 ff.).

According to Messick (1995: 742), there are two major threats to the measurement of this (or any other) test construct: construct underrepresentation and construct-irrelevant variance. Construct underrepresentation exists if the test, although framed as a general reading test, only assesses part of the reading construct, e.g. if it elicits only reading for orientation or careful local reading. Construct-irrelevant variance is present when test scores are due not only to the reading ability of a test taker but at least partially to other sources, such as specialized world knowledge, test wiseness or guessing.

Avoiding these pitfalls as well as establishing sufficient evidence for a valid interpretation of the test results hinges on two things: (a) a clear idea of the test construct and (b) sufficient knowledge of how the input text, the task(s) and the test taker characteristics interact. The latter is the object of research concerned particularly with factors that influence the difficulty of test items. In receptive language assessments, these factors include the propositional density and the topic of the input text (e.g. Freedle & Kostin 1999), the motivation, strategic knowledge and language ability of the test takers (e.g. Jeon & Yamashita 2014; Shiotsu 2010) as well as the characteristics of the test method or items (e.g. In'nami & Koizumi 2009; Ozuru/ Briner/ Kurby & McNamara 2013; Rodriguez 2003).

One item characteristic which may influence the functioning of an item is the language in which the questions are asked (and the answer options are given). Is it the target language (TL), which is the object of the test, or another language that the test takers have in common, usually the LS or the test takers' L1? Over the years, several researchers have looked into this issue (Brantmeier 2006; Cox/ Bown & Bell 2019; Filipi 2012; Godev/ Martínez-Gibson & Toris 2002; Shohamy 1984; Wolf 1993), but the topic has always been on the sidelines of assessment research. A reason for this may be the preponderance of international language assessments such as the Cambridge exams, the Goethe exams or the DELF/DALF exams. These are designed for test takers all over the world who do not necessarily share a common language (except for the one being tested). Hence, the tests are delivered entirely in the TL. At lower levels of language proficiency, this is not unproblematic: Test items in the TL always carry the risk of

being partly or entirely misunderstood (Godev et al. 2002: 204; Gordon & Hanauer 1995: 302), thus limiting a valid interpretation of the test results. To account for this, international exams for lower proficiency levels tend to use well-known item types, which do not need to be explained to the test takers. Also, if at all possible, the language used in the items is often simpler than the reading text itself (Alderson 2000: 86; Green 2014: 113) and in some cases, pictures are used instead of words. Many test takers also familiarize themselves with the specifics of the test beforehand, e.g. by consulting sample tests available online or by taking classes that prepare candidates for certain exams (as is evidenced by the host of material and information available on the internet).

Tests that are written entirely in the TL are not limited to international assessments. Teachers may favor a monolingual approach to foreign language teaching, arguing that exclusively using the TL is more authentic and exposes students to more input, which promotes learning (Godev et al. 2002: 204). Others may fear that using the LS in a foreign language test might confuse students, disadvantage learners who have difficulty in the LS, or introduce a new level of difficulty through the need to translate (ibid.). Finally, teachers may also choose to use target-language assessments in order to prepare students for international examinations. In our own context, compulsory foreign-language learning in Swiss schools, personal discussions with teachers suggest that assessments with questions and answer options in the TL are common practice, and that, from the point of view of those practitioners, the arguments in favor of using the TL hold significant weight.

Yet, the studies which investigated this issue empirically indicate that the use of a strong common language, if it exists, may be a sensible choice in terms of validity. In those studies, most items in the common language of the learners (often the LS) turned out to be easier than items in the TL. This in itself does not necessarily mean that items in the LS lead to more valid test results, but a closer look at the questions and the test takers' answers often points to this conclusion: For instance, Godev et al. (2002: 210 f.) describe several instances where the students' short answers in the LS clearly indicate that they "understood the text well enough to respond correctly to [a certain] question" whereas insufficient command

of the TL made a correct answer to the otherwise same question less probable (see also Wolf 1993: 482).

In cases where items in the TL were found to be easier, a closer look usually revealed reasons such as a direct match between the wording in the L2[4] item and the L2 text (Godev et al. 2002: 209, 211; Wolf 1993: 481 f.) or the presence of transparent cognate words (Filipi 2012: 519). In both cases, it remains unclear whether students who answered correctly did so because they understood the information in the text or because they successfully matched individual words without being aware of their meaning, which would amount to guessing. Shohamy (1984: 158) observed diverging answer patterns in corresponding L1 and L2 multiple choice items, suggesting that students guessed more often whenever they encountered L2 answer options they did not understand. Godev at al. (2002: 210) found evidence of guessing in a number of L2 short answers that did not answer the question in a sensible way but were copied more or less directly from the input text. These findings suggest that students guess more often whenever they encounter items they cannot understand or when they have to give answers they cannot formulate. Such answers are difficult to interpret with respect to the reading construct because guessing is only very inconsistently related to reading proficiency.

Shohamy (1984: 157) argues that the use of the L1 in foreign language assessment may even be considered more authentic "since many students, while processing L2 texts, tend to utilize known elements from their L1 rather than unknown elements from L2". She also maintains that questions and answer options in the LS may offer clues to understand the text better, which she considers to "make the task more natural" (ibid.).

The studies mentioned up to this point mostly focus on low-proficiency learners of the TLs. For learners at higher levels of language proficiency, there is some evidence that the use of the TL may have a lesser effect on the test results (Brantmeier 2006; Shohamy 1984: 155 f.). In a recent study by Cox et al. (2019), however, advanced learners of Russian still performed better on a multiple-choice reading test when the items were in their L1. This finding suggests that there is no clear-cut point in language proficiency

4 Several of the studies cited here use the terms L1 and L2 instead of LS and TL.

development where the language of the items becomes irrelevant for the test results. Instead, according to the authors, whether the LS or the TL should be preferred "is likely dependent on the testing situation and population as well as on practical considerations" (Cox et al. 2019: 134).

3. Method

In the following subsections, we will describe how we investigated the "test language issue" based on three sources of empirical evidence: stimulated recall interviews, questionnaire data and qualitative item analysis. Most of our evidence stems from the Task Lab project, which investigated task and test-taker characteristics in a low-proficiency French reading assessment. The Task Lab questionnaire data is complemented by students' answers to a questionnaire used in a subsequent task development project for large-scale educational monitoring conducted in Switzerland in 2017 (ÜGK[5]).

3.1 The research projects

The Task Lab project was conducted at the IoM between 2014 and 2016. Its primary aim was to investigate the impact of selected task factors on test scores and test behavior, most importantly item type, i.e. multiple-choice questions (MCQ), short answer questions (SAQ) or matching, and item language, i.e. the LS or the TL. These two task factors were investigated in a reading comprehension assessment of French as a foreign language. The target group were German-speaking Swiss pupils at the end of primary school (Grade 6), who had been learning French in a non-intensive two to three-lesson-per-week course for four years. They are expected to reach level A1.2 (i.e. CEFR level A1) by the end of sixth grade according to the national education standards (EDK 2011). The orientation of the project was predominantly quantitative: Around 600 students participated in the main study. During the piloting sessions, qualitative data was gathered by means of stimulated recall interviews (N = 34).

5 *Überprüfung des Erreichens der Grundkompetenzen*, or *Vérification de l'atteinte des compétences fondamentales*; literally "Verification of the achievement of the core competencies".

The assessment comprised 18 reading comprehension tasks, each consisting of a written text input in French and three successive items (54 items in total). Six of the eighteen tasks were matching tasks with items developed in two language variants, i.e. with the same items in German and in French. In the 12 remaining tasks, both the language of the items and the answer format were varied. Thirty-six items were therefore available as multiple-choice as well as short-answer items in both French and German.

During the main data collection, each pupil worked on the French reading tasks for 45 minutes[6], encountering all formats and both languages in about equal measure. The students also completed several cognitive and linguistic component tasks (e.g. a vocabulary test) and a questionnaire. Overall, each student participated in the data collection for 110 minutes during normal school hours. All survey instruments used in the main study were piloted in a field test by 131 students. The tests were delivered on a computer, with the exception of a short paper questionnaire following the reading comprehension test.

Informed by the findings gathered in the Task Lab project, the IoM was later responsible for the task development in the foreign language test of ÜGK 2017, the first nation-wide survey of students' foreign-language skills for educational monitoring purposes in Switzerland. This large-scale assessment targeted the receptive skills of students in their first foreign language at the end of primary school using multiple-choice items in the language of schooling. Pre-piloting of the reading and listening tasks was conducted in three language regions, German-speaking Switzerland and Italian-speaking Switzerland, where French was assessed, as well as French-speaking Switzerland, where the students sat a test of German (more details in Tab. 1). During these pre-piloting sessions, questionnaire data relating to the "test language issue" was collected.

3.2 Stimulated recall interviews

We used stimulated recall interviews (Gass & Mackey 2017) to pre-pilot the reading comprehension tasks developed in the Task Lab project. The

6 Each student encountered 13 out of the 18 tasks because of time restraints.

interviews primarily aimed at finding out how young learners of French proceed when solving computer-based reading comprehension tasks, which strategies they apply, and how their test-taking behavior is influenced by the language of the items and the answer format.

Four previously trained interviewers conducted interviews with a total of 34 pupils. In a 45-minute, audio-recorded session, each participant was confronted with a selection of tasks from the pool described above. Typically, each student completed two to three tasks with three items each. Immediately after each item, the students were asked to explain their approach, their considerations, the strategies used and difficulties they had encountered, if any.

At the beginning of the session, each pupil was informed about the aims of the study and the procedure. While the pupils were reading, the interviewer stepped back. When the pupils indicated that they had finished processing an item, they were interviewed. The interview lasted approximately one to three minutes per item. The task displayed on screen and the answer given by the pupil (the selected answer option or the written short answer, respectively) served as a stimulus for each interview. The interviews were conducted in standard German and/or a Swiss German dialect and were based on a written guideline that was constructed with the research questions in mind. During the interview sessions, the researchers noted down potentially interesting observations to complement the audio recording.

The interviews focused on the students' responses to the items and the thoughts that led to them. Whenever the questions (and where applicable the answer options) were in the TL, the pupils were asked whether they had understood them. In the case of multiple-choice and matching items, the interviewer asked why a particular answer was chosen (rather than the others), and in the case of short answer items, whether there had been difficulties in formulating the answer. If required, the interviewers asked the pupils additional in-depth questions, for example, how often and how accurately they had read the text, where in the text they had found the answers, whether the task instructions had been clear. The last five minutes of each interview were dedicated to more general questions, e.g. questions related to the use of the computer.

In general, we observed that most of the students were happy to share the thoughts they had during the test, and we had the impression that they actually expressed what they thought. For instance, they readily admitted to have merely guessed or to using test wiseness strategies such as copying

text directly from the input text, possibly because they had been informed in advance that their performance in the test had no influence on their school grades.

The interviews were transcribed and coded using the data analysis software *MAXQDA* (VERBI Software 2015). Data analysis was carried out according to the principles of structuring content analysis (Mayring 2010). The coding categories were derived from the research questions (top down) and and inferred from the collected data (bottom up). The results of the data analysis served as a basis for the revision and adaptation of the test tasks for the main study.

3.3 Questionnaires

In the stimulated recall interviews of the Task Lab study, we collected some evidence on what the test takers thought about the "test language issue". To expand on this evidence, we added a question to a short questionnaire given to the participants of the main study: At the end of the reading test, students were asked what they had found easier – the questions and answers in German, their LS, or in French, the TL. They checked a box to indicate their preference and had the opportunity to explain their answer.

Later, in small-scale field tests supporting the development of test tasks for the Swiss educational monitoring survey (ÜGK), we asked sixth-graders in three language regions to speculate what they would find easier (or better[7]) regarding the language of the items: the LS they had just encountered during the field test, or the TL. They, too, were asked to explain their answer. Some of these students solved reading tasks similar to the ones we used in Task Lab, others completed listening tasks.

Overall, we collected the opinions of 936 6th-graders on the test language issue. Of these, 879 also wrote a comment[8]. Table 1 gives more details on the sample.

7 Due to a translation error, the word "easier" was replaced by "better" in some of the questionnaires.

8 Not all of those comments are meaningful for this article. For instance, some students gave their opinion on how they found the test in general, or they made observations we cannot fully understand in hindsight (e.g. in a test of reading: "The speed was okay overall").

Tab. 1: Questionnaire data on the "language issue" (overview)

Project	Assessed skill during data collection	LS	TL	N
Task Lab	Reading	German	French	591
ÜGK dF[a] (field test)	Reading	German	French	45
ÜGK fD (field test)	Reading	French	German	47
ÜGK dF (field test)	Listening	German	French	46
ÜGK fD (field test)	Listening	French	German	78
ÜGK iF (field test)	Listening	Italian	French	129

[a]The acronyms dF, iF and fD stand for the German names of the languages involved in the assessment: d, i and f represent the pupils' LS (German, Italian and French respectively), whereas F and D stand for the target languages of the assessment (French and German).

3.4 Qualitative interpretation of differences in item difficulty

As pointed out in the literature review, the language of the items does not affect the difficulty of each item in the same way, nor do other characteristics, such as item type. To account for this, we examined the items we used in Task Lab, making use of the empirical item difficulties we obtained from the main study.

To determine the item difficulties, a Rasch model was estimated using the R package "TAM" (Kiefer, Robitzsch & Wu 2015; R Core Team 2014), in which the four item variants (defined by MCQ or SAQ[9]; TL or LS) were considered as separate items. The model used responses[10] from 577 participants. Each of the 144 items was answered by at least 84 test takers (120 on average). The items fit the Rasch model sufficiently well according to common standards (e.g. OECD 2014: 151) as all infit values except one fall between 1.20 and .80[11].

9 The six matching tasks are not considered in this section.

10 The MCQ items were scored automatically based on the option chosen by the student. The SAQ items were scored manually by two raters according to detailed scoring guidelines. The raters first scored all answers individually and then discussed diverging scores and reached an agreement in each case. Only dichotomous (0 or 1) coding was used.

11 In fact, the SAQ items generally show overfit (low infit) while the MCQ items show underfit (high infit). This fact has to be taken into account when tests are

The item difficulties thus obtained were grouped so that all four format-by-language variants of one item could be visualized together. We then examined these groups and identified possible reasons for the various patterns of relative item difficulty by referring back to the items, i.e. the input texts, the questions and the students' responses.

4. Results

4.1 Stimulated recall interviews

The analysis of the transcripts of the stimulated recall interviews conducted in the Task Lab project revealed three major topics with respect to the "language issue": the students' individual preferences regarding the language of the items, the comprehension problems they encountered when confronted with items in the TL and the difficulty of writing short answers in French. In the following, these three perspectives will be presented and illustrated with original quotes from the interviews[12].

4.1.1 Students' preferences regarding the language of the items

As mentioned above, during the Task Lab study each student was confronted with items in each language version, i.e. in the LS German and in the TL French. In the interviews we asked the students (N = 34) if they felt bothered or confused by having to switch between German and French when the items were formulated in the LS and whether they preferred the items in the TL or in the LS. Twenty-eight of the thirty-four students we interviewed[13] did not consider this to be a problem and favored the LS: they argued that this allowed them to understand the questions (better), which in turn enabled them to know what to look for in the reading text.

constructed which consist of both item types (cf. Lenz/ Karges & Barras 2019), but it is not relevant for this analysis.

12 The quotes are English translations of the original German transcripts.

13 Due to the qualitative nature of our pilot study, the numbers reported in the following paragraphs should not be understood in a statistical sense. Instead, they should give an idea of the general tendencies we found in our data.

I: If you could choose, would you choose the questions in German or French?
S: German.
I: And it doesn't bother you that you have to switch between languages?
S: No, it doesn't bother me. (Je116[14])

I: We had some questions in French, and some in German. Which do you like better?
S: In German. Because in French you're busted when you don't understand the question. (Je115)

S: (…) when I have a French question, it sometimes happens that I don't understand a word and can't answer the question because I don't know what to do. But I actually understand the whole text. (Je110)

Not everybody was partial to the use of German in a French test. Thirteen students mentioned that the language of the item did not really matter to them and that their test results would remain the same if the items were in the TL[15].

I: Which did you like better: the questions in German or in French?
S: Both the same. […] If you don't understand individual parts [= words], you can put them together and guess what the question might be.
I: Then would you say that you might have given better answers to the German questions than to the French ones?
S: No, not necessarily. (Ge1103)

Only one of the 34 students we interviewed actually said that he would have preferred to have the questions presented in the TL and found the German items confusing. His home language, interestingly, is Portuguese.

I: You had some German and some French questions. What was that like?
S: In French it was a bit better.
I: Why?
S: If they are in German, you have to translate the words on top of it. (Ge198)

It is likely that the linguistic similarity between French and Portuguese made the French items easier for him to understand. Whether this

14 The codes are unique identifiers of individual students used throughout all projects. Their meaning is not relevant for this paper.

15 Since the students who participated in these stimulated recall sessions only worked on a small number of tasks, it is impossible to say whether they would have indeed performed similarly with items in either language.

preference for the TL is typical of students who speak Portuguese or other Romance languages remains to be investigated.

4.1.2 Problems understanding items in the TL

While dealing with the items in the TL, students often reported that they did not fully understand the questions. As a result, they could not always be sure whether they were looking for the right answer.

> I: And the questions?
> S: If they were in German, maybe I would have had them right. I don't know now if that's right. (Ge1106)
>
> S: I don't have a complete answer for this one, because I didn't understand the question very well. (Ge1101)

This was observed 133 times across all students and items[16]. We also observed many instances (N = 127) where the students had difficulty understanding a question because one or several words were unfamiliar to them. They usually had a vague idea of what the question could be about but were insecure about whether their assumption was correct or they only understood individual words of the question. This was usually not enough to answer the question correctly, and the students were quite aware of that.

> *Question: What occupation do Tom's grandparents have? [Quel est la profession des grands-parents de Tom ?]*
> S: […] I don't know exactly what "profession" means.
> I: (…) What do you think the question means?
> S: It's somehow about Tom's grandparents, but I don't know exactly what it says. (Ge1107)
>
> *Question: Why do the young kids rarely have accidents on the lake? [Pourquoi est-ce que les jeunes enfants ont peu d'accidents sur le lac ?]*
> I: (…) Then what do you understand about the question?
> S: Not much. I only understand "the lake". And otherwise I didn't understand the question at all. (Je112)

Interestingly, in a considerable part (N = 77) of those instances, students struggled to specifically understand the question word(s) and therefore did

16 Note that overall, we discussed 222 items in the TL with the students.

not know what they were supposed to answer. Even important question words that occur frequently, such as "why" or "where", posed a problem in some cases.

Question: Why did Hans Kaufmann start the project? [Pourquoi Hans Kaufmann a-t-il commencé le projet?]
I: Why weren't you quite sure?
S: Because I don't know "pourquoi" [= why] exactly… "What" or something. (Je101)

I: If you look at the question: "Où as-tu besoin d'un dictionnaire?" [Where do you need a dictionary?]. Do you understand "où" [= where]?
S: No. (Ge1107)

While discussing the multiple-choice items in our study, we observed 44 times that students struggled not only to understand the question itself but also the answer options provided. Overall, multiple-choice items in the TL were discussed in 99 instances.

I: Do you understand the question?
S: No.
I: Did you rather guess?
S: I don't understand any answer. (Ge1108)

I: Do you know what the other two options mean?
S: No, not really. (Ge1106)

A common mistake we observed in our data was the misinterpretation of French words as German cognates. Such words were translated incorrectly by some students[17] – for example "prof" [teacher] as "Profi" [professional]. This phenomenon may explain a number of incorrect responses.

Question: Why is Vidal's way to school special? [Pourquoi le chemin de l'école de Vidal est-il spécial ?]
I: Do you know what the question is?
S: Why is Vidal's chemistry lab so special?

17 Since we did not systematically ask the students to translate the items or input texts in the target language, we only observed this phenomenon by accident (nine times overall). We have reason to believe that this happened much more often, though.

Question: Who prefers languages over mathematics? [Qui préfère les langues aux maths ?]
I: Do you have a spontaneous idea?
S: Something about math. What is your longest[18] math lesson? (Je114)

During the interviews, we found evidence that at least in some cases, pupils who did not understand the question had actually understood the input text quite well[19]. In the following example, one student was able to write the correct answer in French after the interviewer had translated the French question into German.

I: So you have no idea what it could mean?
S: No.
I: Ok, then I'll tell you, and after that you'll try to answer the question. It means: "In this project the children learn to play floorball. What else do they learn?"
S: Ok.
((S reads, types))
S: So, I wrote that the children learn to play in a team[20]. (Je101)

Whenever the students encountered the types of problem described above, they used various strategies to try and give a suitable answer. For example, if they did not understand a multiple-choice answer option, some students tended to ignore that particular answer option.

I: And why didn't you choose the other options?
S: Concerning the third, I do not know what that means. (Ge196)

S: I read through again. Then I thought it was the one in the middle because I didn't understand the others. (Ge1105)

In many instances, students simply guessed or used test-wiseness strategies such as trying to locating unknown words in the question or answer

18 The word "longest" refers to the German word "längste", superlative of "lang", which the student may have confused with the French word "langues" [languages].
19 Again, we did not elicit this systematically. It could be found 9 times in our data.
20 The student wrote "les enfants apprennent la jouer dans l'equipe" [children learn playing in the team] which can be considered a correct answer to that question.

options in the input texts[21]. Whenever they found a match, the students would try to figure out an appropriate answer with the help of the context provided by the input text. Of course, this strategy was unsuccessful if the unknown word in the question did not appear in the input text, as in the example below.

Question: What did Alicia like the most? [Qu'est-ce qu'Alicia a aimé le plus ?]
I: You read the question and didn't understand it, and then you read the text again. What did you do while reading?
S: I looked what the answer could be.
I: And how did you do that?
S: I just read here, and maybe the word "aime" [to like] appears somewhere in the text.
I: You were looking for the word you didn't understand.
S: Yes.
I: And then you didn't find it.
S: Yes.
I: That's why you had to give up.
S: Yes. (Je115)

4.1.3 Writing short answers in the TL

In the Task Lab study, when students encountered a short-answer question in French, they also had to answer them in French. During the interview sessions, we observed that for 22 out of the 36 students at this low proficiency level, writing even the simplest answers in the TL represented a significant problem. The students found it difficult to focus on content, syntax and spelling at the same time. This bothered them even though they had been told at the beginning of the test that linguistic errors would not be taken into account. Many students also mentioned that they lacked vocabulary and were therefore unable to formulate their answers in French.

We found evidence that some students had something very different in mind than what they actually wrote in French.

21 In our data, we observed 36 instances where students admitted to having guessed and 45 instances where we were able to confirm the use of this strategy. As before, we have reason to believe that we did not identify all instances systematically.

I: Can you read your answer to me?
S: So, Karusu loses his dad at the zoo.
I: And you wrote "Karusu devenu papa" [Karusu became a dad]. (Je105)

S: I wrote: "Pierre Dumont en danger." [Pierre Dumont in danger.] (…)
I: What would you write in German if you could write this answer in German?
S: Pierre Dumont is very dangerous. (Je117)

On the other hand, 10 students stated that it was easier for them to deal with SAQ items in the TL because this gave them the opportunity to copy words or passages directly from the French text or the French question.

I: Was writing in French a problem?
S: No, I could copy that. (Je116)

I: And here you had to write some of the answers in French. Do you think that's difficult?
S: Well, if the text is in French, it's not [difficult] because you can copy a lot of things. (Je104)

Having gained these insights, we decided to further annotate the short answers gathered later in the main study. We identified all answers that were copied directly from the text, i.e. answers which contained three or more words in the same order as they appeared in the input text. Overall, more than a third of the French SAQ answers (37 %) were at least partly copied directly from the French text. In these cases, it depended largely on the item whether the strategy was successful or not: Whereas, for instance, 90 % of the copied answers to item T01-2 led to a correct answer (16 out of 18), because the item elicited a concrete piece of information, this was true for only 11 % (7 out of 65) of the copied answers to item T03-2, which demanded an inference from the content of the text.

We also annotated short answers as "absurd" when they did not in any way answer the question. This was the case for 33 % of all short answers in French. Again depending on the item, up to half of these non-sensical answers had been copied directly from the text. Thus, it appears that some students simply chose a random word or text fragment from the input text or the question when they did not know what was asked and to what they were supposed to provide an answer.

Q: What does Emilie like in school?
A1: Zurich. (Br625)
A2: [a] world. (Mu712)
A3: He preferred at school Thursday. (Vi129)
Q: What do the two want to buy?
A1: dad and boy (Vi154)
A2: and you don't like to go by bike (Bi366)

We also encountered "absurd" answers in German (10 % of all German answers were annotated as such), but these were often comments unrelated to the text itself, like "no idea" [keine Ahnung] or, supposedly, answers copied from nearby students who were completing a different task. These findings show that the students employed different strategies depending on what language(s) they were dealing with.

4.2 Questionnaire

As described in the methods section, the students on the Task Lab main study and on the ÜGK field-test were given a short questionnaire where, amongst other things, they first indicated which language they found easier (or better) for the questions and answers and then justified their choice or left some other comment.

As Table 2 very clearly shows, only a small minority of the students chose the TL when asked what they found easier (or, in two questionnaires, better[22]) (Task Lab: 10 %, ÜGK: 15 %). In the ÜGK field test, where students had only completed items in their language of schooling, another 25–35 % of the students indicated that they had no preference concerning the language of the items. This means that overall, a clear majority of the students in our samples preferred the use of the LS or, at least, was indifferent towards it. Based on these results, we assume that the use of the LS in foreign language assessments (reading and listening) does not confuse most students.

22 Overall, there is no discernible difference between the students' reactions to the "easier" question and the "better" question. Their written justifications are very similar in both cases.

Tab. 2: Distribution of the students' answers in the questionnaires

Project	Skill assessed during data collection	Total	In favor of the LS	In favor of the TL	No preference
Task Lab	Reading	608	537	54	17
ÜGK dF	Reading	45	22	13	10
ÜGK fD	Reading	47	24	10	13
ÜGK dF	Listening	56	34	3	19
ÜGK fD	Listening	78	49	12	17
ÜGK iF	Listening	129	68	16	45
ÜGK all	Listening or reading	355	197	54	104

The reasons the students gave for their choice provide more insight[23]. The most common type of answer is related to the actual languages (LS and TL) concerned, and may not be generalizable in a straightforward manner to any reading or listening test at the proficiency levels in question: Many students argued in favor of the LS either by stating that they were proficient in this language or that they were *not* proficient in the TL.

> Because [Italian] is my language! (ÜGK iF, Ca538)
> Because I'm not good at German. (ÜGK fD, Es139)
> Because I never speak French and I understand almost nothing! (ÜGK dF, Be635)

Some students who preferred the TL or indicated that they were indifferent, asserted that they either found the TL easy or knew both languages equally well.

> Because everything was easy, because I have spoken French since birth. (Task Lab, Br671)
> Because I understand both languages well enough. (ÜGK fD, Fa347)

Some students gave more precise reasons related to the test itself. For instance, students who opted for the LS stated that it helped them understand the questions and answer options. Some students also pointed out

23 All answers cited in the following are literal English translations of the students' handwritten answers.

that reading the questions and answer options in their LS gave them some idea of what the text was about.

> Because that way you can understand more of what they're saying because [the questions in the LS] tell me a lot. (ÜGK iF, Br448)
> Because then I knew what it was about and what the question was. (Task Lab, Si475)

Many students who indicated that they preferred the use of the TL pointed out that words or phrases in the questions or answer options could help them find the answer. With respect to the multiple-choice items, this was mentioned particularly often by the students who had just taken the listening comprehension test in the ÜGK trials. These students argued that seeing a word in its written form may give them clues about the words that were pronounced.

> That way when I heard the text and then read, I understood better, I think. (ÜGK iF, Br441)
> I understand German better when I see it written. (ÜGK fD, Fr358)

With respect to the SAQ items, which were only used in the Task Lab project, many of the 10 % of students who preferred the TL argued that writing short answers in that language allowed them to copy words or phrases from the text. This is the same argument we had already encountered in the stimulated recall interviews (cf. Section 4.1.3).

> Because then, most of the time, you could look for the words in the text. (Task Lab, Bi383)
> Because in French, I could take the answers directly out of the text. (Task Lab, Vi148)

A few students were aware of the benefits of both language versions.

> In French you know because it's my mother tongue. In German, you can locate the words. (ÜGK dF, Bu398)
> In German, when you don't know a word in French. In French: When you cannot translate something. (Task Lab Vi182)

Finally, a small number of students stated that mixing the two languages did not appeal to them because they were not used to it. In our entire sample, however, there are less than 10 instances of this, and not all of them necessarily imply that the test results suffered from mixing the languages.

> Because I get confused between Italian and French. (ÜGK iF, Br413)
> I've practised it more the other way. (ÜGK dF, Ta251)

Finally, a rather small group of students considered the testing situation to be a learning opportunity. These learners tended to prefer the TL because it gave them more opportunity to practice the language.

> Because you can learn more that way. (Task Lab, Zu302)
> It's easier in French but it would be funnier and more exciting to put the questions in German. (ÜGK fD, Fa351)

Interestingly, it was most often the students from Ticino who brought forth this argument. One possible reason, which is also reflected by the comparably large number of students who answered "I don't care" to the initial question, is that these Italian-speaking learners of French had just encountered listening tasks that were decidedly easier for them compared to the other groups of test takers. This is most probably due to the fact that there is a close typological relationship between Italian, their LS, and French, the TL (both being Romance languages).

> That way it's a little more difficult and a bit more entertaining. (ÜGK iF, Br429)
> Because that way you can practice French better, also with the questions. (ÜGK iF, Ca486)

4.3 Qualitative interpretation of differences in item difficulty

As mentioned in the methods section, the following section builds on the results of a Rasch analysis of the reading test in the Task Lab main study. The results of the analysis show that, in general, multiple-choice items are easier than short-answer items, and items in the LS (German) are easier than items in the TL (French). This is the pattern we would expect based on the literature, and it can be observed like that in 13 out of the 36 items. However, for the remaining 23 items, there seem to be three major deviations from this pattern:

- multiple-choice items in French are easier than multiple-choice items in German (six items),
- short answer questions in French are easier than short answer questions in German (six items),
- short answer questions in German are easier than multiple-choice items in French (nine items).

The 12 items which are easier in the TL in at least one item format seem to differ from all other items in one important respect: the overlap between the item (question and/or answer options) and the input text in the TL version. Many of these items contain words in the French questions which can be matched directly to the relevant passage in the text. As a result, short answer questions of this type can be answered correctly by simply copying words or phrases from the input text. Similarly, the TL answer options of multiple-choice items contain words or phrases that can be found verbatim in the relevant passage. This effect can be illustrated particularly well in item T06-2, which is easier in French both in the multiple-choice and in the short-answer version. In this item's input text, the boy Tom presents himself and his family. In the item, students have to indicate the occupation of Tom's grandparents. The correct answer, "paysan" [farmer] appears in the text, in the same sentence as the word "grands-parents", which makes it possible to choose the correct answer option even if neither the meaning of the question nor the answer were quite understood. If, however, the item is in German, the test takers have to know that "Grosseltern" means "grands-parents", and then either be able to identify the word "paysan" and know that it means "Bauer" in German, or – in the multiple-choice version of this item – identify the words "vétérinaire" and "professeur de physique" and discard them as possible answers. A similar argument can be made for the short-answer version of this item and for most of the other items which follow the first two patterns.

The third pattern, where the short-answer item in German is easier than the multiple-choice item in French, is observed in items which seem to roughly belong to two groups. One group is composed of the first items of five tasks. These items ask about the general topic of the given texts, and it seems that some students, when confronted with the multiple-choice question in the TL, failing to understand what was being asked, tried to match words from the text with the answer options. In these items, however, this strategy was less successful because words from the input text usually appeared in all three answer options. Students who encountered the question in their LS knew what they were being asked and thus had the opportunity to show that they understood the general topic of the text. As a result, more correct answers were given in the short-answer items in the LS. The remaining four items which show this pattern contain

multiple-choice answer options which do not reveal much about the content of the question. This seems to make it more difficult to decide on an answer without fully understanding the question. A good example of this phenomenon is item 7-02, which asks "Qui utilise son Natel le plus souvent?" [Who uses their cell phone the most often?]. Since the answer options are merely the names of three people who have a conversation in the input text, they do not give much of a clue as to the content of the answer. Thus, students who do not understand the question cannot even make an educated guess about the answer.

5. Discussion

In this article, we have presented empirical evidence relating to the use of the language of schooling (LS) and the target language (TL) for the questions and answers in tests of foreign language reading and listening intended for young learners in a compulsory school context. Our data is based on assessments at the lower levels of language proficiency, especially reading comprehension targeting the region around A1.2.

In the Task Lab project, we used items in both the LS and the TL. We observed and interviewed students during task pre-piloting and later analyzed the test results to get a clearer idea of how students process items in the two languages. We also gathered the students' opinions on the "test language issue" through stimulated recall interviews and a questionnaire. More data was collected with a similar questionnaire in the more recent ÜGK task development project, where students only worked on items phrased in the LS.

As discussed in Section 2, empirical evidence from earlier studies indicates that for the test questions and answers a language that is familiar to all test takers is preferable with regard to test validity. Our results corroborate these previous research findings. We found almost no evidence for the assumption that the use of the LS in a foreign language assessment might confuse test takers (see also Filipi 2012: 527). On the contrary, using a language in which test takers are more proficient may actually help to better reflect the reading construct: Without the need to spend a lot of time and cognitive resources on understanding the questions, students may be able to concentrate more on the text itself. Fully understanding

the questions may also trigger more authentic types of reading because students can choose more consciously whether, for instance, local or global reading is more appropriate for answering a specific question (see also Cox et al. 2019: 123). After all, reading is hardly ever done in the real world without knowing its purpose (see also Shohamy 1984: 157).

Furthermore, whenever the items were presented in the TL, we found evidence that correct answers less often reflected good reading proficiency, and more often depended on chance knowledge of a specific word, (more or less informed) guessing, or test wiseness strategies. This was confirmed by many students we interviewed and it can also be inferred from the qualitative interpretation of the differences in item difficulty. Admittedly, a reading construct may also include the ability to deal with uncertainty and to infer the meaning of unknown words from the surrounding context, pictures or other elements. In Switzerland, and in many other European countries, developing this kind of strategic competence is part of the foreign language curricula. From this point of view, one may argue that guessing a correct answer from a limited number of clues is a desirable skill that should be assessed. However, guessing what one is supposed to do (i.e. which information is to be found in a reading text) does not seem to be part of a construct worth assessing, because guessing the meaning of a test item is not the same as inferring meaning from context during reading. Furthermore, we doubt that it makes sense to combine the measurement of test-strategic competence and reading comprehension in the same items and, as a result, to obtain test results which can be interpreted as being evidence of *either* good (or bad) reading skills *or* good (or bad) test-strategic skills.

In our Task Lab interviews, we encountered students who did understand the information that an item asked for, but were unable to show this understanding, either because they did not understand the question or the answer options or because they could not write a (sufficiently good) short answer in the TL. The item analysis from the main study corroborates this finding. It shows that TL items were often harder than the same items in the LS, suggesting that the TL items introduced a source of difficulty unrelated to the understanding of the input text.

Of course, the use of the LS can also be problematic: Students with a very weak command of the LS, but good command of the TL may be at a

disadvantage. Indeed, one student we interviewed, whose L1 is typologically related to French, indicated that he preferred the items in the TL, and a small number of students who participated in the main study argued similarly. This would have to be more closely examined, however, because it remains unclear whether these students' performance on the LS items was indeed affected negatively. Otherwise, we found very little evidence that the students' proficiency in the LS was an issue: Even though a number of multilingual students participated in our data collections[24], most of them clearly had more contact with the LS than the TL.

Additionally, the students' answers in the questionnaire used in both projects confirm that most of our participants clearly prefer items in the LS to items in the TL, mostly because they understand the LS better. One may argue that the students' opinions should not be overrated because they could have been motivated simply by the desire for a test that can be solved correctly with minimal effort no matter whether this test accurately reflects a reading or listening proficiency construct. In the results section we presented some answers that do indeed reflect such a tendency. However, this does not automatically exclude the test takers' opinions from being considered for validation purposes: One aspect of validity – face validity – is concerned with whether stakeholders (e.g. test takers) think, based on "subjective judgement rather than […] any objective analysis", that a test is "an acceptable measure of the ability they wish to measure" (ALTE Members 1998: 145). This is often rejected as "not [being] a true form of validity" (ibid.). Nevertheless, Hughes (1989: 27) argues that, if a test which lacks face validity is used, "the candidates' reaction to it may mean that they do not perform on it in a way that truly reflects their ability". Bachman (1990: 288) also concludes that "[t]he 'bottom line' in any language testing situation […] is whether test takers will take the test seriously enough to try their best […]". Thus, if the students in our context find the test unusually difficult (e.g. if the questions about the text are hard to understand), if they feel bothered by certain aspects (e.g. if more than

24 About 10 % of the Task Lab students (N = 57), asked in which language(s) they had first learned to speak, named one or more languages that were neither German nor Swiss German. Another 30 % indicated that they had first learned to speak both German or Swiss German and one or more other languages.

one language is used), or if they think that the test is not giving them the chance to show their true understanding (because they do not understand the questions), then they may not do their best, which reduces the validity of the results (see also Cox et al. 2019: 131 f.; Filipi 2012: 513). Therefore, the test takers' opinions, whether subjective or not, should carry some weight in the validity argument.

Finally, we must point out that our results are limited to a large-scale assessment context and cannot be applied directly to other settings. Both in the Task Lab project and during the ÜGK task development, we targeted a population of test takers that is relatively homogeneous with respect to their language skills: The students are usually reasonably proficient in the LS and lower-level learners of the TL. Thus, we could rightfully assume that our items would be easily understood in the LS whereas the same would not be true for the TL. In our specific context, it also makes more practical sense to use the LS for the test items, because our students do not learn according to the same curriculum: Since there is no core vocabulary and no common textbook, it would be very cumbersome to determine what precisely the test takers would be able to understand in the TL and what not. Also, preparing the learners for the test, as is often the case in the context of international language examinations and also in classroom-based assessment, was not feasible for various reasons.

6. Conclusion

Our results point to the conclusion that, at least for our target group, there is very little evidence that speaks against using items in the LS in reading and listening tests, and a lot in favor of it. In fact, a large majority of students prefer questions and answers in the LS and have no problem switching between the languages. If items in the TL are used, there is a risk that pupils cannot even start engaging with the text because they have not understood the question, or that they cannot demonstrate their understanding of the text because they do not understand the response options or are not able to formulate a short answer in the TL. Failing to understand, they may apply test-taking strategies that do not involve knowledge of the TL – or simply guess their answers. Furthermore, in our study, the analysis of differences between four variants of the same items led to

unexpected findings that mostly have to do with problems in the TL items. All of these sources of failure have very little to do with the ability to carry out the intended reading activities, and therefore introduce construct-irrelevant variance.

Overall, it seems that with the help of questions and answers in the LS, we can measure more reliably what we actually intend to measure: the ability to understand a variety of TL texts in various relevant ways.

Bibliography

Alderson, J. Charles (2000), *Assessing Reading*. Cambridge: Cambridge University Press.

Alderson, J. Charles/ Figueras, Neus/ Kuijper, Henk/ Nold, Guenter/ Takala, Sauli/ Tardieu, Claire (2006), Analysing tests of reading and listening in relation to the Common European Framework of Reference: The experience of the Dutch CEFR construct project. *Language Assessment Quarterly* 3: 1, 3–30.

ALTE Members (1998), *Multilingual Glossary of Language Testing Terms*. Cambridge: Cambridge University Press.

Bachman, Lyle F. (1990), *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bailey, Alison L./ Heritage, Margaret/ Butler, Frances A. (2013), Developmental considerations and curricular contexts in the assessment of young language learners. In: Kunnan, Anthony John (ed.) (2013), *The Companion to Language Assessment*. Chichester: Wiley Blackwell, 423–439.

Brantmeier, Cindy (2006), The effects of language of assessment and L2 reading performance on advanced readers' recall. *The Reading Matrix* 6: 1.

Council of Europe (ed.) (2001), *Common European Framework of Reference for Languages*. Cambridge: Cambridge University Press.

Cox, Troy L./ Bown, Jennifer/ Bell, Teresa R. (2019), In advanced L2 reading proficiency assessments, should the Question Language be in the L1 or the L2? Does it make a difference? In: Winke, Paula & Gass, Susan M. (eds.), *Foreign Language Proficiency in Higher Education*. Cham: Springer International Publishing, 117–136.

EDK (ed.) (2011), *Grundkompetenzen für die Fremdsprachen. Nationale Bildungsstandards*. EDK. [Online: http://edudoc.ch/record/96780/files/grundkomp_fremdsprachen_d.pdf. 24.08.2021].

Filipi, Anna (2012), Do questions written in the target language make foreign language listening comprehension tests more difficult? *Language Testing* 29: 4, 511–532.

Freedle, Roy O./ Kostin, Irene W. (1999), Does the text matter in a multiple-choice test of comprehension? The case for the construct validity of TOEFL's minitalks. *Language Testing* 16: 1, 2–32.

Gass, Susan M./ Mackey, Alison (2017), *Stimulated Recall Methodology in Applied Linguistics and L2 Research* (2nd edn.). New York & London: Routledge.

Godev, Concepción B./ Martínez-Gibson, Elizabeth A./ Toris, Carol C. M. (2002), Foreign language reading comprehension test: L1 versus L2 in open-ended questions. *Foreign Language Annals* 35: 2, 202–221.

Gordon, Claire M./ Hanauer, David (1995), The interaction between task and meaning construction in EFL reading comprehension tests. *TESOL Quarterly* 29: 2, 299–324.

Green, Anthony (2014), *Exploring Language Assessment and Testing: Language in Action*. London, New York: Routledge.

Hasselgreen, Angela (2005), Assessing the language of young learners. *Language Testing* 22: 3, 337–354.

Hughes, Arthur (1989), *Testing for Language Teachers*. Cambridge: Cambridge University Press.

In'nami, Yo/ Koizumi, Rie (2009), A meta-analysis of test format effects on reading and listening test performance: Focus on multiple-choice and open-ended formats. *Language Testing* 26: 2, 219–244.

Jeon, Eun Hee/ Yamashita, Junko (2014), L2 reading comprehension and its correlates: A meta-analysis. *Language Learning* 64: 1, 160–212.

Kane, Michael (2006), Validation. In: Brennan, Robert L. (ed.), *Educational Measurement* (4th edn.). Phoenix: Greenwood, 17–64.

Khalifa, Hanan/ Weir, Cyril J. (2009), *Examining Reading*. Cambridge: Cambridge University Press.

Kiefer, Thomas/ Robitzsch, Alexander/ Wu, Margaret (2015), *TAM: Test Analysis Modules*. [Online: http://CRAN.R-project.org/package=TAM. 24.08.2021].

Lenz, Peter/ Karges, Katharina/ Barras, Malgorzata (2019), Investigating test method effects in French L2 reading items for young learners. In: Huhta, Ari/ Erickson, Gudrun & Figueras, Neus (eds.), *Developments in Language Education: A Memorial Volume in Honour of Sauli Takala*. Jyväskylä: University of Jyväskylä & EALTA, 182–202.

Mayring, Philipp (2010), *Qualitative Inhaltsanalyse: Grundlagen und Techniken*. Weinheim, Basel: Beltz.

Messick, Samuel (1990), *Validity of Test Interpretation and Use*. Princeton, NJ: Educational Testing Service.

Messick, Samuel (1995), Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist* 50: 9, 741–749.

OECD (2014), *PISA 2012. Technical Report*. Paris: OECD Publishing.

Ozuru, Yasuhiro/ Briner, Stephen/ Kurby, Christopher A./ McNamara, Danielle (2013), Comparing comprehension measured by multiple-choice and open-ended questions. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 67: 3, 215–227.

R Core Team (2014), *R: A Language and Environment for Statistical Computing*. Wien: R Foundation for Statistical Computing.

Rodriguez, Michael C. (2003), Construct equivalence of multiple-choice and constructed-response items: A random effects synthesis of correlations. *Journal of Educational Measurement* 40: 2, 163–184.

Shiotsu, Toshihiko (2010), *Components of L2 Reading: Linguistic and Processing Factors in the Reading Test Performances of Japanese EFL Learners*. Cambridge: Cambridge University Press.

Shohamy, Elana (1984), Does the testing method make a difference? The case of reading comprehension. *Language Testing* 1: 2, 147–170.

SKBF (2019), Bildungsmonitoring. [Online: http://www.skbf-csre.ch/bildungsbericht/bildungsmonitoring/. 24.08.2021].

Urquhart, A. H/ Weir, Cyril J. (1998), *Reading in a Second Language: Process, Product, and Practice*. London, New York: Longman.

VERBI Software (2015), *MAXQDA 11*. Berlin: VERBI.

Wolf, Darlene F. (1993), A comparison of assessment tasks used to measure FL reading comprehension. *The Modern Language Journal* 77: 4, 473–489.

Ruth Trüb and Stefan D. Keller

Conceptualising and measuring writing in English as a Foreign Language at primary school

Abstract This article discusses the role and nature of writing in English as a Foreign Language (EFL) at primary school and highlights the importance of conceptualising writing as a meaningful communicative activity in both teaching and research. From the outset, teaching and assessing EFL writing should not only focus on linguistic features but also, and primarily, on the communicative effectiveness of a text. This approach is well established within communicative approaches to language teaching, yet equally important in the context of measuring and researching EFL writing competence. The way in which EFL writing is measured in research can have a considerable washback effect on how it is taught and assessed in schools. By taking the example of the research project 'An Empirical Study of EFL Writing at Primary School', the article illustrates the implications of such a construct definition for measuring EFL writing, in particular with regard to rating scale and task development. It also discusses its implications for teaching and assessing EFL writing in the primary English classroom.

1. The relevance and role of EFL writing at lower and upper primary school

Written language is part of children's lives from early childhood on, be it at home, at school or in the media (Cameron 2001). Due to technical developments such as internet technology and mobile communication networks, written communication has become increasingly important in private, educational and professional environments. According to Weigle (2002), "the ability to write effectively is becoming increasingly important in our global community, and instruction in writing is thus assuming an increasing role in both second- and foreign-language education" (p. 1).

In early foreign language learning, oral skills usually precede the written language and form the basis for developing reading and writing skills (McKay 2006: 221 f.; Szpotowicz & Lindgren 2011: 126). It seems

important, for example, that the learners first learn how to say and use new words before the written form is introduced to avoid mispronunciation and develop confidence in using the oral language (Brewster/ Ellis & Girard 2002; Cameron 2001; Legutke/ Müller-Hartmann & Schocker-von Ditfurth 2014). Special emphasis is placed on developing listening comprehension and speaking skills so that the learners quickly become able to use the language for oral communication (Jäger 2014). In the early stages of learning the foreign language, writing may mainly consist of copying words and simple phrases, which can be used to consolidate vocabulary and sentence structures, develop an awareness of the English written form and practise handwriting (Brewster et al. 2002). However, it is equally essential that the learners, from early on, experience writing as a meaningful communicative activity: "When children write as participants in authentic, meaningful activities, they not only learn about the technical aspects of encoding spoken words in print, but also how writing is used in social situations. As children begin to playfully take up their roles as writers in everyday experiences at home and school, they come to see themselves as writers" (Wells Rowe 2017: 261). In first language acquisition, even very young children who are not yet fully literate often use whatever resources and knowledge they have to produce a wide variety of 'texts' such as letters, menus, shopping lists, signs, plans, maps or receipts which they purposefully use in play and social situations (Worthington & van Oers 2015). They seem to do so with considerable enjoyment and commitment and usually receive plenty of praise for their effort, even if what they have written is difficult to read or does not follow common spelling conventions. It seems paramount, therefore, that even in the first years of foreign language learning at primary school authentic, motivating and meaningful communicative writing activities be integrated in teaching in order to maintain and develop a positive attitude towards literacy and the act of writing itself. At the later stages of primary school, learners become increasingly able to write longer texts if enough guidance is provided (Brewster et al. 2002; Cameron 2001). The learners may gain some first experience with planning, drafting, editing and revising a text and develop an awareness of different text types or genres (Cameron 2001). According to Boscolo (2009), research studies have shown that a key factor for maintaining motivation to write at upper primary and lower

secondary school is that writing be taught through meaningful activities in a supportive and collaborative writing environment and that the learners experience writing "as a tool by which various communicative objectives can be attained" (p. 309 f.). Hasselgreen, Kaledaite, Maldonado Martín and Pizorn (2012) similarly emphasise the importance that children at upper primary school be given the opportunity to share personal experiences and write for an audience. They argue that, with support such as pre-writing activities and appropriate feedback, writing can both be a personally rewarding experience for the children and "a major source of language development" (Hasselgreen et al. 2012: 19). It seems essential, therefore, that EFL writing at upper primary school is also conceptualised and implemented as a meaningful activity that has a clear communicative function, both in teaching and in assessment. An overly narrow focus on linguistic aspects such as orthography, syntax or grammar would not do justice to the communicative function and purpose of writing. Such a narrow focus would not help learners to see writing as a meaningful social activity and a desirable skill to learn. A frequently discussed question is what role spelling should play in primary EFL writing (see e.g. Hempel/Kötter & Rymarczyk 2017: 43 ff.). From our perspective, correct spelling is relevant where written products are to be 'published' or presented to some audience, but the pressure of formal correctness should not hinder the pupils' learning or create the impression that learning to write is the same as learning to spell. If EFL writing primarily focuses on learners' deficits, or on learners avoiding mistakes, it might both damage the learners' skills development and their motivation to write. It appears central, therefore, to establish a classroom environment where the learners can experience writing as a personally rewarding activity, and where orthography is regarded as a resource for writing as much as vocabulary, genre knowledge or writing strategies. In order not to overwhelm the young learners with overly high orthographic demands, it seems advisable to first help them copy words and phrases correctly from a text model in order to create their own personal texts, to teach them how to look up words in a dictionary or school book, and to develop their language awareness by drawing their attention to specific features such as the capitalisation of the 'I', the spelling of question words, or silent letters. This seems to be in line with the concept of 'focus-on-form' in communicative language teaching

as suggested by Ellis (2015), which "provides learners with the opportunity to take 'time-out' from focusing on message construction to pay attention to specific forms and the meanings they realize" (p. 4).

2. Theoretical background and challenges for teaching and researching EFL writing

Such a communicative and functional notion of language use stems from insights from second language acquisition research and developments such as communicative and competency-based language teaching. According to McKay (2006), there is general agreement in current second language acquisition research that

> in foreign and second language classrooms, children's language learning flourishes when there is a focus on meaning, and when their teachers and other visitors give them opportunities to interact in ways that reflect the wider discourse communities relevant to the language they are learning. Children learn to use language because the interesting activities in which they are engaged absolutely necessitate (from the child's point of view) cooperation and interaction. (p. 41)

The two approaches 'communicative' and 'competency-based' language teaching share a very similar view on language and language learning. Both approaches regard language as a means of communication with a particular function or goal (Richards & Rodgers 2014). Whereas communicative language teaching mainly focuses on the ability to communicate effectively and stresses the importance of using and learning the language in meaningful, authentic situations, competency-based language teaching focuses more on what the learners 'can do' with the language, i.e. the actual outcomes or competencies the learners should develop (Richards & Rodgers 2014). According to Richards and Rodgers (2014), these competencies "consist of a description of the essential skills, knowledge, attitudes, and behaviors required for effective performance of a real-world task or activity" (p. 156).

Both approaches use the term 'communicative competence' to describe their main aim of language learning (for a detailed discussion of this concept see, e.g., Bachman & Palmer 1996; Canale & Swain 1980; Hymes 1972). In terms of writing competence, it seems important that writing not only be regarded as a linguistic but also a pragmatic activity that needs to

fulfil a communicative function in concrete, everyday situations (Becker-Mrotzek & Böttcher 2014: 47). This emphasis on communicative competence is also found in the Common European Framework of Reference for Languages (CEFR) (Council of Europe 2001) and the educational standards in European countries which are based upon it. The CEFR defines communicative language competence as consisting of linguistic, sociolinguistic and pragmatic competences (Council of Europe 2001: 13). Hence, besides linguistic aspects, it also includes social aspects of language use such as politeness, and pragmatic aspects such as logical structuring or the awareness of genre-specific features and the communicative function of a text (Council of Europe 2001). The formal skills or 'resources' which underlie the ability to communicate are seen as subservient to the learners' ability to express their own ideas (D-EDK 2016: 11).

Such a communicative and competency-oriented view of language and language learning has concrete implications for teaching materials design, instruction and research. A major challenge is to create tasks and situations in which even elementary and pre-intermediate learners can acquire a complex skill like writing in the context of meaningful and communicative activities, and to identify relevant genres for writers at primary school to serve as models for writing. The term 'genre' refers to "abstract, socially recognized ways of using language for particular purposes" (Hyland 2003: 18). It comprises a shared understanding of what a specific text or discourse type usually looks like, which enables the user to read and write in an efficient and effective way (Hyland 2007). Generic competence, however, "is not simply about the ability to reproduce discourse forms; it is the ability to understand what happens in real-world interactions and to use this understanding to participate in real-world communicative practices" (Paltridge 2001: 7). Hallet (2016) argues that genres can be used to develop communicative competence at all ages, be it in form of a three-sentence-narrative or two-sentence-argument at primary school, or more complex texts at secondary school:

> From the beginning of foreign language teaching, pupils should be encouraged to produce meaningful written messages in social interactional situations. In concrete terms: Even in early foreign language learning at primary school, learners should be given opportunity to tell short stories, produce simple informative or explanatory texts or exchange arguments in a limited form. [unofficial translation] (p. 89f.)

He emphasises that the concrete implementation of a genre can be aligned with the learners' language level and that individual learners' differences will be evident in the level of complexity, elaboration and adequacy of their texts (Hallet 2016: 89–90).

A second challenge is developing instruments of educational measurement which are able to measure these differences in text quality. The instruments should be sensitive not only to aspects of language such as vocabulary, grammar or orthography but also capture functional qualities such as the ability to tell an interesting story or address a reader with a personal opinion. If these key elements of EFL writing are not relevant in the measurement of educational progress, they are in danger of being marginalised both at the level of classroom instruction and curriculum development.

The following sections describe how these challenges have been dealt with in the research project 'An Empirical Study of EFL Writing at Primary School'. First, the aims, context and methodology of the study will be briefly outlined, followed by a description of how the rating scales and tasks for the project were developed. Text samples will then provide insight into how well the learners were able to accomplish the tasks and, finally, there will be a brief discussion of implications for teaching, teacher education and materials development.

3. An Empirical Study of EFL Writing at Primary School: Description of an empirical research project

3.1 *Aims, context and methodology of the study*

The research project 'An Empirical Study of EFL Writing at Primary School' aimed at investigating the English writing skills of children in Year 6 at primary schools in Switzerland. It was carried out in the Canton of Aargau, where the teaching of English as a foreign language was introduced at primary school in 2008 (Husfeldt & Bader-Lehmann 2009) and where learners start learning English in Grade 3 when they are eight to nine years old. The study included 322 learners from 19 schools. It intended to conceptualise writing as a communicative activity in the way outlined above and measured writing competences both in terms of communicative/functional and linguistic qualities. In contrast to previous studies that

measured EFL writing at primary schools in Switzerland (Bayer & Moser 2016; Gnos 2012; Kreis, Williner & Maeder 2014), this study not only focused on how well the pupils achieved the goals set by the curricula but also intended to give a more detailed account of the quality and characteristics of the learners' texts. Its main research questions were:

(a) How proficient are the learners in EFL writing by the end of primary school, and what qualities do their texts display?
(b) How is primary EFL writing taught in the Canton of Aargau, and what is the learners' perception of EFL writing?
(c) What is the effect of different individual and educational factors on the learners' writing competence?

In order to answer these questions, the pupils completed two writing tasks, an e-mail and a story, which were rated independently by two raters using semi-holistic rating scales that were specifically developed for this project. The final scores were adjusted for rater severity, task difficulty, and difficulty of the rating scales using Many-Facet Rasch Measurement (FACETS, version 3.80.4, Linacre 2018) and a standard setting was carried out to align the learners' results to the CEFR language levels. Moreover, the learners and teachers were asked about their experience in the writing classroom to gain insight into the kind of instruction they used. Since many teachers had previously expressed their uncertainty about how to teach EFL writing at primary school in an age-appropriate way, it seemed desirable to inquire about teaching practices employed in order to make recommendations later on how these might be further used, developed or adapted. The study also investigated individual and educational factors such as the learners' self-efficacy, the frequency of text composition in class or the use of English at home and their effect on the learners' writing competence (not discussed in this article). The following sections describe how the definition of writing competence as described above has influenced the rating scale and task development in this project.

3.2 *Developing rating criteria for communicative EFL writing at primary school*

A well-defined construct definition is a key component of and an important basis for rating scale and task development. Hasselgreen et al. (2012)

emphasise that "assessment criteria for writing must reflect the consensus of what good writing is. And descriptors based on these criteria must reflect the age and ability of the writers for whom they are being developed" (p. 20). One should be aware that the selection of assessment criteria is a central decision for the validity of a test since they implicitly or explicitly represent the developer's view of writing competence (Weigle 2002). Not only the learners and their written products but also the rating scales, tasks and raters have an effect on the final result and, hence, the estimation of the learners' writing competence (Grotjahn & Kleppin 2017; Weigle 2002).

3.2.1 Holistic, semi-holistic and analytic rating systems

A first important decision was whether to apply a holistic, semi-holistic or analytic rating system. In holistic rating, the texts are read quickly and assessed on the basis of an overall impression on one global scale (Grotjahn & Kleppin 2017: 129). 'Semi-holistic rating' refers to the use of rating scales for an overall assessment of particular dimensions of writing competence such as content, style or language correctness (Schipolowski & Böhme 2016). In some contexts, these rating scales are also called 'analytic scales' (e.g. Grotjahn & Kleppin 2017; Weigle 2002). However, in order to clearly distinguish them from what will be referred to as 'analytic rating', the term 'semi-holistic rating' is preferred. The term 'analytic rating' is used to describe an assessment on the basis of a series of precisely defined, often dichotomous criteria such as reached/not reached or present/missing (Schipolowski & Böhme 2016) or scales that are based on discourse analytic measures such as the percentage of error-free sentences, average word length or number of spelling mistakes (Knoch 2009).

In our study, we opted for a semi-holistic approach for several reasons. First, one of the aims of the study was to provide detailed information about the learners' writing ability and text quality. A holistic rating system, therefore, would not have provided sufficient information. Secondly, it seemed important that the rating system would comply with a competency-oriented view of language learning that focuses on what the learners can do rather than their deficiencies. A semi-holistic approach with well-formulated descriptors appeared to better fulfil this criterion

than an analytic rating system that, among other aspects, would base its estimation of writing competence on aspects such as counts of spelling mistakes or grammatical errors. Neither did it seem possible to sufficiently operationalise a complex task such as communicative writing with only dichotomous criteria (McKay 2006: 267). A third aspect to consider was a possible washback effect in the schools. If the research project applied an analytic rating system, it would seem likely that the teachers follow an approach like this and assess their learners in a similar way. We saw here the danger of leading the teachers away from competency-oriented assessment, a concept which is currently being introduced in many cantons in Switzerland (e.g. BKS 2020; VSA 2016). Therefore, semi-holistic rating was favoured over a holistic and analytic approach.

3.2.2 Developing rating scales for the two genres e-mail and story

Starting point for developing the semi-holistic rating scales for the study were insights from genre analysis (Augst/ Disselhoff/ Henrich/ Pohl & Völzing 2007; Brinker/ Cölfen & Pappert 2018; Hyland 2004), an analysis of existing rating scales for assessing writing, in particular EFL writing at primary school (Brock 2015; Hasselgreen et al. 2012), and the collated representative samples of descriptors of language competences developed for young learners aged 11–15 (Szabo 2016). On this basis, the five dimensions *task completion*, *text structure and cohesion*, *syntax and grammar*, *vocabulary* and *language mechanics* were selected for use in the study. The five dimensions were further specified according to the genres employed (see Fig. 1):

Rating scales for the e-mail:
- **Task completion** (coverage, level of detail, genre-specific elements of an e-mail)
- **Text structure and cohesion** (coherence, cohesion)
- **Syntax and grammar** (complexity, correctness)
- **Vocabulary** (range)
- **Language mechanics** (orthography, punctuation)

Rating scales for the story:
- **Task completion** (communicative effect/creativity, level of detail)
- **Text structure and cohesion** (coherence, cohesion)
- **Syntax and grammar** (complexity, correctness)
- **Vocabulary** (range)
- **Language mechanics** (orthography, punctuation)

Fig. 1: *Dimensions of text quality for the genres e-mail and story*

As far as possible, the descriptors for the rating scales were developed on the basis of CEFR descriptors as referred to in Szabo (2016). Since the CEFR levels are rather broad descriptions of language competence, the Swiss national standards for foreign language learning and the cantonal curricula further differentiate the existing levels into e.g. A1.1/A1.2 and A2.1/A2.2 (D-EDK 2016; EDK 2011). They specify that the pupils should reach CEFR level A1.2 by the end of primary school in Year 6. This level, therefore, defined the midpoint of a 5-point rating scale (see Tab. 1):

Tab. 1: *5-point rating scale for assessing writing at levels A1 and A2 (for the detailed descriptors see the online material accompanying this volume (Link on page 137))*

0	1	2	3	4
Below	Approx. level A1.1	Approx. level A1.2	Approx. level A2.1	Above

The CEFR level descriptors provided a useful framework for formulating rating scale descriptors for the linguistic dimensions. There were, however, central dimensions of text quality that could not be based on existing descriptors. In particular, the pragmatic and genre-specific dimensions had to be developed by the research team. These were dimensions that measured whether and how well the texts fulfilled their communicative function and genre-specific aspects such as addressing the reader in an e-mail or the narrative structure of a story.

For the genre e-mail, the dimension *coverage* judged whether the reader was provided with the expected information, and the dimension *level of detail* assessed in how much detail this was accomplished. The dimension *genre-specific aspects of an e-mail* captured aspects such as greeting the reader, asking for information, concluding sentences (e.g. 'Please write back soon'), complimentary closes (e.g. 'Best regards'), the signature or smileys.

Regarding the genre story, the main pragmatic and genre-specific dimensions were *communicative effect/creativity*, *level of detail* and *coherence*. The dimension *communicative effect/creativity* captured to what extent the story fulfilled the function of entertaining the reader, e.g. by creating tension, capturing the reader's attention or taking a surprising twist. *Level of detail* gave account of the elaboration of the story and *coherence* focused on the logical organisation and genre-specific structure of the text with elements such as setting the scene, development/complication of the story, resolution and the story ending.

As far as possible, the descriptors for the linguistic rating scales (*cohesion*, *complexity* and *correctness of syntax and grammar*, *vocabulary, orthography* and *punctuation)* were aligned to the CEFR descriptors relevant for young learners aged 11–15 (Szabo 2016). As can be seen in Appendices A and B, the descriptors are formulated in a way that tries to highlight positive achievement rather than what the learners cannot do. They, hence, try to reflect a competency-based view of language learning that does not regard low levels as 'insufficient' or something negative but as first steps on the way to reaching higher language levels. Even at the lowest levels, where negative formulations can hardly be avoided, careful consideration was given to using non-offending language.

3.2.3 Revising and editing the scales

The initial rating scales underwent a thorough revision process. First, the scales were repeatedly discussed with different experts and trialled with 24 texts from a pilot study. In this revision process, the two initial dimensions *syntax* and *grammar* turned out to be ambiguous since they did not reliably differentiate between complexity and correctness. They were, hence, replaced by the dimensions *complexity of syntax and grammar* and

correctness of syntax and grammar. In addition, the wording of several descriptors was adjusted because it was ambiguous and lacked clarity. After this initial revision process, the two raters participated in an intensive rater training that followed a 'combined rater training and scale revision approach' as suggested by Harsch and Martin (2012). During the rating process, the raters took notes of any ambiguities they experienced. This information, together with the results of an interrater-reliability analysis, was used to develop a common understanding of the rating scales and to discuss any changes that seemed necessary. Special cases of uncertainty (e.g. whether missing reporting clauses for direct speech were to be rated among coherence or cohesion) were recorded in a separate document. A second piloting and intensive rater training with the revised scales and with benchmark texts was carried out before the main data collection.

Eventually, the texts from the main data collection were rated with an interrater-reliability between .78 and .94 (ICC, two-way mixed, single measure, absolute agreement) for all dimensions. Therefore, it was possible to rate even seemingly difficult dimensions such as *communicative effect/creativity* in a reliable way (ICC = .82). The final version of the rating scales with the detailed descriptors can be found in the online material accompanying this volume (Link on page 137).

3.3 *Developing tasks for communicative EFL writing at primary school*

A first step in task development was the selection of suitable genres for measuring the learners' EFL writing skills. According to Furger and Lindauer (2013), it is important not only to rely on one type of genre when measuring writing skills because the learners may display different levels of writing competence depending on the genre employed. The curriculum (D-EDK 2016) as well as the CEFR (Council of Europe 2001) and the collated representative samples of CEFR-related descriptors for young learners (Szabo 2016) mainly refer to descriptive texts, correspondence and simple narrative texts at the target levels A1 and A2. Therefore, it seemed reasonable to select a descriptive text in form of an e-mail and a narrative text to measure the learners' writing skills.

The selection of suitable genres was followed by the actual task design. If writing is seen as an activity in which the written language is used in a

communicative and meaningful way, it seems important in terms of content validity that this also be reflected in task design. According to McKay (2006), "young learners learn best through activities that are concrete and meaningful, and evidence of their language learning is most likely to be present in language use assessment tasks that have similar characteristics to those in the child's real world" (p. 100). This seems to be very much in line with the concept of performance assessment, which "involves either the observation of behavior in the real world or a simulation of real-life activity" (Weigle 2002: 46). Therefore, the tasks used in this study tried to reflect real-life situations and were set in a context that was familiar to the children. In the first writing task (Appendix A), the learners were asked to write an e-mail and introduce themselves to two children from Canada in the context of a (fictional) class collaboration project. In the second writing task (Appendix B), they had to write a picture story set in the context of a family meal and invent their own story ending.

A further aspect of task design was to create a close connection between the two tasks and the rating criteria. Since the ability to communicate effectively was part of the construct definition and represented in the rating criteria, the corresponding task had to give the learners the opportunity to display this ability. Therefore, the instruction for the e-mail did not only give directions on what information the learners should include but also invited them to ask the recipients questions. Similarly, the pictures of the story did not predetermine the whole story but gave the learners room to invent their own story ending and, hence, display their creativity. These aspects are reflected in the rating scales for the dimension *task completion*.

One of the main challenges of task design was to find and specify a narrative task that did not overwhelm the learners at level A1.1 and at the same time give high achievers the chance to display their abilities. In contrast to first language writing instruction, where narrative tasks seem to be predominant (Furger & Lindauer 2013), primary EFL writing appears to first focus on descriptive texts and correspondence since these two genres can be used with limited language resources and do not necessarily require wide-ranging vocabulary. When writing stories, much more imagination, creativity and personal experience come into play. Without appropriate support and a carefully selected stimulus, writing a story might be too difficult for learners at level A1 because they would lack much of

the vocabulary needed for expressing their ideas in the foreign language. Therefore, it seemed important to provide enough guidance so that also learners at lower levels would be able to successfully complete the task. This was done by selecting a story that was set in a family home and mainly required vocabulary in the word fields *family*, *rooms*, *food* and *activities at home*, which are usually dealt with in the first years of foreign language learning (e.g. D-EDK 2016). The main direction of the story was given through pictures but the story ending kept open so that the learners still had the opportunity to express their creativity.

The feasibility of the tasks was tested in a pilot study with a small sample from the target population of learners. Forty-three pupils including one dyslexic child and three children with individual learning objectives in the language of schooling were asked to give feedback on the perceived level of difficulty of the tasks after they had written the texts. Even though more than 25 % of the learners perceived the story as challenging, only three learners, or 7 %, indicated that the task was slightly too difficult for them. Therefore, the story seemed to be demanding but doable for most learners. In order to provide insight into how well the learners accomplished the task, some sample texts from the main data collection will be presented in the next section.

3.4 *Text samples*

The following texts illustrate the range of writing skills that could be observed in the study with reference to the levels of competence as described in the rating scales (see the online material accompanying this volume (Link on page 137)). The examples are taken from the narrative task 'story' (see Appendix B). The learners who wrote these texts were all in Grade 6 and in their fourth year of learning English. None of them was a native English speaker. Large differences in terms of text length can be observed, ranging from 52 words in Text 1 to 316 words in Text 5. Of more interest, however, seem to be the differences in text quality.

> the dad is food and *er sieht das der* boy not hir is the dad gou and *sucht den* boy the dad loock in this book the mom *sieht dous er nicht do* is the mom *sagt geh sucht* Dad *und er sit das er den BLiest und dann wird er saur.*

Text 1: *Approaching A1.1*

The first text (see Text 1) was written by a learner with learning difficulties. Even though much of the text is written in German (italics) and several aspects such as the communicative effect, level of detail, syntax and grammar, vocabulary or punctuation are very limited or not yet existent, it can be seen that the learner knows some words in English (e.g. 'dad', 'food', 'boy', 'gou' or 'loock') and tried to formulate English sentences. He or she uses 'and' to link sentences and the orthography is good enough so that most English words are comprehensible.

> The fater and the mater food, dont the Kid ried a Book. The fater say: „come" The Kid came, the fater ried th Book, the mater say the Kid „going and bring de fater „Yes mom" say the Kid. The Kid coms not beak. „Mom comon its cool", the mater coing and then read the Book.

Text 2: *Level 1 (approx. A1.1)*

The language in the second text (see Text 2) is generally comprehensible but the author does not fully succeed to develop a storyline. Some parts seem to be missing and the reader sometimes has to guess what might have happened between the scenes. There is some simple direct speech such as 'Mom common its cool' but besides this, there is not much detail that would make the story vivid. In terms of language, there are few very basic cohesive devices ('and' and 'and then'). The text consists of very simple sentences and grammatical structures, often used inconsistently with some reduction and omission of elements (e.g. 'the mater coing'). Vocabulary and orthography are limited (e.g. 'The fater and the mater food') but the errors generally do not affect comprehension. The learner already seems to know a lot about punctuation, using full stops, commas, colons and quotation marks, usually following the punctuation rules of the German

> It was a hard day, the man and the woman was at the desk he want to eat. But were are the doughter? the man go to the sleeping room of the doughter. Then when the man the doughter find, he sais: come arn you mast go to eat. The dougther go to her mom. But the dad look in the book of the doughter. A long time ago, the mom sais doughter go and look where are your dad. The doughter go agan in his bed room, when they was in his bed room he see his dad. The dad read in the book of the doughter! Dad mom sais you mast go tho eating, and why do you read in my book in my book!?...

Text 3: *Level 2 (approx. A1.2)*

language.

Text 3 shows some more elaboration. There is a simple storyline and the text contains some elements that catch the reader's interest such as the introductory sentence 'It was a hard day', the question 'But were are the doughter?' or the story ending 'and why do you read in my book in my book!?'. In comparison to Text 2, there are more cohesive devices ('but', 'then' and 'when') and also some pronouns such as 'he' and 'her'. However, the use of cohesive devices and reference words does not yet seem to be firmly in place, as can be seen in the incorrect use of 'A long time ago', 'his', 'they' and 'he' in lines 4 and 5. The learner already uses some more complex grammatical structures and sentences, e.g. subordinate clauses and some past tense forms, but still inconsistently and with limited control (e.g. 'where are your dad' or 'they was'). The text consists of mainly basic vocabulary. Short and common words are written with reasonable phonetic accuracy but there are still basic errors (e.g. 'come arn' = come on; or 'tho' = to). Basic punctuation is often used in a correct way but quotation marks for direct speech are missing.

It was time for dinner and dad was so hungry but mom tells him that he have to search his son and the he can eat. Dead was so hungry and the meal looks so delicius that dad went to search his son. After 2 minutes dad findes his son in his room, his son was reading a book but dad calls him to eat. As the son was seating ont the chair his Mother asks him: „ Where ist dad? Go and search him! And the son beginns to search him then he finds him in his room reading his book. The son asks his dead: „What are you doing. Im reading your book. But are you not hungry. saids the son. It haves spaghetti. No im not hungry. But it haves also Pizza, saids the son. Ok im hungry. Then at the table where they eat dad ask were ist the Pizza. Wich Pizza, it has no Pizza. Dad was said and the son was laughing.

Text 4: *Level 3 (approx. A2.1)*

Text 4 contains several elements that make the story interesting. The fact that dad first has to search for his son before he can eat, even though he is so hungry and the meal looks so delicious, clearly creates some tension. Also, the way how the son tries to persuade his father to come to dinner, cheating on what food there will be, shows creativity and the ability to entertain the reader. The learner clearly succeeds to tell a story. In comparison to Text 3, the author of this text has a good command of simple cohesive devices and reference words. The learner already uses some complex sentences and grammatical structures, e.g. the third person '-s' and the present and past continuous, but still makes basic errors (e.g. 'he have' or 'it

haves'). The text contains mainly simple but also some specific vocabulary (e.g. 'delicius'). The expressions 'It haves spaghetti' and 'it has no pizza' show some influence from the German language. In terms of orthography, the learner shows a good command of common words and even of some difficult words such as 'laughing', even though there are still errors such as 'seating' instead of 'sitting' or 'im' instead of 'I'm'. There are different elements of punctuation but sometimes even basic punctuation such as full stops or question marks are missing or not correct.

Hi, I'm Tom and I wanna tell you a funny story that happened last week. My mom and dad were eating lunch. I was in my room while reading a fascinating book. It was so intresting that i forget the time. My dad saw, that i didn't was at my seat and asked my mom. She told him to go and tell me that lunch was finished. I was a bit disappointed that I had to stop reading but suddenly i got up and went to our kitchen. My dad saw the book, that I were reading and took a look at it. He was also intrested so he started reading. My mom and me were siting at the tabel and suddenly I realised that my dad wasn't there. My mom told me to go and look where he was. I stand up and went to my room. I was a bit angry cause I just had started eating.
At the moment I got in my room, I saw my dad at the floor and reading my book. That wasn't right! I had to stop reading but he was alaud to read? No! I told him to get up and go to the kitchen. But he ignored me. Suddenly my mother came because she was waiting a long time for us. He got up and went to the kitchen. My mom wanted to put my book back in my self, but she saw an intresting picture and started to read. Me and my dad were therewhile sitting by the table and waiting for her but she didn't came. We went to my room and took a look where she has beent and saw her on the floor by reading my book. That was so funny because she were angry when we were reading the book. But it was so intresting that everyone was fascinated by it.

Text 5: *Level 4 (approx. A2.2 and higher)*

The last example (Text 5) shows one of the most elaborated texts from the study. The story contains all elements of a simple narrative structure: First, there is a clear setting of the scene with an introduction to the main characters and the time and place where the story takes place. Then the story develops, culminating in a considerable complication when the son gets angry because his father does what he himself has been denied and even refuses to follow him back to the kitchen. The mother first seems to be able to resolve the problem but then gets absorbed in the book herself, which causes a second complication that is finally resolved when the narrator realises the humorous side of the situation. In particular at the beginning of the second paragraph, the author succeeds to create tension and make the story interesting: 'I saw my dad at the floor and reading my book. That wasn't right! I had to stop reading but he was alaud to read? No! I told

him to get up and go to the kitchen. But he ignored me.' In addition, the different scenes, people and actions are described in detail with the effect that the reader can well imagine the situation and understand why the people react the way they do. The text is linguistically well linked and even uses some difficult cohesive devices (e.g. 'suddenly', 'at the moment I got in my room' or 'while sitting and waiting…'). Even though there are some grammatical errors (e.g. 'she didn't came'), the learner shows a good command of simple and complex grammatical and syntactical structures. There is mainly simple but also some specific vocabulary such as 'disappointed', 'suddenly', 'alaud to' or 'ignored me'. Orthography and punctuation are mostly correct.

In general, the text samples reveal a wide range of text qualities, both in terms of communicative-functional and linguistic text features. It becomes clear that limiting writing to word and sentence level in primary EFL classes and research contexts would not do justice to the wide range of abilities that can be observed. Neither would the use of only linguistic assessment criteria do justice to the communicative effectiveness and function of the texts. It hence seems paramount that such insights be considered when teaching and measuring EFL writing at primary school.

4. Conclusion and outlook

This paper aimed at demonstrating the process of planning and implementing a study of EFL writing at primary school. We discussed how EFL writing at primary school might be conceptualised. It was argued that writing should be defined and implemented as a meaningful activity with a clear communicative function and that such a construct definition is essential for developing young learners' communicative writing competence and maintaining a positive attitude towards writing.

For the study, a system of rating scales and two tasks which operationalised the construct of writing were developed. The rating scales operationalised the different dimensions of communicative writing competence at proficiency levels relevant for primary EFL learners (CEFR levels A1 and A2) and focused on pragmatic, simple sociolinguistic, and linguistic aspects. Special emphasis was placed on communicative-functional dimensions such as to what extent a text fulfils its communicative function, e.g. to inform or entertain the reader. The rating scales for linguistic dimensions were developed on the basis of CEFR descriptors as

referred to in Szabo (2016) and careful consideration was given to formulating the rating scale descriptors in a positive and competency-oriented way. The two tasks, an e-mail and a story, were closely linked to the assessment rubrics. We highlighted the importance of designing tasks in a way that gives the learners the opportunity to display not only linguistic but also communicative-functional competences. The story, for example, contained two empty picture frames at the end to encourage the learners to find an original, individual ending to their story. For learners at primary level, this involves taking risks with the language, stretching their repertoire of expressive tools to the limit in order to be creative with the language. While such a focus on meaningful and creative language use seems to be highly desirable, we also emphasised that the tasks should not overwhelm the learners. It was argued that careful consideration needs to be given to the selection of appropriate stimulus material which elicits everyday language, is situated in a familiar context and gives the learners conceptual guidance, e.g. through visuals.

Text samples from the study illustrated how learners at different levels coped with the demands of the narrative task and how wide their range of abilities was. We saw that even learners at lower levels with limited language proficiency were able to write texts which – for an attentive reader – fulfilled basic communicative functions.

The research study will provide further insights which are hoped to be useful for developing EFL writing at primary school. By providing the distribution of learners across ability levels and further information about the quality and characteristics of the learners' texts, the study will give some indication of what can be expected from primary EFL learners by the end of Year 6 (this will be published in the dissertation study associated with this project). Additional insights are expected from questionnaires and interviews about how EFL writing is currently taught in schools and perceived by teachers and learners. Empirical evidence like this will form an important basis for making concrete suggestions in terms of teaching methodology and materials design, which are outlined briefly in the following.

First, it seems important to supplement the development of oral skills with meaningful writing activities that have a clear communicative function and are able to develop and maintain the learners' motivation to write. Learners should be given opportunities to gain experience with

different genres and think about the process of writing and various writing strategies. This may involve studying model texts and using process-oriented approaches in which learners draft their texts and then receive feedback from peers and teachers on how to develop their texts. This feedback would also need to take into account communicative-functional aspects such as creativity, originality, or adequacy. Finally, the approach taken in this project suggests assessing learners' products in a competency-oriented way, using communicative, genre-specific assessment criteria such as the ones developed in this study.

Implementing a methodology like this entails a number of challenges which teachers, teacher educators, and material developers will have to deal with. First, there will be the challenge of designing age- and language-level-appropriate tasks that offer the opportunity to use the language in a communicative way and at the same time provide enough support so that all learners have a good chance to complete them successfully. This will go hand in hand with the question of how to cater for learners as diverse as observed above and, lastly, it needs to be discussed how rating scales such as those used in the study can be employed and adapted for the assessment of written products in schools.

As outlined in the introduction, we believe that such an emphasis on the communicative use of language in combination with appropriate support and guidance is paramount for the further development of EFL writing at primary school. We hope that this paper, along with the concrete examples of tasks and rating scales used in the study, will be an inspiration for teachers, teacher educators, material developers, and researchers to further think about and discuss the role and nature of EFL writing at primary school.

Bibliography

Augst, Gerhard/ Disselhoff, Katrin/ Henrich, Alexandra/ Pohl, Thorsten/ Völzing, Paul-Ludwig (2007), *Text – Sorten – Kompetenz. Eine echte Longitudinalstudie zur Entwicklung der Textkompetenz im Grundschulalter*. Frankfurt am Main: Peter Lang.

Bachman, Lyle/ Palmer, Adrian (1996), *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.

Bayer, Nicole/ Moser, Urs (2016), *Evaluation der Englischkompetenzen im Kanton Aargau. Englischkompetenzen auf der Primarstufe und auf der Sekundarstufe I. Schlussbericht zuhanden des Departements Bildung, Kultur und Sport des Kantons Aargau*. Zürich: Institut für Bildungsevaluation. Assoziiertes Institut der Universität Zürich. [Online: https://www.ag.ch/de/weiteres/aktuelles/medienportal/medienmitteilung/medienmitteilungen/mediendetails_73793.jsp. 28.07.2021].

Becker-Mrotzek, Michael/ Böttcher, Ingrid (2014), *Schreibkompetenz entwickeln und beurteilen* (5th edn.). Berlin: Cornelsen Schulverlage GmbH.

BKS (2019), *Handreichung – Kompetenzorientiertes Unterrichten und Beurteilen*. Aargau: Departement Bildung, Kultur und Sport des Kantons Aargau. [Online: https://www.schulen-aargau.ch/media/schulen-aargau/unterricht/pruefen-beurteilen/bksvs-hr-beurteilen-vs.pdf. 28.07.2021].

Boscolo, Pietro (2009), Engaging and motivating children to write. In: Beard, Roger; Myhill, Debra; Nystrand, Martin & Riley, Jeni (eds.), *The SAGE Handbook of Writing Development*. London: SAGE Publications Ltd, 300–312.

Brewster, Jean/ Ellis, Gail & Girard, Denis (2002), *The Primary English Teacher's Guide* (New ed.). Harlow: Pearson Education.

Brinker, Klaus/ Cölfen, Hermann/ Pappert, Steffen (2014), *Linguistische Textanalyse: Eine Einführung in Grundbegriffe und Methoden* (8th edn.). Berlin: Erich Schmidt Verlag.

Brock, Rainer (2015), *Handbuch Writing (2014 und 2015). Kommentierte Schreibperformanzen 2014 und 2015: Gesamtbeurteilungen. Richtlinien zur Bewertung von Schreibleistungen (6. und 7. Schulstufe)*. Salzburg: Bundesinstitut für Bildungsforschung, Innovation & Entwicklung des österreichischen Schulwesens (BIFIE). [Online: https://www.bifie.at/node/3148. 20.08.2021].

Cameron, Lynne (2001), *Teaching Languages to Young Learners*. Cambridge: Cambridge University Press.

Canale, Michael/ Swain, Merrill (1980), Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics,* 1(1), 1–47.

Council of Europe (2001), *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.

D-EDK (2016), *Lehrplan 21 – Sprachen*. Luzern: Deutschschweizer Erziehungsdirektoren-Konferenz.

EDK (2011), *Grundkompetenzen für die Fremdsprachen: Nationale Bildungsstandards*. Bern: EDK. [Online: https://edudoc.ch/record/96780. 27.08.2021].

Ellis, Rod (2015), The importance of focus on form in communicative language teaching. *Eurasian Journal of Applied Linguistics,* 1(2), 1–12.

Furger, Julienne/ Lindauer, Thomas (2013), Schreiben in einer interaktiven Web-Umgebung. Schreibkompetenzen und Schreibmotivationen modellieren und erfassen. In: Rosebrock, Cornelia & Bertschi-Kaufmann, Andrea (eds.), *Literalität erfassen: bildungspolitisch, kulturell, individuell*. Weinheim: Beltz Juventa, 40–54.

Gnos, Christina (2012), *Englischkompetenzen am Ende der 6. Klasse: Überprüfung der Lernziele. Sprachstandserhebung im Kanton Luzern*. Luzern: Dienststelle Volksschulbildung des Kantons Luzern.

Grotjahn, Rüdiger/ Kleppin, Karin (2017), Kriteriale Evaluation von Schreibkompetenzen. In: Akukwe, Bettina/ Grotjahn, Rüdiger & Schipolowski, Stefan (eds.), *Schreibkompetenzen in der Fremdsprache. Aufgabengestaltung, kriterienorientierte Bewertung und Feedback*. Tübingen: Narr Francke Attempto Verlag, 117–157.

Hallet, Wolfgang (2016), *Genres im fremdsprachlichen und bilingualen Unterricht. Formen und Muster der sprachlichen Interaktion*. Seelze: Klett, Kallmeyer.

Harsch, Claudia/ Martin, Guido (2012), Adapting CEF-descriptors for rating purposes: Validation by a combined rater training and scale revision approach. *Assessing Writing,* 17(4), 228–250.

Hasselgreen, Angela/ Kaledaite, Violeta/ Maldonado Martín, Natàlia/ Pizorn, Karmen (2012), *Assessment of Young Learner Literacy Linked to the Common European Framework of Reference for Languages*. Strasbourg: Council of Europe Publishing. [Online: http://www.ecml.at/tabid/277/PublicationID/63/Default.aspx. 28.07.2021].

Hempel, Margit/ Kötter, Markus/ Rymarczyk, Jutta (2017), *Fremdsprachenunterricht in der Grundschule in den Bundesländern Deutschlands*. Berlin: Peter Lang.

Husfeldt, Vera/ Bader-Lehmann, Ursula (2009), Englisch an der Primarschule: Erfahrungen aus der Schweiz. In: Engel, Gaby; Groot-Wilken, Bernd & Thürmann, Eike (eds.), *Englisch in der Primarstufe – Chancen und Herausforderungen. Evaluation und Erfahrungen aus der Praxis*. Berlin: Cornelsen Verlag.

Hyland, Ken (2003), *Second Language Writing* (5th edn.). Cambridge: Cambridge University Press.

Hyland, Ken (2004), *Genre and Second Language Writing*. Ann Arbor: University of Michigan.

Hyland, Ken (2007), Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing*, 16(3), 148–164.

Hymes, Dell Hathaway (1972), On communicative competence. In: Pride, John Bernard & Holmes, Janet (eds.), *Sociolinguistics. Selected Readings*. Harmondsworth: Penguin, 269–293.

Jäger, Astrid (2014), Die Förderung kommunikativer Fähigkeiten im Englischunterricht der Grundschule. In: Böttger, Heiner (ed.), *Englisch. Didaktik für die Grundschule* (2nd edn.). Berlin: Cornelsen Schulverlage GmbH, 112–122.

Knoch, Ute (2009), *Diagnostic Writing Assessment. The Development and Validation of a Rating Scale*. Frankfurt am Main: Peter Lang.

Kreis, Annelies/ Williner, Michaela/ Maeder, Christoph (2014), *Englischunterricht in der Primarschule des Kantons Thurgau. Schlussbericht der Evaluation*. Kreuzlingen: Pädagogische Hochschule Thurgau. [Online: http://www.sprachenunterricht.ch/themenbereich/forschungsberichte-und-evaluation. 28.07.2021].

Legutke, Michael/ Müller-Hartmann, Andreas/ Schocker-von Ditfurth, Marita (2014), *Teaching English in the Primary School*. Stuttgart: Klett.

Linacre, John Michael (2018), *A User's Guide to FACETS: Rasch-Model Computer Programs [Software Manual]*. Chicago: Winsteps.com.

McKay, Penny (2006), *Assessing Young Language Learners*. Cambridge: Cambridge University Press.

Paltridge, Brian (2001), *Genre and the Language Learning Classroom*. Ann Arbor: University of Michigan Press.

Richards, Jack/ Rodgers, Theodore (2014), *Approaches and Methods in Language Teaching* (3rd edn.). Cambridge: Cambridge University Press.

Schipolowski, Stefan/ Böhme, Katrin (2016), Assessment of writing ability in secondary education: Comparison of analytic and holistic scoring systems for use in large-scale assessments. *L1 Educational Studies in Language and Literature,* 16, 1–22.

Szabo, Tunde (2016), *Collated Representative Samples of Descriptors of Language Competences Developed for Young Learners Aged 11–15 Years*. Strasbourg: Council of Europe. [Online: http://www.coe.int/en/web/portfolio/overview-of-cefr-related-scales. 28.07.2021].

Szpotowicz, Magdalena/ Lindgren, Eva (2011), Language achievements: A longitudinal perspective. In: Enever, Janet (ed.), *ELLiE. Early Language Learning in Europe*. Manchester: British Council, 125–143.

VSA (2016), *Kompetenzorientiert fördern und beurteilen*. Solothurn: Volksschulamt Kanton Solothurn. [Online: https://www.so.ch/fileadmin/internet/dbk/dbk-vsa/Foerdern/Schullaufbahn/2016_brosch%C3%BCre_kompetenzorientiert_beurteilen.pdf. 28.07.2021].

Weigle, Sara Cushing (2002), *Assessing Writing* (6th edn.). Cambridge: Cambridge University Press.

Wells Rowe, Deborah (2017), The value of writing in early childhood. In: Kucirkova, Natalia; Snow, Catherine; Grover, Vibeke & McBride, Catherine (eds.), *The Routledge International Handbook of Early Literacy Education. A Contemporary Guide to Literacy Teaching and Interventions in a Global Context*. London, New York: Routledge, 251–263.

Worthington, Maulfry/ van Oers, Bert (2015), Children's Social Literacies: Meaning Making and the Emergence of Graphical Signs and Texts in Pretence. *Journal of Early Childhood Literacy,* 17(2), 147–175.

Illustrations

Der Schmöker. Images by Erich Ohser. Public Domain. From Plauen, E. O. (1949), *Vater und Sohn*. Band 1. Konstanz: Südverlag.

Appendices A + B contain additional information as an attachment. Please contact orders@peterlang.com and transmit the following activation code: PL22Dx61G. You will then receive the attachment by e-mail.

Appendix C: Writing task e-mail

Write an e-mail

Imagine that your class has started an e-mail project with a school in Canada. On the photos, you can see Sophie and Jacob from Canada. They are in grade 6, like you. Write an e-mail to them and introduce yourself.
Stell dir vor, deine Klasse macht ein E-Mail Projekt mit einer Schule in Kanada. Auf den Fotos siehst du Sophie und Jacob aus Kanada. Sie sind in der 6. Klasse, genau wie du. Schreibe ihnen ein E-Mail und stelle dich vor.

Write about the following aspects and add more if you can.
Schreibe über folgende Aspekte. Füge wenn möglich weitere hinzu.

- Your name and age
  Dein Name und dein Alter
- What school subjects you like and why.
  Welche Schulfächer du gerne hast und weshalb.
- What you like doing in your free time. Describe two activities in detail.
  Was du in deiner Freizeit gerne machst. Beschreibe zwei Dinge etwas genauer.
- What would you like to know from Sophie and Jacob? Ask questions.
  Was möchtest du von Sophie und Jacob wissen? Stelle ihnen Fragen.

Appendix D: Writing task story

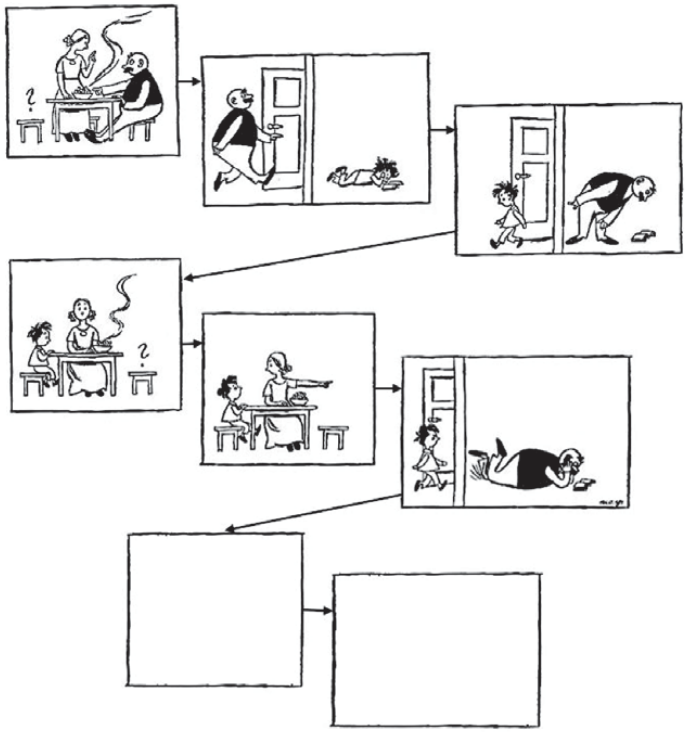
Write a story

Look at the pictures. What happened at this family meal last week? What did the characters say and do? How did the story end?
Schau dir die Bilder an. Was geschah in dieser Familie letzte Woche beim Essen? Was sagten die verschiedenen Personen und was machten sie? Wie ging die Geschichte zu Ende?

On the basis of these pictures, write a coherent story with a witty ending. Include enough details so that the story becomes vivid and interesting for the reader. Write the story in the past tense.
Schreibe auf der Grundlage dieser Bilder eine zusammenhängende Geschichte mit einem witzigen Ende. Schreibe so detailliert, dass die Geschichte für den Leser lebendig und interessant wird. Schreibe die Geschichte in der Vergangenheitsform.

The fascinating book

Das spannende Buch



Sandie Mourão, Carolyn Leslie, Maria Afredo Moreira
and Estela Monteiro

Battling against a traditional assessment culture: The case of early English learning in Portugal

Abstract This chapter addresses the particular case of Portugal and its culture of assessment in early foreign language (FL) education and the implications this has for teachers working with younger children. After briefly addressing the principles and procedures for assessment in early FL learning, we provide information about the Portuguese education system, its English education in the 21st century, teacher education and teacher placement. These details are described thoroughly as they are relevant for the reader to understand the constraints set out in the rest of the chapter. We continue with a description of a study, its methodology and data collection tools. The discussion around selected results centres on the issues perceived as problematic related to traditional assessment practices which focus on assessing literacy skills rather than oral skills. Our conclusion suggests that the specificities of the Portuguese education system and its reliance on specialist teachers of English brought from other levels of education make this issue particularly difficult to overcome. The recommendations regarding adequate teacher education are relevant for other countries with similar practices.

1. Assessment in early language learning

Assessment, seen here as the "principled ways of collecting and using evidence on the quality and quantity of people's learning" (Rixon & Prošić-Santovac 2019), has long been a cantankerous issue in language education. It was only at the end of the 1990s that an agenda for assessment in early language learning began to take shape (Rea-Dickins 2000) and its pedagogical relevance is now recognized.

Traditional assessment methods, associated with "paper and pencil tests", are considered "problematic" (Pinter 2006: 131) for young FL

learners, mainly because they do not show what children can do with the language they are learning. Typical objectives, and thus activities for the young learner classroom, contribute to the development of attitudinal goals and motivation towards FL learning through approaches which include singing songs, reciting rhymes and riddles, listening to stories and engaging in playful game-like activities – all of which provide age-appropriate, meaningful opportunities for oral language use. Unfortunately, age-appropriate assessment practices have not accompanied the fast-moving introduction of English at ever earlier ages. Rixon and Prošić-Santovac (2019) suggest that this is due to a variety of reasons which include inadequate planning during policy implementation; considering assessment at odds with the playful approaches implied in many early FL learning programmes or a concern that inappropriate assessment may be detrimental to young learners' motivation.

Nevertheless, as the 21st century moves into its third decade and teaching methodologies develop, age-appropriate assessment practices are beginning to take clearer forms. 'Assessment of learning' includes a variety of procedures and tools for teachers to enable them to judge children's learning. These can be either formative or summative in approach and include, but are not restricted to:

- classroom-based structured assessment activities, allowing children to engage in authentic tasks to demonstrate skills and abilities (Ioannou-Georgiou & Pavlou 2003; McKay 2006),
- systematic, planned observations guided by rubrics (Cameron 2001; Diehr 2009; Ghosn 2019; McKay 2006),
- spontaneous, incidental observations ideally gathered into anecdotal records, which are usually short, non-judgmental narrative accounts (Cameron 2001; Ghosn 2019; Ioannou-Georgiou & Pavlou 2003),
- document analysis, based on in-class or out-of-class paper-based activities (Cameron 2001),
- traditional tests, which may be teacher-made.

'Assessment for learning' (Black & Wiliam 1998; Popham 2008) is an essential component of formative assessment, and involves the learners

themselves engaging in reflecting on learning and recognizing what they can or cannot do, with a view to thinking how they might improve and get better. This involves assessment procedures such as self-assessment, peer-assessment and portfolio assessment, procedures which are considered possible in the young learner classroom (CILT 2001; Ellis & Ibrahim 2015) and contribute to a learner-centered approach to teaching as well as contributing to more autonomous language learners.

Jang (2014: 112) lists a number of recommendations for appropriate assessment practices in early language learning. We are in agreement with much of what she says, and the following manifesto for assessment in early language learning is adapted from her list. Assessment should:

- mirror the ways children learn and enjoy learning,
- be congruent with our educational philosophy and be useful to meet our teaching goals,
- identify strengths and areas that need improvement,
- serve as a tool for self-regulation,
- shape children's life-long learning and affective attitude towards assessment.

School-based assessment in Portugal is traditional in approach – test-oriented, directed towards grading and selecting students, rather than towards providing feedback to improve learning (Conselho Nacional de Educação 2015). An OECD report from 2014 suggests that "formative assessment is not used systematically" (OECD 2014: 12). Our own personal involvement in teacher education in Portugal has alerted us to the fact that the recognized age-appropriate approaches to assessment of English described above are not as prevalent as could be hoped. This resulted in our undertaking a nationwide study 'Assessment in Primary English Education in Portugal: Perceptions and practices'. This study is based on a selection of results which look in particular at the mis-match between learning objectives which focus on orality, and actual assessment practices which focus on the development of literacy. Before introducing the reader to the study, an overview of the Portuguese education system is provided for those who are not familiar with the context.

2. The Portuguese education system

In this section, the Portuguese education system is introduced together with a short historical overview of English as a FL in the curriculum and information about teacher preparation to teach English to children under 10 years old. Teacher placement is also relevant for this chapter, so a short description of how this occurs is also included.

2.1 *Compulsory state education in Portugal*

Formal, compulsory education in Portugal begins in September of the year a child turns six years old and comprises three sequential cycles of Basic Education together with secondary education (as can be seen from Fig. 1). The focus of this chapter is on 1st cycle education, with children from 6–10 years old and comprised of Grades 1–4. Teachers in 1st cycle education are generalist teachers, responsible for teaching all subjects except English, which is taught by a peripatetic specialist teacher. From the 2nd cycle, education is subject specific, and children only have specialist subject teachers.

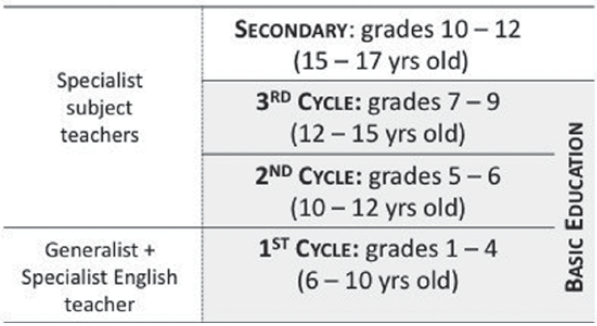


Fig. 1: The education system in Portugal

The public-school network in Portugal has been organised around school clusters since 2006. Schools of one or more education levels are grouped together under a centralised leadership, the stated objective of which is to facilitate transition between cycles, as well as to overcome geographical and social seclusion (OECD 2014). It is most common to group the first three cycles, along with pre-primary institutions, although sometimes a

secondary school is included. A school cluster can include any number of 1st cycle schools depending on their geographical location. Some 1st cycle schools can be quite far away from both the central school cluster building and each other, sometimes covering areas of over 50 kilometres.

2.2 *Early English language education*

The teaching of English as a FL has been part of Portuguese education system since the educational reform of 1989, when English or French were both options as a first FL in the 2nd cycle. However, from the mid-1990s, English became more popular and by the 21st century, it was the main choice of most children entering this cycle of education. In September 2012, it became compulsory to study English in the 2nd cycle, and French, Spanish and sometimes German became an option upon entering the 3rd cycle. In September 2015, English was introduced into the national curriculum in Grade 3 of the 1st cycle.

This recent appearance of English in the 1st cycle national curriculum is deceiving, for English had been part of 1st cycle activities in a less formal context since the beginning of this century. In 2001, the Portuguese Ministry of Education (ME) issued a law establishing that "1st cycle schools may, according to available resources, provide the initiation of a foreign language, with an emphasis on oral expression" (Decree-Law 6/2001: 260 – own translation). This legal framework, however, entailed some constraints – English was to be seen as an extracurricular activity, carried out after school hours, attendance was optional and there should be no associated cost for parents. Later, in 2005, the ME announced a more formal 'Generalisation Program of English Teaching in the 1st Cycle' which integrated the teaching of English as an extracurricular activity into Grades 3 and 4 of the 1st cycle and was supported by a set of national guidelines (Bento, Coelho, Joseph & Mourão 2005). Nevertheless, this renewed legislation presented a clear tension between the principles established in the legal documents, outlining a need to promote "the level of training and qualification of the future generations by acquiring fundamental skills […], namely in the generalisation of English teaching since the 1st cycle" (Decree-Law 14753/2005: 9785 – own translation), and the curricular organisation which remained non-statutory.

In 2006, the ME (Despacho nº 12591/2006) settled that 1st cycle schools were obliged to provide English in 1st cycle from Grade 3, even though parents were not required to enrol their children, as attendance remained voluntary. Later in 2008, this compulsory provision was reaffirmed and extended to Grades 1 and 2. The profile of the teachers of English employed during this time was varied. Many were originally trained to teach in 2nd and 3rd cycles, or even secondary education and, though specialist teachers of English, lacked the pedagogical knowledge required to work with children under 10 years old. Most were employed by external bodies (e.g. local town councils, or legalized after-school activity-centres), were on short-term contracts, and very badly paid. It should be emphasized that there were no programme changes made to English in the 2nd cycle, which was still considered the legitimate start to learning English in the Portuguese curriculum.

Official figures state that between 2008 and 2014 the percentage of children learning English in Grades 3 and 4 of the 1st cycle was over 85 %, reaching as high as 91.2 % in the academic year of 2012/2013 (DGEEC 2014). Anecdotal evidence from 2nd cycle teachers of English at that time, lamented a change in children's attitudes towards English. The only in-depth study to look at the impact of the 'Generalisation programe' on 2nd cycle English education included a survey of 407 teachers of English (Bernardo 2014). A total of 85 % of these teachers indicated that children did not appear more motivated towards English, neither were their results any better than before the programme had begun.

Reasons for children's less than satisfactory attitudes upon reaching the 2nd cycle were highlighted between 2006 and 2012 in annual inspection reports and include the facultative nature of English in the 1st cycle, under-qualified teachers, the precarious nature and professional status of 1st cycle teachers of English, lack of parental understanding and support for the programme, and no clear approach to transition within the school clusters (Bastos & Brites 2012; Gregório, Perdigão & Casa-Novas 2014). By the time English was formally introduced into the 1st cycle curriculum in September 2015, its status as a subject was extremely low. Mainstream 1st cycle teachers and families saw it as an activity associated with informal learning and undertaken outside curriculum time while children waited to

be picked up from school. Formal testing had been forbidden, and anecdotal evidence suggests that teachers tended not to use any other methods of collecting evidence of learning, despite the existing guidelines (Bento et al. 2005).

2.3 *Curricular English in the 1st cycle*

Since 2015, curricular English has been governed by a set of national standards, which define the skills and domains to be developed, as well as designating a set of topics to be covered in each grade. These standards outline directives regarding listening, spoken production and interaction, reading and writing, lexis and grammar, and the intercultural domain, with exit profiles for both Grade 3 (equivalent to pre-A1) and Grade 4 (equivalent to A1). The standards also make it plain that the main priority is to develop listening and spoken production and interaction skills,

> Learning in the 1st cycle privileges orality. It is a learning that is gradually consolidated, beginning with listening and repetition, progressing to simple situations of spoken production and interaction in conjunction with reading and writing. (Bravo, Cravo & Duarte 2015: 3 – own translation).

Expectations, stated in the exit profiles, for Grades 3 and 4 with regards to reading and writing, the topic of this volume, are restricted to identifying known words and reading short sentences with these words in Grade 3, as well as copying words to label images or complete sentences. In Grade 4 this extends to understanding short sentences and simple texts with known words and completing short texts with known words, as well as writing about oneself and one's preferences (Bravo et al. 2015).

There is, however, no syllabus or similar document produced by the Portuguese ME that includes:

> The aims, selection and sequence of contents to be covered, mode of delivery, materials to be used, learning tasks and activities, expected learning objectives or outcomes, and assessment/evaluation schemes. (UNESCO-IBE 2013: 53)

As such, specific aims and objectives have not been outlined, and official directives for assessment are non-existent. The only document for 1st cycle English that remotely resembles a syllabus is the guidelines employed during the extracurricular English activities between 2005 and 2014 (see

Bento et al. 2005). As a result, English teachers rely heavily on textbooks that are used at all levels of education[1].

2.4 *The preparation of teachers of English for the 1st cycle*

Following the compulsory introduction of English in Portuguese 1st cycle education in 2015, as described above, a new recruitment group for teachers of English was created by the Portuguese ME for placement purposes. However, with less than a year to qualify, teachers were accepted with a range of profiles. Just one group of teachers had direct access to this recruitment group, those that had taken a four-year pre-Bologna bachelor's degree in basic education with a specialization in English and Portuguese, together with teachers who had a post-Bologna master's in Teaching English and another foreign language in Basic Education. Most applicants needed to requalify and as a result, several short-term qualifying courses opened specifically to enable teachers to re-train. These included an online course, 'Certificate in Primary English Language Teaching' (CiPELT), run by the British Council, and a short-term, complimentary course taught through higher education institutions (HEIs).

The CiPELT and the complimentary courses were both targeted at specialist teachers of English, already qualified to teach English in other cycles of education, as well as generalist teachers with a C1 level in English. The complimentary courses provided differentiated modules, depending on the teacher's original training: these included primary English methodology (for all teachers), English language and culture (for generalist teachers), and children's language development (for specialist teachers). From 2015 onwards, an 18-month pre-service master's qualification 'Teaching English in 1st Cycle of Basic Education' also opened in 13 HEIs with a two-semester practicum in 1st cycle schools. This master's is targeted to candidates that come from a degree course of English-related subject disciplines (e.g. language, linguistics, literature, culture etc.) and reinforces

1 Information on adopted textbooks is available at https://www.dge.mec.pt/lista-de-manuais-escolares-adotados (28.07.2021).
In the school year of 2018/19, *Let's Rock* (Abreu & Esteves 2015), one of six text books, was adopted by 2490 1st cycle schools, private and public, representing around half of those in mainland Portugal.

the relevance of a policy regarding the preference for specialist teachers of English in 1st cycle education.

2.5 *The placement of teachers of English in the 1st cycle*

Education in Portugal is centralized, and the large majority of teachers are part of the civil service with teacher placement, for those teachers who are not permanently associated with a school cluster, organised annually through a nation-wide placement system. Teachers who are not placed can be contracted through local placement tenders; however, these are shorter, yearly or part-time contracts. With unfavourable demographics and cuts in public services, including education, the majority of teachers of English in 1st cycle are contracted annually, which results in instability for both teachers and the children in their care. As contracted teachers, teachers of English in 1st cycle usually start working in the school cluster at the beginning of September and will miss the preparations for the academic year which took place in the previous June and July. Their precarity also results in lack of status amongst the school staff, and 1st cycle teachers of English usually have little chance of making changes to decisions taken by colleagues before they arrived at the school cluster.

All teachers in Portugal, including the teachers of English, are allocated to a school cluster which then places them into specific 1st cycle schools. Because English is taught only in Grades 3 and 4, and there are only two sessions of 50–60 minutes a week, each teacher of English can have up to nine class groups to ensure a complete timetable. A full 1st cycle class may have up to 26 children and mixed grade groups of Grades 3 and 4 together are also very common. Teachers of English in 1st cycle can therefore be responsible for around 230 children.

Logistically, as peripatetic teachers, 1st cycle teachers of English not only move between classes but may also have to move between 1st cycle schools, as the different groups of children may not be in the same building, but in several different schools all belonging to the same school cluster. As mentioned earlier, some of these 1st cycle schools can be tens of kilometers apart and require that teachers drive between them as part of their job. In addition, they may also be the only teacher of English in their school cluster, so it can be a rather solitary job.

3. The study

The study 'Assessment in Primary English Education in Portugal: Perceptions and practices', from which we are taking some of the data, began in October 2017 and terminated in December 2019. It includes the following objectives:

- to analyse perceptions, approaches and practices used in assessment of primary English education in Portugal;
- to identify constraints in relation to the national standards and a transformative vision of language education.

It has been our intention in this study to move from the overarching to the specific, to gather information from a variety of sources, and relate this to the specific history of English language education in 1st cycle education in Portugal. As such we looked at school cluster assessment criteria documents – public documents which aim to inform stakeholders, provide a common reference for schools and their teachers, and which should recognize the specificities of each school subject and its level. We have also obtained data from teachers of English via an online survey, attempting to move closer to actual classroom practice. However, recognising that perceptions may not actually be practices, we have also interviewed teachers of English and analysed their individual assessment instruments. What follows is a more detailed description of the study, selected results, and discussions that are relevant to this edited volume.

3.1 *Method*

We adopted a mixed-methods approach to the study, as our aim was both to collect quantitative national data on assessment and qualitative data regarding perceptions and evidence of practices. A total of three data collection procedures were used in the study: a grid to analyse cluster assessment documents, a teacher survey, teacher interviews, and content analysis of the assessment instruments used by teachers in the classroom.

3.1.1 The school cluster assessment criteria documents

The first stage of data collection involved the analysis of state school cluster assessment criteria documents from mainland Portugal, which

were accessed online through school cluster websites. The number of websites consulted represented 79 % of those that have 1st cycle schools. Assessment criteria documents for 1st cycle English were found on just 212 websites (32.5 %). These documents related to three different academic years and were analysed taking into consideration the Council of Europe's descriptors of language competencies for young learners (Council of Europe 2016) and national guidelines (Bento et al. 2005; Bravo et al. 2015). This analysis included a focus on the scope (key competencies, skill areas and domains), assessment instruments and processes, objectives, relationship with the 1st cycle assessment criteria and other relevant information. Inter-rater reliability was ensured, and results were expressed as percentages.

3.1.2 The teacher survey

A teacher survey, which ensured anonymity and confidentiality of respondents, was designed with the following objectives: to characterise teacher perceptions of assessment in 1st cycle English education and to understand the reasoning behind choices regarding assessment in primary English education. We also wanted to indirectly collect evidence of approaches and practices regarding assessment in primary English education and identify constraints in relation to national standards and a transformative vision of language education.

The survey, created using SurveyMonkey©, was written in Portuguese, and distributed online with the help of the Portuguese Association of English Teachers, HEIs and Facebook, and teachers of English in the 120-recruitment group employed during the academic year 2017–18 were invited to respond. The survey consisted of 42 open and closed questions, organised into seven sections designed to collect personal and professional information, data on teachers' current professional situation, information related to both general and specific assessment procedures, information on assessment instruments and processes, and teachers' attitudes and opinions on assessment in 1st cycle English education. In addition, a final section invited respondents to record any further comments or observations on assessment in primary English education and to take part in post-questionnaire interviews. A 5-point Likert Scale was used to analyse attitude

statements and results were expressed as percentages. Quantitative analysis was also carried out on closed questions and results were again expressed as percentages. Open questions were analysed qualitatively by reading and re-reading responses and identifying common themes.

3.1.3 The teacher interviews

A semi-structured interview was created to gain a further insight into responses from the teacher survey and four pilot interviews were carried out face to face in May 2018. Slight adjustments were made on the basis of these pilot interviews and the final version was used as the actual interview. At the time of writing this chapter, 10 state school 1st cycle teachers from the North and the Centre of the country had been interviewed, either face to face or via skype. Interviewees were asked to sign a declaration consenting to the audio-recording of the interview, and the anonymity of the interviewee and the school cluster were guaranteed. The interview consisted of 22 questions and was divided into four sections relating to personal information, assessment procedures and practices, and general questions on assessment of English in 1st cycle education. Teachers were asked to bring assessment documents with them or send them by email, and during the interviews they were asked to explain how these were used. Interviews were then transcribed and quotes from interviews and examples from assessment documents have been used to provide a greater understanding of the results.

3.2 Results and discussion

This section looks at each of the data collection procedures and discusses the results with a view to highlighting the mis-match between a required focus on the development of oral skills and the actual use of assessment through written tests.

3.2.1 Analysis of school cluster assessment documents

As mentioned previously, a total of 212 school cluster assessment criteria documents were analysed according to the criteria described above with a view to understanding each school cluster's expectations for teachers of English in the 1st cycle. There are no national guidelines for these

documents and so each school cluster creates their own at the beginning of the academic year and makes it available to the school community. These documents are supposed to guide practice as well as inform stakeholders.

In general, the analysed cluster documents provided guidelines which were common to both Grades 3 and 4, with only 10 % of documents referring to year-specific assessment criteria. The documents varied in content, length and specificity. Not all had a specific section for 1st cycle English, neither did they all include reference to the skills and domains, their weighting, the assessment tools that should be employed, or a set of descriptors indicating exit profiles.

Regarding mention of skills and domain areas, approximately 70 % of the analysed cluster documents made mention of the four traditional skills: listening, spoken production, reading and writing. Spoken interaction was mentioned less frequently (61 %), and lexis, grammar, and the intercultural domain the least frequently, 42 % and 32 % respectively. Our concern here is an almost equal balance given to literacy skills, to the detriment of oracy that should be the main focus.

Just over three-quarters of the analysed school cluster assessment documents included assessment instruments for English and the most common was the written test, referred to on 77 % of occasions. Other tools were alluded to less frequently, e.g. classroom tasks (43 %), observation notes and grids (39 %). Less common still were student-led procedures such as peer- and self-assessment (12 %), portfolios (5 %) and questionnaires (1 %). The actual weightings for tests were mentioned in a third of these analysed documents, and the majority indicated tests were worth over 40 % of the learners' final mark.

Although 43 % of the analysed school cluster assessment documents showed that different assessment instruments were used in the assessment of English and other school subjects, just under a third of documents indicate that the same assessment instruments were used across the 1st cycle, and these are usually in the form of written tests. This serves to highlight how learners are being assessed on their reading and writing skills in the FL, in common with assessment practices for all the other curricular areas.

Our analysis of the school cluster assessment documents suggests that, overall, expectations regarding assessment of English in the 1st cycle neither match the national standards nor recommendations in the literature.

It would appear that this group of learners are principally assessed on their reading and writing skills to the detriment of their spoken production and interaction skills and through a highly traditional, summative format.

3.2.2 The teacher survey

Responses from the teacher survey funnel the information provided from the cluster assessment documents and impart a view of possible classroom practices in relation to the cluster directives. The survey also provided a window into teachers' thinking and positioning on the issue of assessment.

Responses were collected between January and September of 2018, and only complete surveys were analysed for this chapter, totaling 134 in all. Unfortunately, this is not representative of the approximately 1,000 teachers of English in the 1st cycle. Nevertheless, it gives an insight into a possible teacher profile and professional attitudes towards the assessment of English in 1st cycle education in Portugal.

The Survey Respondents. As was expected, the respondents were mostly women (94,8 %), around two-thirds were between 31 and 40 years old, and just under a third were between 41 and 50, representing a younger than average group of teachers[2]. Just over three-quarters had more than 10 years of teaching experience, with nearly half of these reporting over 15 years. The vast majority were also qualified to teach older learners (94 %), confirming they were retrained specialist teachers of English. In fact, just over 10 % of the respondents revealed that they also taught in the other cycles of basic education and in secondary education.

Just under 20 % of the respondents were those teachers with direct access into the recruitment group, 50 % had retrained through the complimentary course, and 19 % had taken the online CiPELT course. The remainder had the accepted master's qualification. A total of 53 % of the respondents had been teaching in the 1st cycle since the beginning of the curriculum change in September 2015, but a far greater number (around 80 %) had taught English in 1st cycle during the years before English became curricular. Nearly 70 % of the respondents were on annual contracts, which

2 Recent OECD (2019) figures report the average age for teachers in Portugal at 49, with 47 % being over 50 years old.

meant that they were likely to move to another school cluster in the following year. Further, just over half the respondents had nine class groups, and a handful indicated having 10 or 11. Regarding mixed grade groups, 66 % registered having at least one such group.

Regarding training in language assessment, 35 % declared they had had no specific training. Nevertheless, respondents indicated high levels of self-assurance about assessment, showing they felt confident or extremely confident regarding their capacity to assess learning (99 %), their capacity to adapt assessment tools from other sources (98 %) and their capacity to design their own assessment tools (96 %).

Assessment Procedures and Tools. Respondents were asked to register if they agreed with the assessment procedures in their school cluster. Just under two-thirds revealed they did, 25 % agreed in part and 12 % disagreed. Around a third of the respondents (49) provided justifications for their opinions, and a significant number (20) included reference to:

- what they considered was inappropriate use of formal approaches to assessment, e.g. tests and written assignments;
- their concern that English was being assessed in the same way as other subjects in the 1st cycle, or in the same way as English in other cycles, thus ignoring the specificity of learning a FL at a young age.

Regarding the skills and domain areas, nearly all the respondents indicated assessing the four skills (listening, spoken production, reading, and writing). However, listening and writing were selected by slightly more respondents (98.5 and 97.8 %) than reading and spoken production (95.5 and 94.8 %). As with school cluster documents, fewer respondents selected spoken interaction (85.2 %), lexis and grammar (70.9 %), and the intercultural domain (59.7 %).

Figure 2 shows the teacher survey responses alongside the school cluster document analysis in relation to instruments and processes. The survey respondents indicated they used a wider range of instruments and processes and classroom activities with classroom tasks, observation grids and tests considered the most important procedures to assess learners. Learner-led assessment (self-/peer-assessment) was also signalled far more by survey respondents (just over 60 %), as well as document analysis (40 %) and portfolio assessment.

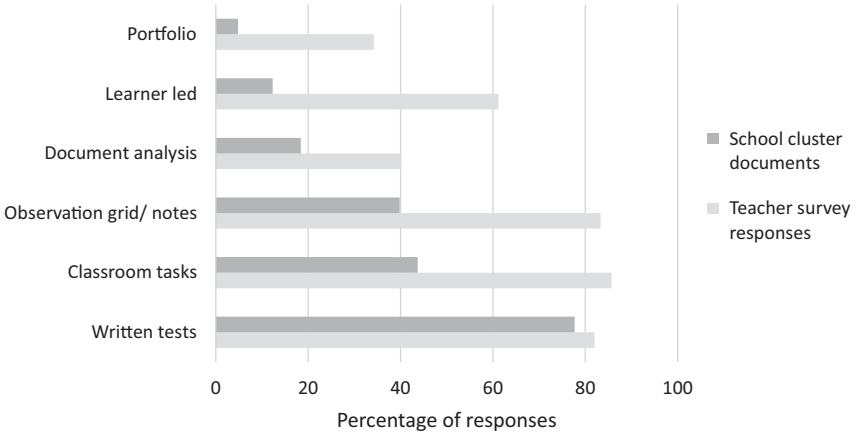


Fig. 2: Assessment instruments and processes in school cluster documents and teacher survey responses

Respondents were asked to indicate which assessment instruments they used to assess the different skills and domains. Classroom tasks were selected most widely for all the skills and domains. This contradicts the data from the school cluster assessment documents, which designated using tests more often.

Tests were employed for the assessment of all skills, but over 85 % of the respondents indicated using tests in particular for the assessment of writing, grammar and lexis, and listening. Reading, a skill which is developed at word level in Grade 3 and sentence level in Grade 4, is easily included in classroom activities, e.g. matching word cards with picture cards, circling target language in songs, rhymes and short texts or reordering cut up words into sentences (Pinter 2006: 71). These activities lend themselves to paper and pencil-based tests as well, yet fewer than half the respondents indicated using tests to assess reading, but instead preferred using classroom tasks.

In a classroom which values orality, one would assume that the systematic use of observation grids or notes would be important for collecting evidence of learners' ability to listen, produce and/or interact in English. However, these instruments were not as prevalent as expected in association

with these skills. Respondents indicated they used systematic observation in particular for collecting evidence of spoken production (78.4 %) and reading (70.8 %). Listening and spoken interaction were selected by fewer respondents (approximately 58 % each).

Issues with Assessing Reading. It may be worth discussing the data which suggests that reading is more regularly assessed during classroom activities and systematic observation. Our interpretation of this is that teachers, in their survey response, are considering 'reading aloud' or 'oral reading' as an activity to assess reading. School cluster documents also referred to 'reading' in association with oral skill development, presumably referring to reading aloud. McKay (2006: 234) states that it is considered inappropriate to over-use reading aloud as "an informal assessment tool in the classroom", mainly as it can have negative effects on learners' confidence and motivation, due to its focus on pronunciation. In addition, reading aloud is not considered part of the 'spoken production' category in the CEFR Companion Volume (Council of Europe 2018), where spoken production is described as a "'long turn', which may involve a short description or anecdote, or may imply a longer, more formal presentation" (p. 68). The descriptors developed for young learners (Council of Europe 2016) make reference to learners' saying, retelling, reciting etc. which are certainly not practiced through reading aloud. Reading should involve matching the spoken and written forms of words and short routine phrases at this level (Pinter 2006).

Teachers' Opinions and Attitudes. 80 % of the survey respondents agreed or totally agreed that the way they assessed their 1st cycle learners was in consonance with their personal convictions around assessment in this context. Fifteen respondents left a comment to this question. All but three reflected on the excessive use of tests and formal assessment approaches which is evident in their school clusters, an example being:

> The assessment criteria for Grades 3 and 4 are grouped together and give too much value (or the same value) to writing to the detriment of the other three skills [listening, spoken production and reading], and worksheets and summative assessment to the detriment of routine classroom tasks and observation grids. (Survey respondent – own translation)

Several comments highlighted the imposed directives, which these respondents disagreed with, or the attitudes of their mainstream colleagues, who they stated did not recognize the relevance of other forms of assessment beyond the written test:

> I neither agree nor disagree, I do not have sufficient autonomy to assess as I want to, as I have to follow the department / school cluster directives. (Survey respondent – own translation)

> There is so much work to do which starts with sensitizing the mainstream 1st cycle teachers and parents. Different assessment 'formats' are not always understood as assessment because assessment is seen as exclusively summative and written. If English has another format, it is seen as a pastime with no curricular relevance. (Survey respondent – own translation)

Our analysis of the teacher survey suggests that classroom practice go rather further than the school cluster assessment documents dictate. There is still an emphasis on using tests, mostly to assess writing, lexis and grammar, listening, and to some extent reading. Assessment of reading is mostly associated with classroom tasks and direct observation. Results point to teachers generally feeling satisfied with the way they assess English in the 1st cycle.

3.2.3 The teacher interviews

All 10 teachers interviewed were highly experienced teachers, all but one with previous experience as a 1st cycle English teacher, either as an extracurricular or curricular activity. Four interviewees were on yearly contracts, thus still in a professionally precarious position. Upon arrival at their designated school clusters, assessment documents were already approved, so they did not have an opportunity to participate in the definition of assessment criteria and weighting. Still, most indicated they agreed with the weighting given to literacy skills, which ranged from 30 to 50 %.

As we anticipated, what is actually happening in practice does not follow the school cluster assessment criteria documents. The interviewed teachers seem to be very aware that teaching and its assessment should focus on oral skills, for all but one used observation grids for spoken production and interaction activities. They were also aware of the relevance of formative assessment, as again all but one used self-assessment activities, and two used peer assessment. However, generally, student self-assessment

was undertaken at the end of a unit or term and focused on affective domains, behavior, and overall performance. Even though formal testing of the learners' listening, reading, and writing occurred once or twice per term, it was evident that these teachers tried to diversify their assessment strategies and instruments – e.g. taking notes on learners' writing, or using structured assessment tasks to collect evidence of development in reading and writing. However, these interviewed teachers recognised the prevalence of traditional approaches to assessment and the testing of literacy skills in their practice, a 'tradition' difficult to counteract, but also 'easy' to implement and one they considered 'objective'.

> Students shouldn't do tests and continuous assessment is more correct. But (…) it's understood in school that students have to do tests, that's how parents prefer things and that's what the 1st cycle teachers expect. There needs to be change but a gradual change because a test is a 'moment' and some students get very stressed with tests. It's much better to assess their daily work than assessing them on a few moments. (Interview with Teacher C1 – own translation)

> Tests are the normal system that people use. It has always been important that the teacher has some written evidence – something visible, palpable as a base for assessment. (…) It's important that the students have a more serious moment of assessment, so they understand how things work and that they have to work for a particular objective and for this formal assessment they study and work at home. (…) And it's an important time for the students – it is a motivation engine. (Interview with Teacher C2 – own translation)

As was noted in the teacher survey results, the interviewed teachers were critical of being 'forced' to follow prescriptive assessment procedures that did not make much sense to them but were imposed by the school clusters – for example, having to give two tests a term, or implementing the same written test for all classes on the same day or at the same time. One interviewee reflected on the influence of other cycles, both on a personal level – bringing traditional practices with her from her former days as a teacher of secondary students – but also because English was seen as a specific subject to be assessed equally throughout the school cluster:

> We are still contaminated by that which is done in other cycles, we have a path to be explored and mentalities to change – ours, my own, parents', the class teachers' (Interview with Teacher S1 – translated)

Others stated that after so many years of English as an extracurricular activity, which had lowered the status of English, traditional testing and

standardized procedures were seen as ways to encourage learners to study and take English 'seriously':

> Children are not conditioned to work without the purpose of improving grades. There is a long way to go. (Interview with Teacher S1 – translated)

> It's necessary for them to know that English is important because some students and parents think that English in Grades 3 and 4 is just to 'play' like the extracurricular activities (…) So it's important to tell them – this is a test, we're not here to play around. (Interview with Teacher C3 – translated)

Those teachers who were critical of formal testing admitted they missed those 'playful activities with the children', adding that too much testing was harmful to children. They believed the child's achievement should be regarded as a whole and that children should not be faced with formal assessment tasks focusing on discrete language items. They described more informal ways of assessing reading and writing, which included taking notes while children completed these tasks (also evident in the teachers' survey results). However, it is not clear whether all teachers fully understood what reading and writing (and even spoken production) is actually about. The following excerpt confirms the mis-match, or misunderstanding, discussed earlier between reading and reading aloud.

> When assessing reading and writing, a test is not needed; an assignment would suffice, as they already do it when they [orally] present their work (…) they are already working on writing, are already working on reading (…) When we are assessing speaking, students are already doing reading. (…) I have to check reading, pronunciation, whether they read quickly or with difficulties, whether they hesitate, and this is not easy (…). Reading is not easy to listen to. (Interview with Teacher M1 – own translation)

Upon analysis, the different activities designed to assess reading and writing were appropriate to the learners' level and to the national guidelines, e.g. matching words or sentences with images; numbering sequences of events; reproducing / adapting modelled texts; completing sentences with grammar or lexical items or dialogues with words or sentences already given; labelling objects, colours, or sketches expressing feelings; ordering scrambled letters or parts of a sentence, etc. However, there was also evidence of remnants of a structural approach in the strong emphasis on form or accuracy, for some teachers mentioned taking points away for learners' spelling errors, despite the fact that writing with correct spelling is not an

objective for 1st cycle English. These teachers found it difficult to distance themselves from a grammar-centred approach and the focus on accuracy associated with their earlier practices as specialist teachers in 3rd cycle and secondary school.

Many of the assessment instruments came from the textbooks used in English lessons. These included listening tests, observation grids for speaking production and children's self-assessment documents. Additionally, even though other instruments came from teacher-focused social network groups on Facebook, these groups were often associated with a textbook author who promoted their own materials. Analysis of the textbooks was not included in this study; however, due to its strong regulatory function, the approach to learning in the textbook has an important role to play in the teachers' actual practices, "[b]y setting out what needs to be done, what should not be done is simultaneously dictated" (Littlejohn 2012: 293).

Our analysis of the teacher interviews suggests that some teachers try to diversify and move away from traditional assessment procedures which focus on writing in particular. They use what experience they have, join informal communities of practice, like Facebook groups, and appear fairly confident in their abilities to create their own assessment instruments, justifying choices according to external constraints like parental pressure, while implicitly recognizing their own difficulties in moving beyond a psychometric approach to assessment.

3.3 *The actual battle – a discussion around results*

It has been evidenced throughout the chapter that the use of paper and pencil tests is predominant, and reading and writing are given more emphasis overall regarding skill development, despite national guidelines in Portugal for 1st cycle English indicating otherwise. This results from a prevailing culture of traditional expectations regarding assessment and our study highlighted the difficulties felt by the 1st cycle teachers of English on a variety of levels. Mainstream colleagues exert pressure into conformity with 'established' norms and procedures and the fact that there is no official syllabus for curricular English, with clear guidelines on teaching and assessment procedures to appropriately help children in their knowledge

construction processes in a FL, does not support easy adjustments. In addition, parental pressure for testing and grades is a force to be reckoned with, as parents find tests easier to understand and interpret than narrative descriptors based on evidence collected through observations or other more formative approaches to teaching and assessment for learning. Children often request tests in English, as they are accustomed to them in other school subjects. Furthermore, there is the influence of the established assessment practices from other cycles that convey the message that fairness is equal to decontextualised assessment throughout the school cluster, regardless of level, learning purposes, children's readiness and individual teacher practices.

Nevertheless, traditional cultures of assessment were not the only issue blocking teachers' path towards age-appropriate assessment practices. The survey respondents and interviewed teachers also felt that the recourse to traditional forms of assessment was caused by their difficult working conditions. These include multiple classes, low numbers of contact hours with children, isolation as well as precarity and low professional status. This merits further discussion.

The large number of classes and students per class make it harder to collect evidence of learning through diversified assessment instruments. Our results suggest that even though classroom tasks are a popular procedure to collect evidence, observation grids or notes as an instrument to accompany observation of classroom tasks are less used. This may indicate that classroom tasks are paper-based, thus focus on collecting evidence of writing in particular, exacerbating the focus on writing to the detriment of the other skills. The short amount of time allotted to English in the 1st cycle (two lessons of 50–60 minutes) adds to the difficulty, as it provides little room for the implementation of more formative, diversified and individualized assessment, so instruments are curtailed to the minimum, resulting in the over-use of a written test. In addition, opportunities to reflect with others about teaching and assessment practices are few due to the peripatetic nature of their positions in 1st cycle schools – moving between schools does not support the creation of moments for interaction, even informal ones when the school cluster is big enough to have more than one teacher of English in the 1st cycle.

The precarity of being contracted teachers, probably the most culturally specific issue, is a serious constraint to transformation. Not only are contracted teachers perceived as lower-status alongside the permanent teaching staff at the school cluster, but belonging to a recruitment group which, as yet, has not found its niche within the education system, brings about a variety of challenges. These teachers are 'in a no man's land', as they are neither teachers of English nor 1st cycle teachers. As such, teachers of English in the 1st cycle have no clear professional identity, and this deepens their isolation as a socio-professional group.

Precarity also results in late placements which culminate in teachers arriving after major decisions have been made regarding important pedagogic issues. Their low-status as a contract teacher, but also of a subject which is widely misunderstood by 1st cycle mainstream teachers and 2nd cycle teachers of English, provokes distrust. Alongside this issue, the fact that the majority acquired their qualification for teaching in this level through short-term qualifying courses results inevitably in incomplete training. Compounded with the problems associated with anything that is not seen as 'naturalised' professional practice, the result is traditional testing, focusing on judging children's acquisition of literacy skills over their ability to use English to interact with others in the classroom.

Curricular English in 1st cycle is still in its infancy and there is much to do. Policy as yet does not support appropriate assessment practices; training is required and the educational community needs time to acknowledge and conciliate that which is new. The original purpose of this study was to better understand the constraints of this new curriculum change, regarding assessment practices, but it was also an endeavour to support this new socio-professional group of teachers in their complicated trajectory to become accepted and find professional communities of practice in schools. This, in turn, will contribute more positively to guiding the process of knowledge construction in early English education in Portugal. At the end of the interviews, some teachers expressed the need to find this community, to be listened to and heard and to become better teachers of English: "This conversation has enabled me to unburden and think about

things I'm not happy with yet. I want to improve!" (Teacher interview S1 – translation). As this chapter evidences, there is still a long way to go, and our hopes are that this study may constitute a small stepping-stone in the right direction.

4. Lessons learnt from the Portuguese battle

No context is equal, and Portugal's specificity is unique. What could be done that would better prepare the education system for introducing English as a curricular subject in primary Education; what can be learned from Portugal's recent curriculum change?

Teacher profiles: Though the teachers of English in 1st cycle are invariably experienced, their consolidated teaching practices, when associated with older learners, are detrimental to fostering appropriate language learning processes in children. These teachers still need training, but more experiential, collaborative and school-based training. They also need recognition of their professional status – they are as important as all other teachers in the education system.

Teacher education and sensibilisation: Specialised teacher education for 1st cycle foreign language education requires the integration of modules which focus on approaches to language assessment. Assessment for learning continues to be problematic in practice and is uncommon in the majority of the Portuguese education system, thus, when introducing English as a curricular subject it would be beneficial to sensitize school directors and classroom generalist teachers to the differences between the teaching and assessment of English in a low-exposure context and the teaching and assessment of Portuguese, as well as showing these professionals that there are multiple approaches to classroom based assessment.

Material and resources: The absence of a syllabus and national guidelines for assessing English in 1st cycle exacerbates the problem in Portugal. Instead of expecting every school cluster to develop their own approaches to assessment, a set of national guidelines would ensure that the teachers of English were supported in successfully assessing all skills and domains, encouraged to use assessment procedures which did not rely upon paper and pencil tests and, as important, afforded professional credibility.

Acknowledgments

The study was funded by national funds through the Foundation for Science and Technology, I.P., within the projects: UID/ELT/04097/2013; UID/ELT/04097/2016; CEECIND/01822/2017; UID/ELT/04097/2019; UIDB/04097/2020; UIDP/04097/2020; and UIDB/01661/2020.

Bibliography

Abreu, Cláudia R./ Esteves, Vanessa R. (2015), *Let's Rock!* Porto: Porto Editora.

Bastos, Cristina/ Brites, Isabel (2012), *Actividades de Enriquecimento Curricular no 1º Ciclo do Ensino Básico INGLÊS: Relatório final de acompanhamento* APPI/CAP 2011/2012. [Online: https://appi.pt/storage/app/media/docs_appi/Relatorio-APPI_CAP-2011-2012.pdf. 28.07.2021].

Bento, Cristina/ Coelho, Raquel/ Joseph, Nicola/ Mourão, Sandie (2005), *Programa de Generalização do Ensino do Inglês no 1º ciclo do Ensino Básico – Orientações programáticas, materiais para o ensino e a aprendizagem*. Lisboa: DGIDC/ Ministério da Educação.

Bernardo, Vera M. B. C. (2014), *O Impacto do Programa de Generalização do Inglês na Aprendizagem da Língua Inglesa no 2ª Ciclo*. Unpublished Doctor's Thesis, Universidade de Beira Interior, Portugal.

Black, Paul/ Wiliam, Dylan (1998), Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 92(1), 81–90. [Online: http://electronicportfolios.org/afl/InsideBlackBox.pdf. 17.08.2021].

Bravo, Conceição/ Cravo, Ana & Duarte, Eulália (2015), *Metas Curriculares de Inglês 1º, 2º e 3º ciclos*. Lisboa: Ministério de Educação. [Online: https://www.dge.mec.pt/sites/default/files/Basico/Metas/ING/eb_metas_curriculares_ingles.pdf. 28.07.2021].

Cameron, Lynne (2001), *Teaching English to Young Learners*. Cambridge: Cambridge University Press.

CILT (2001), European Language Portfolio – Junior version. London: CILT, the National Centre for Languages. [Online: https://www.ecml.at/Portals/1/ELP_Portfolios/d0855489-1020-4fac-9722-e2a654e7c596.pdf?ver=2011-10-09-184800-560. 17.08.2021].

Conselho Nacional de Educação (2015), *Recomendação. Retenção escolar nos ensinos básico e secundário*. Lisboa: Conselho Nacional de Educação.

Council of Europe (2016), *Collated Representative Samples of Descriptors of Language Competencies Developed for Young Learners Aged 7–10 years*. Strasbourg: Language Policy Unit, Council of Europe.

Council of Europe (2018), *Common European Framework for Reference: Companion Volume with New Descriptors*. Strasbourg: Language Policy Unit, Council of Europe.

DGEEC (2014), *Atividades de Enriquecimento Curricular: 2013/2014*. Lisboa: Ministério de Educação e Ciência. [Online: http://www.dgeec.mec.pt/np4/99/%7B$clientServletPath%7D/?newsId=161&fileName=Relat_rio_AEC_2013_2014_Final.pdf. 28.07.2021].

Diehr, Bärbel (2009), Young Learners' Use of English: Imitation or Production?. In: Stewart, Tim (ed.). *Insights on Teaching Speaking in TESOL*. Alexandria, VA: TESOL, 53–66.

Ellis, Gail/ Ibrahim, Nayr (2015), *Teaching Children How to Learn*. Peaslake: Delta Publishing.

Ghosn, Irma-Kaarina (2019), Integrating Developmentally Approriate Assessment of Learning in a Young Learner Classroom. In: Prošić-Santovac, Danijela & Rixon, Shelagh (eds.) (2019), *Integrating Assessment in Early Language Learning and Teaching Practice*. Bristol: Multilingual Matters, 52–86.

Gregório, Carmo/ Perdigão, Rute/ Casas-Novas, Teresa (2014), *Integração do ensino da língua inglesa no currículo do 1º ciclo do Ensino Básico [Relatório Técnico]*. Lisboa: Conselho Nacional de Educação.

Ioannou-Georgio, Sophie/ Pavlou, Pavlos (2003), *Assessing Young Learners*. Oxford: Oxford University Press.

Jang, Eunice (2014), *Focus on Assessment*. Oxford: Oxford University Press.

Littlejohn, Andrew (2012), Language teaching materials and the (very) big picture. *Electronic Journal of Foreign Language Teaching*, 9(1), 283–297.

McKay, Penny (2006), *Assessing Young Language Learners*. Cambridge: Cambridge University Press.

OECD (2014), Education Policy Outlook – Portugal. [Online: https://www.oecd.org/education/EDUCATION%20POLICY%20OUTLOOK_PORTUGAL_EN.pdf. 17.08.2021].

OECD (2019), *TALIS Country Notes – Portugal*. [Online: https://www.oecd.org/education/talis/talis-2018-country-notes.htm. 28.07.2021].

Pinter, Annamaria (2006), *Teaching Young Language Learners*. Oxford: Oxford University Press.

Popham, William J. (2008), *Transformative Assessment*. Alexandria, VA: ASCD.

Rea-Dickins, Pauline (2000), Assessment in early years language learning contexts. *Language Testing* 27(2), 115–122.

Rixon, Shelagh/Prošić-Santovac, Danijela (2019), Introduction: Assessment and early language learning. In: Prošić-Santovac, Danijela & Rixon, Shelagh (eds.) (2019), *Integrating Assessment in Early Language Learning and Teaching Practice*. Bristol: Multilingual Matters, 1–18.

UNESCO – IBE (2013), *Glossary of Curriculum Terminology*. Geneva: UNESCO-IBE. [Online: http://www.ibe.unesco.org/fileadmin/user_upload/Publications/IBE_GlossaryCurriculumTerminology2013_eng.pdf. 28.07.2021].

Legislation

Decreto-Lei nº 176/2014, de 12 de dezembro, publicado na I Série, nº 240 do Diário da República, pp. 6064–6068.

Decreto-Lei nº 6/2001, de 18 de janeiro, publicado na I Série, nº 15 do Diário da República, pp. 258–265.

Decreto-Lei nº 91/2013, de 10 de julho, publicado na I Série, nº 131 do Diário da República, pp. 4013–4015.

Despacho nº 12591/2006, de 16 de junho, publicado na II Série, nº 115 do Diário da República, pp. 8783–8787.

Despacho nº 14460/2008, de 26 de maio, publicado na II Série, nº 100 do Diário da República, pp. 23194–23198.

Despacho nº 14753/2005, de 5 de julho, publicado na II Série, nº 127 do Diário da República, pp. 9785–9787.

Part II: Fostering reading, spelling, and writing

Gee Macrory

‘Commas in the air’: Young children’s experiences of learning the orthographies of French and Spanish as a foreign language

Abstract This chapter reports upon a small-scale research project in an English primary school, exploring young children’s experiences of a new orthography, in this case French or Spanish. The aim of the project was to explore their attitudes towards this, but also to discern the extent to which they were aware of rules. Forty-four children aged from seven to eleven were interviewed in focus groups; interviews were also held with eight teachers, to investigate their views on literacy in the primary language classroom. Findings from the children suggested that they were aware of differences between the new orthography and English, but possessed only a hazy understanding of any rules. The teacher interviews suggested an emphasis on an oral approach, with enjoyment of lessons seen as a priority. They expressed concern about possible detriment to the children’s English and lack of curriculum time. One possible inference is that the lack of time results in a predominantly oral and ‘fun’ approach. This may risk trivialising languages, resulting in a limited time allocation which then affects the choice of pedagogy, creating a vicious circle.

1. Introduction

Language learning and teaching in the UK has arguably never been as important as it is now. We live in an increasingly globalised, multicultural and multilingual world where cross-cultural communication is vital to peace and harmony. With this perspective, the UK government in 2000 commissioned a report into the UK’s capability for the 21st century in terms of language learning (Nuffield 2000), the outcome of which was a National Languages Strategy (DfES 2002) with a set of key recommendations. Subsequent government and policy changes have inhibited greatly the implementation of these, such that now, more than twenty years into that new century, we face a situation of increasing urgency. A notable and much vaunted aspect of the strategy was the proposal that primary

languages become an entitlement at age seven from January 2010; a subsequent review of the primary curriculum (Rose 2009) recommended that this in fact should from September 2010 be a compulsory element of the curriculum. By this stage, considerable progress had been made towards developing capacity for this provision (see Cable, Driscoll, Mitchell, Sing, Cremin, Earl, Eyres, Holmes, Martin & Heins 2010; Driscoll, Jones & Macrory 2004; Wade, Marshall & O'Donnell 2009). However, a change in government in May 2010 resulted in vastly reduced funding and an effective abandonment of the primary languages project until the decision was taken to make it a statutory part of the curriculum from September 2014. The delay produced by such governmental policy changes is all the more worrying, given the broader context of language learning in the UK, where motivation and attainment in foreign languages in the UK are the subject of increasing concern (Coleman 2009; Hagger-Vaughan 2016; Ofsted 2015; Tinsley 2013; Tinsley & Board 2015; Tinsley & Board 2016; Tinsley & Board 2018; Tinsley & Doležal 2018). In a situation where a relatively recent survey notes that the percentage of pupils taking a GCSE (General Certificate of Secondary Education, the exam taken at age 16) has dropped from 76 % in 2002 to 47 % in 2017 (Tinsley & Doležal 2018: 3) it can be of little surprise that hopes are pinned upon an earlier start in order to redress the situation. An earlier start and thus longer exposure might create the success that would motivate learners to continue past the age of 14, when languages cease to be compulsory.

Starting early with foreign language learning, however, must be a positive experience for all involved, notably the young learners themselves, if this is to result in a longer term successful outcome. Central to any discussion surrounding the effectiveness of a language policy, is, of course, the issue of pedagogic approaches.

2. Early language learning and literacy

One fundamental question in language teaching pedagogy is without doubt that of the balance of spoken and written language, perhaps all the more so in the case of young learners, who are still mastering the written version of their first language (L1). In terms of language *policy*, it is worth noting what is actually expected. The most recent version of the National

Curriculum (Department for Education 2013a), inter alia, expects children at Key Stage 2 (age 7–11) to do the following in foreign language lessons:

- explore the patterns and sounds of language through songs and rhymes and link the spelling, sound and meaning of words
- develop accurate pronunciation and intonation so that others understand when they are reading aloud or using familiar words and phrases
- read carefully and show understanding of words, phrases and simple writing
- broaden their vocabulary and develop their ability to understand new words that are introduced into familiar written material, including through using a dictionary

The aims just listed suggest a role for the written language in the classroom. However, research into second language literacy, points to potential difficulties for learners. One key issue that has attracted much attention is that of transfer from L1 (Koda 2005). Clearly there is a difference between moving from a non-alphabetic writing system, such as morphemic or syllabic, to one that is alphabetic and premised upon phoneme-grapheme correspondences. However, this is not to say that moving between alphabetic systems is necessarily easier. While learners may have to move from one alphabetic system to another (such as English to Arabic), even within the same orthography, such as the Roman alphabet, the task is far from simple. To a large degree the graphemes are the same, although some languages using the Roman alphabet may, for example, use diacritics particular to that language, such as the acute accent <é> in French, the tilde in Spanish <ñ> or the umlaut in German <ü>. But what does differ is the phoneme-grapheme correspondences (PGCs) of each language and, importantly, orthographic depth. The concept of shallow and deep orthographies is relevant here, in that some orthographies such as Spanish are seen as shallow on account of more regular GPCs, whereas English and French are considered to be deep as their GPCs are much more complex (Sampson 1985). Koda (2008) suggests that the distance between the first and second language affects the learners and indeed, various studies appear to bear this out. Muñoz's (2014) study of Grade 3 Spanish-Catalan beginner learners of English found that 50 % of their perceived difficulties with learning English were focussed on spelling. A further finding, that

three years later, spelling still accounted for 32 % of difficulties, underlines the potential barrier to effective language learning that the orthography can pose. While, unsurprisingly, much research has focussed on the learning of English, given its dominance globally (Enever 2018), research into the learning of other languages suggests that even moving from a deep to a shallow orthography can pose problems. Rafat (2016) found that for English-speaking beginners of Spanish, different GPCs such as <v>-/b/ presented difficulties for the learners, resulting in the transfer of the pronunciation associated with the grapheme in English.

In the UK, studies have largely been carried out with secondary aged pupils, perhaps understandably given that only in 2014 was Languages made a compulsory subject in England, such that only in the summer of 2018 did children complete four years of language learning. Furthermore, the implementation of this curriculum reform has suffered from a somewhat variable approach However, studies carried out with older learners have indicated some difficulties that learners experience with the written form of French. French and English are similarly deep in terms of orthography, but it may be that some similarity at this level is not necessarily facilitative. Erler (2004) found that by the end of their first year of learning French at secondary school, learners had very limited awareness of the spelling-sound rules for principal vowel sounds and for the general rule of silent final consonants. A number of research projects have been undertaken by Woore, who, for example, in 2009 followed 85 learners through to their second year of learning French and found little evidence of progress in their decoding skills (Woore 2009). A later in-depth study of 12 learners found little evidence of knowledge of the GPCs of French (Woore 2010). More recently, he investigated 31 15–16 year old learners of French in a school in England, in order to explore their ability to decode familiar versus unfamiliar words. Participants decoded a large majority of both familiar and unfamiliar French words inaccurately but decoded the spelling bodies of familiar words significantly more accurately than identical spelling bodies in unfamiliar words, leading Woore (2018: 467) to speculate that the learners may be relying upon retrieving stored pronunciations for familiar words as a compensatory mechanism. He further notes that "[i]n the longer term, however, this is an inefficient basis for phonological decoding in French [….] [and a] pedagogical implication of the above

finding is that teachers may gain a misleading impression of learners' L2 decoding proficiency if they over-rely on simple texts comprising mainly familiar words" (2018: 467). A further implication of the problems apparently encountered by learners is outlined by Erler and Macaro (2011), who undertook a study with a large sample of 11–14 year old learners of French in England. Their finding (2011: 512) that "disaffection with continuing French has been shown here to be linked with poor decoding skills, low self-efficacy and the students' perception of having no control over the language or the learning" suggests a potentially crucial link between reading ability and motivation to continue studying the language, with clear implications for the broader context of language learning in the UK outlined earlier in this chapter.

The pedagogical implications of research are not clear-cut, however, as research by Graham et al. (2014) indicates. A recent project undertaken by Graham, Courtney, Marinis, and Tonkyn (2014) explored the differential impact of oracy and literacy approaches for the teaching of French to learners in Years 5, 6 and 7 of schooling. They found no clear advantage for either approach, although the authors observed that the literacy activities in which learners were engaged remained largely at word and sentence level. This led them to speculate that literacy activities consisting of word-level work were insufficient to bring a clear advantage for a literacy-based approach; they also noted that there was little evidence of instruction in grapheme-phoneme correspondences (2014: 11–12). The doctoral study of Porter (2014: 338) does, however, recommend introducing Modern Foreign Languages (MFL) literacy as an integral part of an early start to language learning, although she cautions that this should be done sensitively as it may present difficulties to learners. What, however, characterises current pedagogic approaches? In their large-scale three year study, Cable et al. (2010) observed that in primary languages generally, oracy had been given priority over literacy and they suggested that the limited amount of time spent on literacy activities was possibly a result of the shortness of lessons and the relatively limited confidence and expertise noted among some staff. In a study of beginning teachers, all of whom had chosen primary languages as a specialism, Macrory (2019) found that their experience on school placement echoed previous research such as that of Graham et al., noted above. A total of 55 beginning teachers were

involved in the research project, including 23 postgraduates who had two school placements, and 32 undergraduates who had had three or four school placements. The beginning teachers reported limited experience of observing or having the opportunity to teach literacy activities and very little indeed when it came to the more specific issue of phoneme-grapheme correspondences in the new language, which was usually Spanish or French, and occasionally German. In a study of 18 pupils in Year 6 (final year of Primary school, aged 10–11), Chambers expressed surprise at the lack of writing in particular, noting (2019: 28) that "it is perhaps strange that writing played such a minor role in their Primary school experience, not least given the underpinning support and consolidation it can provide to listening, speaking and reading". The same pupils interviewed a year later, interestingly, showed a preference for their secondary school experience, leading him to suggest that they were by now ready to feel that they were learning more and progressing, and also to speculate that their Primary school teachers' focus was likely to have been on making MFL a positive, enjoyable experience. However, a study by Wingate (2018) suggests that the focus on light-hearted activities is not exclusive to the primary phase. She observed and video-recorded 15 lessons in German, Spanish and French in Key Stage 3 (11–14 age range) and reported a preponderance of activities that required no more than reactive behaviour, repetition and reproduction, such as chorusing words or phrases, reading aloud, or writing down rules (2018: 448). She concluded that "it is fair to speak of an MFL classroom culture of low expectations, lack of challenge and light entertainment. It is possible that this culture has gradually developed as a result of the negative perceptions of MFL in England and of dwindling student numbers and that it reflects teachers' desire to make language learning look easy and fun" (2018: 452). This reflects an Ofsted (Office for Standards in Education, the inspection regime in England) report into Key Stage 3 that described this phase of schooling as the 'wasted years' (Ofsted 2015); this formed part of the background into the 2016 report by the Teaching Schools Council into MFL pedagogy at Key stages 3 and 4 (i.e. 11–16 years of age). Although focussed mainly upon the upper years of schooling, the report also states that "much of what we say is also relevant to pedagogy in the upper end of primary" (Teaching Schools Council 2016: 8), going on to define curricular content as the knowledge to be

covered and mastered by pupils, saying that, "in languages this includes grammar, vocabulary and phonics (relationship between writing and pronunciation, or phoneme-grapheme correspondences)" (2016: 8).

This brings us to the issue of what constitutes knowledge on the part of young children. How much explicit awareness should young children have about the new orthography? We might ask to what extent young children themselves have a view on this. Millonig, Stickler and Coleman (2019: 394) note that in the extensive list of publications of the European Centre for Modern Languages covering 1999–2015, there is not a single study investigating Primary school pupils' perceptions of the language learning process. Similarly, Kolb (2007) noted how little is known about how *children* (my emphasis) perceive their learning activities and how they conceptualise the process of learning a foreign language. She conducted research in two German Primary school classes (43 learners, aged 8–9, who were in their third year of learning English as a second language), collecting data from classroom recordings, learner interviews and children's texts. The results of the study show that even young learners can engage in reflective activities and that they are remarkably aware of the learning process (2007: 227). More recently, Muñoz (2014) observes that few studies have looked into children's viewpoints of learning a foreign language. She notes that studies conducted in the young learners' classroom have focused on aspects such as age-appropriate teaching methods, attitudinal and learning outcomes, and the effects of second language learning on the first, but that research into children's development of consciousness about language and language learning still lags behind. She asks, "[f]or example, to what extent are YLs aware of the learning processes in which they engage, of their own skills as learners, or of the conditions that are favourable for their FL learning?" (2014: 24). As noted above, the children in this study were very able to highlight the lack of transparency of English orthography. However, they also showed an early awareness of the conditions that help them learn English, of classroom management issues and of learning-effective activities. Asked in Year 6 about what helped them learn English best, data from 28 child participants in the longitudinal study indicated that they most enjoyed games, songs, listening and speaking activities, in that order. In contrast, they learned most from vocabulary, form-focused, speaking and listening activities. Muñoz (2014: 36) remarked that it was "interesting to

note that the choice of activities with which they learnt the most does not necessarily match the choice of activities learners enjoyed most […] [and that] this mismatch […] reflects an early awareness of learning processes and the conditions that favour them", allowing her to conclude that by the end of primary education pupils identify form-focused activities as good for their learning.

The place of form and explicit knowledge about language is currently not without controversy in the primary curriculum in England. In May 2013, a test of spelling, punctuation and grammar (SPaG) was introduced for children in the final year of Primary school, where children aged 10-11 are asked to, for example, identify the adverb in a sentence or correctly place a certain number of commas in an unpunctuated sentence (Department for Education 2018). Clark (2010), anticipating change in the curriculum for English in both primary and secondary schools, proposed a recontextualised curriculum for grammar, argued that for the teaching of grammar to have any salience or purpose at all, it had to be integrated into the curriculum as a whole. However, Safford, Messer, McLachlan, Walker, Ghosh, Laird, Disney & Wright (2016) explored teachers' responses to teaching grammar to this statutory test format, and how teachers implemented rapid curriculum change in their classrooms. The research sought to learn the perspectives of teachers as they adjusted to these new English assessments and new expectations for children's language in the Primary school. Drawing upon 16 teacher interviews and an online survey of 170 teaching staff, she concluded that since the introduction of the statutory SPaG test in Primary schools, time spent teaching decontextualised and contextualised grammar had increased significantly and that grammar was now taught explicitly and formally as a classroom literacy routine; she also noted that teachers disagreed about the extent to which explicit grammar teaching and testing have a positive impact on pupils' language and literacy skills. On the other hand, Bell (2016) undertook a case study of developing teacher attitudes, beliefs and content knowledge at a Primary school in the North-West of England, focusing on grammatical terms and concepts. His findings suggest that, while much work remains to be done in developing teachers' knowledge base, attitudes are largely supportive of teaching children grammar terms and concepts.

Even more pertinent to the issue of orthography is the current focus on grapheme- phoneme correspondences as the key to early reading. It has now been 15 years since the Rose Report of 2006 (DfES 2006), commissioned by the government, recommended that the approach to the teaching of L1 reading in Primary schools in England should focus on systematic synthetic phonics. While this has not been without controversy (see Ellis & Moss 2014), Ofsted (2012, 2016) noted that most schools were making explicit use of synthetic phonics programmes. We might therefore presume that most, if not all Primary schools in England have an awareness of the role of grapheme-phoneme correspondences in learning to read, at least in English as a first language. A further issue is that a phonics screening test was introduced in 2012 in order to determine whether six-year-old children were meeting an appropriate standard in phonic decoding. This is a check that includes both real words and pseudowords, twenty of each. The test items are phonemically regular and in order to meet the standard required by the government, children have to read at least 32 of the items correctly without support or prompting, to identify children struggling with phonic skills. Like the SPaG test, this has proved somewhat controversial, and the government's own evaluation carried out by Walker, Sainsbury, Worth, Bamforth and Betts (2015) for the National Foundation for Educational Research (NFER) found that "the evidence suggests that the introduction of the check has had an impact on pupils' attainment in phonics, but not (or not yet) on their attainment in literacy" (2015: 11), although research by Duff, Melonig, Bailey and Snowling (2015) found that the check was strongly correlated with other literacy skills and was sensitive in identifying at-risk readers. However, irrespective of the value of such a test, it seems clear that the direction of travel in Primary schools in England has been very much towards increased expectations in terms of children's metalinguistic awareness and ability to consider not just function but form within language.

Such an alteration in pedagogic approach, however, does not exist in a vacuum. Official policy as expressed through, for example, the requirements of a national curriculum are mediated through more local school policies and teacher beliefs (Lüth 2019). Not only are teachers' beliefs in pedagogic approaches embedded in personal history and experience but they may in

turn influence children's beliefs. Millonig et al. (2019) undertook a study in Austria, in two separate Primary school classrooms of year two pupils aged seven to eight. They investigated through children's drawings their perceptions of foreign language learning, in this case English. They found that the teaching approaches used in Austrian Primary school classrooms did not always follow the ministerial guidelines for foreign language learning, and reflected individual teachers' classroom practice. The overall findings from the two case study classrooms also suggested that pupils' perceptions of English were influenced by the teaching approaches and strategies used in the classroom by their teachers (2019: 242). That there is complexity and nuance in the implementation of policy seems without doubt. In addition to the issue of teacher beliefs and their possible relationship with children's views, Gayton (2018: 394) points to a possible disconnect between languages teams and senior management in valuing the subject area. This suggests that in England at least there may be inconsistency between the value placed upon foreign language learning and that afforded to L1 phonics, grammar, spelling and punctuation. One implication of this is that, alongside an investigation of the perceptions of young children, the views of their teachers need also to be sought.

3. Aims of the research

The study that is reported on here was conducted in a Primary school in the North West of England in March 2018. It was carried out in order to investigate young children's experiences of a new orthography in the primary classroom. In particular, we were keen to explore their attitudes towards the new orthography. Previous research by the author (Macrory 2019) suggested that relatively little attention was paid to literacy in the primary language classroom, despite the focus on this in first language literacy. We were also interested in the extent to which the children were aware of rules and what they found helpful in their language lessons. We were mindful of the need to frame the project in a positive way, focussing on what might be helpful in order to avoid any potential criticism of their language lessons, but, importantly to try to tease out any possible implications for future practice. We decided to target the age range 7-11 (Years 3, 4, 5 and 6, together known as Key Stage 2 in the UK), as this is

the age range in which primary language teaching has been compulsory since September 2014. Since the research was carried out in March 2018, this meant that the children in Year 6 had almost completed the four years of language teaching now required.

We also decided to investigate their teachers' perspectives on the importance of literacy in the primary language classroom, and their views of the role that it does and/or could play in effective language learning. We chose to focus on two groups of teachers: firstly, the two Languages teachers themselves, in order better to understand what motivated their pedagogic approach, and secondly, the generalist class teachers, in order to explore their understandings of what happened in language lessons and the extent to which, and how, they supported this in other parts of the curriculum.

4. Research methods

A total of 44 children were interviewed in focus group meetings facilitated by both the author of this article and a teacher in the school, who worked as a class teacher but who, as a German specialist in the secondary phase originally, had also worked closely with us at the university as an Associate Lecturer. Focus groups were chosen as a method as some learners often find speaking more comfortable than writing, and may produce orally more extended answers (Mackey & Gass 2015: 255), and we deemed this to be highly relevant to young children. Additionally, the presence of this teacher as a familiar adult (in addition to being a full-time Year 5 class teacher, she also ran a lunch time German club) was deemed important to allay any concerns the children might have had about talking to an unfamiliar visitor. The focus groups were composed as follows: two groups of six in each of Years 3, 4 and 5 and one group of eight in Year 6. The children had different experiences according to the year group they were in, as a result of the school's language policy. In this particular school, Years 3 and 4 learn Spanish and Years 5 and 6 learn French. In addition, all the children learn a small amount of French between the ages of five and seven, i.e. prior to Key Stage 2. This means that the children in Years 3 and 4 had been learning Spanish, but the children in Years 5 and 6 were learning French but had previously received two years of Spanish teaching.

Semi-structured interviews were held with two language teachers, five class teachers, and one learning support teacher. The latter offered to be interviewed out of interest in the project and was duly included. The two specialist language teachers were both trained as language teachers, albeit for secondary originally (at the time of their training, it was not possible to specialise for primary language teaching), and had degree-level subject knowledge. They were employed full-time by the school. One had the title of 'Head of Department' and taught Languages full-time, whereas the other taught Languages some days of the week but also taught music. They were available to be interviewed individually due to the nature of their timetables, as was the support teacher, whereas the five classroom teachers were interviewed in two separate groups of three and two. The interview guide was structured to reflect the research questions and provide an opportunity to explore viewpoints and clarify understandings (Newby 2010: 340). We deemed it important to explore the beliefs and attitudes of the language teachers, as these are likely to have an impact on the pedagogy adopted. The class teachers, however, were included in order to get a sense of the ecology of the school environment as a whole.

The interviews and focus groups were conducted in English and recorded using VoiceRecord (freely available to download) on an i-Pad, and the analysis was carried out using an inductive approach (Newby 2010: 464).

5. Findings

5.1 *Children's perceptions*

As noted above, we were keen to explore the children's attitudes towards the new orthography, and we were also interested in the extent to which they were aware of rules and what they found helpful in their language lessons. However, we began by asking some more general questions about their perceptions of their language lessons. What was evident was their enjoyment of these, with games and songs particularly welcomed by all the groups, with repetition and worksheets also featuring for some. A typical comment from a Year 4 child learning Spanish was, "really like it, we do fun things and play games". This was not exclusive to the younger

children by any means, with the oldest (Year 6) children also citing games, songs and repetition as helpful.

More pertinent to the aims of the project, the children were then asked several questions about their views of French/Spanish spellings in order to elicit some opinions or understanding of the orthography. For example, they were asked what they thought of the spelling, how they tried to read words written down, any examples of spelling that they found tricky or easy and also, what their teacher does that is helpful.

It was clear that they were very aware of the differences between the foreign language(s) and English. They described these as variously 'weird', 'hard' and 'tricky', and also offered some comparison between French and Spanish. Here it may be useful to recall that the children in Years 5 and 6 were learning French and had previously studied Spanish for two years in Years 3 and 4. However, as all the children had had some exposure to French before this, while in Year 1 and 2 (albeit not particularly systematically), even the younger children were able to make some observations. The children were all asked what they thought about French and Spanish spelling and a number of features appeared to have been salient to them. First of all, a number of children referred to what they saw as 'silent letters', in particular citing the fact that the grapheme <h> in French and Spanish was not pronounced. In Spanish, they tended to remark upon the pronunciation of <j> in Spanish; in both languages they commented upon the pronunciation of <ll>. The inverted <¿> and <¡> in Spanish was noted by several children, as these are different punctuation markers to those used in English. Only one child commented upon the pronunciation of the grapheme <z> in Spanish. Similarly, there were only two children who referred respectively to the pronunciation of <j> and <ch> in French. What the children had observed, however, were the accented graphemes in both languages. A number of them had some interesting, if not amusing, comments to make. For example, some children in Year 3 commented variously that,

You have to remember the little thing like the comma in the air.
Some of the letters have a line on top, I think it's to make the sound a bit different maybe.
Spanish words are a bit easier to learn than French… they're a bit shorter and only some of them have the lines on top.
Usually 'A's in French have little dots on the top.

The children in Year 4 equally made some observations:

They have other things on top of the letters that we don't have.
We have to add the little accents that you don't have in English.

Year 5 too had some observations such as the following:

In our language we hardly use a comma – in French they do that quite a lot.
I find it confusing because they sometimes have weird little signs on top.
I find the spelling more confusing than the actual speaking because there are lots of commas on top of the letters.

Children in Year 6 also observed that:

I think it's quite hard… as their apostrophes lean to the left and then to the right and then there's an up and down arrow with a different pronunciation.
It's quite hard that you have all the different accents. … it would be easy to write one or two lines like you have to but there's upside down squiggles.

5.2 Rules and patterns

The children were asked about any spelling rules they thought they knew. As noted above, they did have some awareness of differences from English and in Year 4, for example, some children were able to articulate that in 'José', the <j> is like an /h/ sound and one child observed that the <z> "is like a <th>". However, other data from Years 5 and 6 suggest that they were generally unsure about rules, as the following extract from Year 5 suggests:

Researcher: How might <j> sound in French?
Child: Je.
Researcher: How did you know that?
Child: I know 'je m'appelle….' I figured it out.

Also from Year 5:

R: What are the bits you find trickiest?
C: Definitely the accents because you don't know when to put them in.
R: Has that been explained to you?
C: Not yet.

Year 6

C: The phonics are different – they look like how you would say it but they sound different – like cat is C-H-A-T but you'd pronounce the <ch> *as /ʃ/.*
R: Is that a rule?
C: I don't know.

The reason why the rules of French spelling were still somewhat mysterious for the learners might be that the children had previously studied Spanish. What was striking in all the focus groups, however, was the positive view that the children held of rules and the fact that they were able to point to teaching strategies they saw as helpful.

In Years 3 and 4, children offered comments such as:

I think Spanish spelling rules are as important as the English spelling rules.
If you got a new word, you'd be able to pronounce it if you knew all the letters in different words.

The children in Years 5 and 6 were even more articulate about rules, saying that:

If there's a rule to it, you can think of that rule …it's easier to learn rules than words.
I think it's good (knowing a rule) because the next time you learn, you know what to do instead of having to ask again.
Sometimes it's helpful for people to explain why it's like that or if it's a rule or if you do it all the time.
If you come across the same thing again and you know it's a rule and you know to do it all the time whereas if you obviously don't know you might get that same thing wrong.

It was also clear that the children were aware of what their teachers were doing that was helpful.

Year 4

A <j> *is two kinds of /l/s ….Mrs X shows us and it helps us to remember if she keeps pointing it out in lessons.*
Sometimes we have to notice because our teacher tells us that some words have different sounds.
The <j> *is like a/* h/ *sound…the teacher told us that*

Year 5

> *In French* <ch> *is like /ʃ/ – Mrs Y told us…it's quite helpful 'cos you have an idea of what it might be.*

A final observation in this section is that two children in Year 3 hinted at the possibility that their own language background might play a part in their perceptions: one child offered, "I know Polish and my dad says then writing Spanish is easier" and another observed that, "It's not like Arabic … you write in English but the words are actually French but you're writing them in English letters … you don't have a language like Arabic or Chinese".

To summarise this section, the children were very aware of some differences between English and French/Spanish and tried very hard to express their views. However, what is striking is the disparity between their metalinguistic awareness and appropriate metalinguistic terminology. Two observations can be made here: firstly, the lack of progression in their ability to discuss salient features of the orthography such as the accents is very noticeable; secondly, there is a remarkable contrast between this and the expectations embedded in the National Curriculum for English, which requires in Year 3, for example, that children learn the following terminology: preposition, conjunction, word family, prefix clause, subordinate clause, direct speech, consonant, consonant, letter, vowel, vowel letter, inverted commas (or 'speech marks'); in Year 4, they learn: determiner, pronoun, possessive pronoun, adverbial; Year 5 requires: modal verb, relative pronoun, relative clause, parenthesis, bracket, dash, cohesion, ambiguity; and Year 6 requires: subject, object, active, passive, synonym, antonym, ellipsis, hyphen, colon, semi-colon, bullet points. Prior to Key Stage 2 (ages 7-11), in Year 1, five to six year olds must master: letter, capital letter, word, singular, plural, sentence, punctuation, full stop, question mark, exclamation mark, followed in Year 2 by: noun, noun phrase, statement, question, exclamation, command, compound, suffix, adjective, adverb, verb, tense (past, present), apostrophe, comma. While on the one hand, we might find the children's comments amusing (the 'comma in the air') and also suggest that they have indeed absorbed this terminology as required, the disparity is clear. Although the children are arguably inventive and drawing upon previous knowledge and terminology (Department for Education 2013b), they are clearly not aware of the appropriate terminology to use in any discussion of their foreign language orthography.

The children were able to express some views about the place of rules in their learning. Not only were they lacking the appropriate terminology, but they were also somewhat hazy about the rules themselves. Yet at the same time, they were able to articulate the potential value and role of knowing the rules. Equally, they showed awareness of what their teachers were drawing attention to and also in fact welcomed the games and songs that more generally assisted their learning.

6. Teachers' findings

6.1 *Class teachers*

Interviews were conducted with one group of class teachers, composed of one each from Years 3, 5 and 6 and then a second group of two class teachers, one from Year 3 and one from Year 4. Some clear themes emerged from the interviews. Firstly, they had inferred from children's comments that they enjoyed their language lessons and were of the view that this enjoyment was an important element in their learning. Although they all admitted that they were unsure as to what actually happened in foreign language lessons, their understanding was that the approach was predominantly oral. Teacher M noted that it was "largely speaking" and teacher K that the "focus (was) on using the language, not writing it down". Teacher N remarked that "writing things down in a grammatically correct sentence isn't really the approach". Indeed, some commented on the approach being at word or phrase level, rather than at sentence level, with comments such as "I think children learn phrases but aren't sure which are separate words" (Teacher A). Teacher M thought it possible that the children "can say a word but not recognise it written down". The downplaying of literacy was linked to the issue of enjoyment – Teacher K remarked that "you've got to keep it fun, keep them engaged, more a fun approach as writing might turn them off…. they do lots of writing elsewhere, they don't need that pressure". An underlying concern was the grasp of writing in *English* that the children had, with the view that they "had to be a certain level in phonics" (Teacher P), and Teacher K saying that "so many struggle with English, so having to do it in a different language would be difficult for them". However, the teachers were not hostile to the idea of a more systematic approach to the new orthography, rather they expressed

concern about lack of curriculum time, with, for example, Teacher K taking the view that an introduction to the GPCs of the foreign language was "highly likely to be helpful, if the time were there".

A further individual interview was conducted with the teacher responsible for children with additional learning needs and/or learning difficulties, who volunteered out of interest in the project. As a qualified teacher of dyslexia, she was able to offer some thoughtful observations. In citing examples of some children learning some French that they could repeat but not explain to her what it meant, she saw the possible value of seeing the language written down. Equally, she could see a place for introducing the children to foreign language phonics, noting that "most children have their L1 embedded and it would be a familiar approach, quiet enlightening as they'd see a foreign word and could actually say it". She went on to say that she thought "by having to write a word in a foreign language, think about it, might strengthen their overall phonological awareness skills which might help their spelling in *any* language".

6.2 *Language teachers*

Individual interviews were held with each of the two qualified language teachers in the school. Both teachers' remarks tended to confirm the views expressed about the oral approach to pedagogy by the class teachers. Teacher R was quick to assert that "I don't make a point of teaching the spelling at all as the emphasis is on the oral". However, she then went on to say that "I do really think that the visual backup is very important for visual learners…but not putting emphasis on the spelling, just as a backup". This emphasis on the visual aspect was again alluded to in the comment that "they're not actually physically spelling it, they're seeing it… from a reading point of view, they're going to need it". When asked about drawing the children's attention to GPCs, she said "just where it comes up… I'll look and say 'What pronunciation will we need to remember?' So we'll definitely talk about that as much as possible so that when they are reading the words, they don't pronounce it wrong". Equally, she offered: "I do draw their attention to things like double <l> in French and Spanish and say to them 'What do you remember about that double <l> sound… what pronunciation are we using here?' " Thus, while

the role she saw for the writing system was perhaps more about supporting pronunciation, she also reiterated that the focus was on the spoken word and listening, expressing concern that "some are already quite challenged with the English spelling and we only have a short time, 25–30 minutes a week", reflecting to some degree the reservations expressed by the class teachers.

These reservations were echoed by the second Languages teacher (Teacher C) interviewed who held the role of Head of Languages. She reiterated the concern that at approximately 30 minutes per week, there was not enough time in the curriculum for languages. Like her colleague, she favoured an oral approach, and for example, said that she would "just introduce a picture, I wouldn't show them the word, they just speak and hear the word and then later maybe matching the word and the picture… I don't do a lot of written work, just labelling pictures and answering some questions". Asked if she thought the children had any sense of the spelling patterns of French and Spanish, her response was "probably not, no", yet went on to say that "if I am writing on the board, I would draw attention to the /ʃ/ sound or the fact that <ç> makes the /s/ sound, so I would draw attention to that… we put that accent on it because… so I *am* doing it". This appeared to surprise her slightly, but she did go on, like her colleagues, to note as a challenge that "a lot of them struggle to spell in English, so it is better to be secure in L1 first". Her primary motivation, however, was summed by her final remark that "what I want to do is just make them love languages… the more technical side of it is for high school".

To summarise this section, then, the teachers together conveyed a strong message that in this school at least, children's enjoyment of their language lessons was paramount. Concern about lack of curriculum time and the children's grasp of English (phonics in particular) appears to underlie a hesitant approach to foreign language literacy. Nevertheless, the languages teachers themselves acknowledged that there were indeed occasions when they drew children's attentions to features of the new orthography, albeit in a rather unsystematic way, prompted more by pedagogic moments than by any curriculum planning.

7. Discussion and implications

What can we make of these findings? This was, of course, a small-scale study in only one Primary school, which does not permit generalisation. Nevertheless, we can draw some tentative conclusions that could inform further research. First of all, the children's positive perceptions of their activities such as songs and games reflect other research, such as that of Muñoz (2014) who found the same preference for these activities. Her research also suggested that the 28 children in her study were aware that they learned most from form-focussed activities rather than the ones they enjoyed the most, prompting her to argue that this mismatch reflected an early awareness of language learning processes and what facilitated these. While the children in our study did not express a parallel view, arguably their views on the potential usefulness of rules does reflect a similar awareness. However, while this to some degree hints at a discrepancy, in fact it might also suggest that the children's perceptions are, as Millonig et al. (2019) found, influenced by the strategies adopted by their teachers. The children in our study clearly welcomed the ludic approach adopted, and it is possible that the appreciation of rules was also influenced by some of the strategies used by their teachers, albeit somewhat unsystematically. Not only were the children aware of what they found helpful on the part of their teachers, but they also showed a level of metalinguistic awareness, although the absence of appropriate terminology in relation to the accented graphemes was particularly noticeable, leading them to draw upon previously learned terms such as 'comma' and 'apostrophe' as well as using terms of their own choosing such as 'sign', 'dot', 'line' and 'squiggle'. That such terms were still in use by Year 6 (and the data was collected at the end of March, only three months before the end of Primary schooling and therefore towards the very end of four years of Languages teaching), points both to a discrepancy between the expectations embedded in the curriculum for L1 English and to a marked lack of progression in acquiring appropriate terminology in the foreign language. In turn, it suggests that the occasions when teachers drew attention to features of the orthography were welcomed by the children and could benefit from a more systematic approach.

These findings sit in marked contrast to the development of the broader curriculum in England. As noted earlier, the curriculum and assessment requirements for L1 English reflect a clear move in the direction of more explicit knowledge about language. The impact of such changes was noted by Safford (2016) whose findings suggested grammar was now being taught explicitly and formally as a classroom literacy routine. The support for this from teachers themselves is unclear, as Safford's findings that teachers were in disagreement as to the usefulness of this is in contrast to Bell's (2016) finding that teachers were largely in support of grammatical terms and concepts. Certainly, the views of the two Languages teachers interviewed for the current study suggest a certain ambivalence, as their initial assertions that they adopted an oral approach were followed by examples of how they had in fact drawn children's attention to some features of the orthography. Of the other teachers interviewed, only the special needs teacher (who had a particularly good understanding of phonological development) saw a possibly beneficial role for the introduction of phonics in a foreign language. The five class teachers were at least equivocal about the place of literacy in the foreign language classroom, concerned as they were about the importance of the enjoyment that they saw an oral approach bring to the children as well as concern about any possible detrimental impact upon L1 English. Given the current testing regime in England, this is hardly surprising, of course. Related to this may also be the desire to protect the children in some way. The remark quoted above, by Teacher M, that "you've got to keep it fun, keep them engaged, more a fun approach as writing might turn them off… they do lots of writing elsewhere, they don't need that pressure" is a very telling one, suggesting that language lessons offer some kind of respite from the rest of the curriculum. What, however, does this tell us about the *place* of foreign languages in the curriculum?

The lack of time was clearly a source of frustration to the Languages teachers and underlay much of the resistance of the class teachers to the idea of incorporating literacy activities. That the time issue was a more widespread concern was suggested by one teacher, who remarked that there was no time to follow up in class time what might have been taught in the French or Spanish lesson. The lack of time is pointed to by Dobson

(2018: 82) who, in a recent historical perspective on foreign languages in England laments the fact that the total lesson time reported by the European Commission in 2012 for English secondary students for the first foreign language was lower than the European average by an hour per week, noting too that provision in Primary schools was not yet sufficient to allow pupils to build successfully upon this in the next phase of education. Indeed, Ofsted (2016: 46–47) found in 106 Primary schools inspected that "the majority (over two-thirds) spent less than an hour on it… (and) [a]round four in ten practitioners identified pressures on time as one of the biggest barriers to effective teaching of the subject in Primary schools". Tinsley & Doležal also refer to the lack of consistency, saying that "[l]anguages remain a marginal subject which many Primary schools find challenging to deliver alongside many other competing demands" (2018: 3).

Dobson also points out that conditions in other European countries cannot be replicated in the UK, as "the universal presence of English in the media provides constant reinforcement for learning beyond the classroom". On the basis of that, one might well argue that more time should be allocated to language learning in the UK than in other countries, not less. Dobson's remarks regarding the universal presence of English are given further weight by Gayton (2018: 385), who reminds us that "French as a foreign language in the United Kingdom is not as pervasive or accessible as English as a foreign language in France. In such an environment, largely removed from either direct or indirect contact with the language and its related culture(s), the role of the teacher in introducing them into the classroom is crucial".

We might, however, wish to argue that these conditions pertaining in the UK mean that this task should not be wholly the responsibility of individual languages teachers. In the current study, there was general admission on the part of the class teachers that they did not really know what happened in Languages lessons. This indicates a possible disconnect between these and the broader curriculum. A further issue is raised by the comments made by the two children in Year 3 regarding Arabic and Polish as influencing perceptions of Spanish or French orthography. While this was admittedly only two children, it nevertheless serves as a reminder that

a new language is not grafted on to a blank slate. Rather, a much more holistic and cross-curricular approach may be what is needed. Indeed, the teachers' concerns about the potential detriment to English L1 skills could well be allayed by findings such as those of Murphy, Macaro, Alba and Cipolla (2015), who found a positive impact of learning a second language in Primary school on developing L1 literacy skills. Yet, at present, there is a risk that the real lack of time forces foreign language learning into a narrow silo, characterised by 'fun' and 'enjoyment', reminiscent of the 'light entertainment' alluded to be Wingate (2018).

Such relegation has serious implications. Gayton's research found a disconnect between Languages teams and senior management teams in valuing the subject area (2018: 394). While this research was carried out in a secondary context, it is not beyond possibility that a subject seen as 'fun' and a possible respite from the rest of the curriculum can be perceived as trivial and unworthy of substantial time. Thus, we may have here a vicious circle, whereby the lack of time accorded results in teachers' prioritising enjoyment, which in turn devalues the subject in the eyes of those in charge of the curriculum. Particularly within a context of high accountability for subjects deemed 'core', notably English and Mathematics, and a national scenario of decreased interest in language learning, we risk exacerbating the problem by failing to devote the ample time in the curriculum that is needed for substantive and engaging language lessons. Re-positioning language learning from the outset as a serious endeavour might help us to address the language learning crisis alluded to by so many (see, for example, Lanvers & Coleman 2017). As noted at the start of this chapter, we live in an increasingly globalised, multicultural and multilingual world where cross-cultural communication is vital to peace and harmony. We owe it to our children and young people that they are able to participate fully in this world.

Bibliography

Bell, Huw (2016), Teacher knowledge and beliefs about grammar: A case study of an English Primary school. *English in Education,* 50: 2, 148–163.

Cable, Carrie/ Driscoll, Patricia/ Mitchell, Rosamond/ Sing, Sue/ Cremin, Teresa/ Earl, Justine/ Eyres, Ian/ Holmes, Bernardette/ Martin, Cynthia with Heins, Barbara (2010), *Language Learning at Key Stage 2: A Longitudinal Study*. London: Department for Children, Schools and Families (DCSF Research Reports DCSF-RR198).

Chambers, Gary (2019), Pupils' perceptions of Key Stage 2 to Key Stage 3 transition in modern foreign languages. *The Language Learning Journal*, 47: 1, 19–33. DOI:10.1080/09571736.2016.1172331.

Clark, Urszula I.L. (2010), Grammar in the curriculum for English: What next? *Changing English*, 17: 2, 189–200. DOI:10.1080/13586841003787332.

Coleman, James A. (2009). Why the British do not learn languages: Myths and motivation in the United Kingdom. *The Language Learning Journal*, 37: 1, 111–127.

Department for Education and Science (DfES) (2002), *Languages for All, Languages for Life: A Strategy for England*. London: H.M.S.O.

Department for Education and Science (DfES) (2006), *Independent Review of the Teaching of Early Reading* (Rose Report). London: H.M.S.O.

Department for Education (DfE) (2013a), *National Curriculum in England: Languages Programmes of Study*. [Online: https://www.gov.uk/government/publications/national-curriculum-in-england-languages-progammes-of-study/national-curriculum-in-england-languages-progammes-of-study. 01.11.2021].

Department for Education (DfE) (2013b), *National Curriculum in England*: *English Programmes of Study: Key Stages 1 and 2*. [Online: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/335186/PRIMARY_national_curriculum_-_English_220714.pdf. 01.11.2021]

Department for Education (2018), *Key Stage 2 Tests: 2018 English Grammar, Punctuation and Spelling Test Materials*. [Online: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/709997/STA187970e_2018_ks2_English_GPS_Paper1_questions.pdf.pdf. 24.08.2021].

Dobson, Alan (2018), Towards 'MFL for all' in England: A historical perspective. *The Language Learning Journal*, 46: 1, 71–85. DOI: 10.1080/09571736.2017.1382058.

Driscoll, Patricia/ Jones, Jane/ Macrory, Gee (2004), *The Provision of Foreign Language Learning for Pupils at KS2*. DfES Report.

Duff, Fiona J./ Melonig, SilvanaE./ Bailey, Alison M./ Snowling, Margaret J. (2015), Validity and sensitivity of the phonics screening check: implications for practice. *Journal of Research in Reading,* 38: 2, 109–123. DOI:10.1111/1467-9817.12029.

Ellis, Sue/ Moss, Gemma (2014), Ethics, education policy and research: The phonics question reconsidered. *British Educational Research Journal,* 40: 2, 241–260.

Enever, Janet (2018), *Policy and Politics in Global Primary English*. Oxford: Oxford University Press.

Erler, Lynn (2004), Near-beginner learners of French are reading at a disability level. *Francophonie,* 30, 9–15.

Erler, Lynn/ Macaro, Ernesto (2011), Decoding ability in French as a foreign language and language learning motivation. *The Modern Language Journal,* 95: 3, 496–518.

Gayton, Angela M. (2018), A context-specific approach to L2 motivation in Anglophone settings: A first step towards theory development. *The Language Learning Journal*, 46: 4, 384–397. DOI: 10.1080/09571736.2015.1130081.

Graham, Suzanne/ Courtney, Louise/ Marinis, Theo/ Tonkyn, Alan (2014), *Primary Modern Languages: The impact of teaching approaches on attainment and preparedness for secondary school language learning. Final report*. Nuffield Foundation.

Hagger-Vaughan, Lesley (2016), Towards 'languages for all' in England: The state of the debate. *The Language Learning Journal*, 44: 3, 358–375.

Lanvers, Ursula/ Coleman, James A. (2017), The UK language learning crisis in the public media: A critical analysis. *The Language Learning Journal*, 45: 1, 3–25. DOI: 10.1080/09571736.2013.830639.

Koda, Keiko (2005), *Insights into Second Language Reading: A Cross-Linguistic Approach*. Cambridge: Cambridge University Press.

Koda, Keiko (2008), Impacts of prior literacy experience on second-language learning to read. In: Koda, Keiko & Zehler, Annette M. (eds.), *Learning to Read Across Languages: Cross-Linguistic Relationships in First- and Second-Language Literacy Development*. Abingdon: Routledge.

Kolb, Annika (2007). How languages are learned: Primary children's language learning beliefs. *Innovation in Language Learning and Teaching*, 1, 227–241.

Lüth, Jessica (2019). *Modern Language Education Policy and its Implementations in Primary Schools in Urban European Contexts*. Unpublished PhD thesis, King's College, London.

Mackey, Alison/ Gass, Sue M. (2015). *Second Language Research: Methodology and Design* (2nd edn.). London: Routledge.

Macrory, Gee (2019), New Orthographies in the Primary Languages Classroom: A Challenge for Teacher Education. In: Zain, Subhan & Garton, Sue (eds.) *Early Language Learning and Teacher Education.* Clevedon: Multilingual Matters, 98–115.

Millonig, Diana J./ Stickler, Ursula/ Coleman, James A. (2019), Young pupils' perceptions of their foreign language learning lessons: The innovative use of drawings as a research tool. *The Language Learning Journal*, 47: 2, 229–245. DOI: 10.1080/09571736.2016.1270348.

Muñoz, Carmen. (2014), Exploring young learners' foreign language learning awareness. *Language Awareness,* 23: 1–2, 24–40.

Murphy, Victoria A./ Macaro, Ernesto/ Alba, Sonia/ Cipolla, Claudia (2015), The influence of learning a second language in Primary school on developing first language literacy skills. *Applied Psycholinguistics,* 36, 1133–1153.

Newby, Peter (2010), *Research Methods for Education.* Harlow: Pearson Education Ltd.

Nuffield Foundation (2000), *Languages: The Next Generation*. London: Nuffield Foundation.

Ofsted (2012), *Moving English Forward: Action to Raise Standards in English*. London: H.M.S.O.

Ofsted (2015), *Key Stage 3: The Wasted Years*? London: H.M.S.O.

Ofsted (2016), *The Annual Report of Her Majesty's Chief Inspector of Education, Children's Services and Skills, 2015–2016*. London: OFSTED.

Porter, Alison (2014), *An Early Start to French Literacy: Learning the Spoken and Written Word Simultaneously in English Primary Schools*. Unpublished PhD thesis, University of Southampton.

Rafat, Yasaman (2016), Orthography-induced transfer in the production of English-speaking learners of Spanish. *The Language Learning Journal*, 44: 2, 197–213.

Safford, Kimberley (2016), *Teaching grammar and testing grammar in the English Primary school: The impact on teachers and teaching of the grammar element of the statutory test in Spelling, Punctuation and Grammar (SPaG) in England*. UKLA and The Open University. DOI: https://doi.org/10.1080/1358684X.2015.1133766.

Sampson, Geoffrey (1985), *Writing Systems*. London: Hutchinson Education.

Teaching Schools Council (2016), *Modern Foreign Languages Pedagogy Review: A Review of Modern Foreign Languages Teaching Practice in Key Stage 3 and Key Stage 4*. [Online: https://www.tscouncil.org.uk/wp-content/uploads/2016/12/MFL-Pedagogy-Review-Report-2.pdf. 24.08.2021].

Tinsley, Teresa (2013), *The State of the Nation: Demand and Supply of Language Skills in the UK*. London: British Academy.

Tinsley, Teresa/ Board, Kathryn (2015), *Languages for the Future: Which Languages the UK Needs Most and Why*. London: British Council.

Tinsley, Teresa/ Board, Kathryn (2016), *Language Trends 2015/16: The State of Language Learning in Primary and Secondary Schools in England*. Reading: Education Development Trust.

Tinsley, Teresa/ Board, Kathryn (2018), *Language Trends 2016/17: Language Teaching in Primary and Secondary Schools in England. Survey Report*. London: British Council.

Tinsley, Teresa / Doležal, Neela (2018), *Language Trends 2018. Language Teaching in Primary and Secondary Schools in England. Survey Report*. London: British Council.

Wade, Pauline/ Marshall, Helen/ Sharon O'Donnell (2009), *Primary Modern Foreign Languages Longitudinal Survey of Implementation of National Entitlement to Language Learning at Key Stage. 2 Final Report*. London: National Foundation for Educational Research for DCFS.

Walker, Matthew/ Sainsbury, Marian/ Worth, Jack/ Bamforth, Heather/ Betts, Helen (2015), *Phonics Screening Check Evaluation: Final Report*. London: National Foundation for Educational Research for DfE.

Wingate, Ursula (2018), Lots of games and little challenge – a snapshot of modern foreign language teaching in English secondary schools. *The Language Learning Journal*, 46: 4, 442–455. DOI: 10.1080/09571736.2016.1161061.

Woore, Robert (2009), Beginners' progress in decoding L2 French: Some longitudinal evidence from English modern foreign languages classrooms. *The Language Learning Journal*, 37: 1, 3–18.

Woore, Robert (2010), Thinking aloud about L2 decoding: an exploration of the strategies used by beginner learners when pronouncing unfamiliar French words. *The Language Learning Journal*, 38: 1, 3–17.

Woore, Robert (2018), Learners' pronunciations of familiar and unfamiliar French words: What can they tell us about phonological decoding in an L2? *The Language Learning Journal*, 46: 4, 456–469. DOI: 10.1080/09571736.2016.1161062.

Annika Kolb

Story apps – new ways in teaching reading in primary EFL?

Abstract The multimodal nature of story apps offers new opportunities for independent reading in the foreign language even at primary school level. Along with the written text, they provide animations, sound and interactive elements which could be helpful for primary school students who have limited linguistic abilities and reading experience in the foreign language. In a research project on story apps in a German primary EFL classroom, three major types of support for the reading process could be identified: providing comprehension support, sustaining reading motivation and supporting the development of reading strategies. Based on these affordances of story apps it is outlined to what extent story apps can contribute to a learner-oriented approach to reading education that caters for individual students' interests and needs.

1. Introduction

This paper sheds light on the potential of story apps for individualized reading in the primary foreign language classroom. It presents a research project that implemented an extensive reading approach to give students the possibility to read stories in the foreign language English on their own and choose texts on different linguistics levels and according to personal interests. It was investigated how the features of story apps can support text comprehension and the development of reading strategies. The analysis of different types of support that story apps can provide for reading activities in the primary foreign language classroom shows to what extent story apps can help to implement a learner-oriented approach in reading education.

2. Reading in the primary EFL classroom

Picturebooks are an essential component of primary EFL methodology. They are supposed to foster children's interest in literature, provide motivating language learning opportunities (Bland 2013; Mourão 2015) and might support the development of literary competences (Frisch & Alter

2018). However, they are most commonly presented by the teacher in a storytelling scenario (e.g., Brewster & Ellis 2012: 186–202). Frequently, it is the teacher who reads the story to the children as a class, whereas the children rarely get the opportunity to read unknown picturebooks by themselves.

Despite the benefits of storytelling, this teacher-centred approach also has two major disadvantages. First, research has shown that the use of the written language is beneficial for learners' language development, even at the beginning stage of their learning (Duscha 2007; Frisch 2015). It prevents the children from developing their own strategies for putting language into writing ("invented spellings", cf. Rymarczyk 2008). Furthermore, language competence develops in an integrated way, which means competences that have been acquired in one language can be transferred to another language. Reading strategies that have been developed in children's first language can – at least partially – be applied when reading texts in the foreign language (Frisch 2015, Reckermann 2018). The use of writing in the foreign language can in turn foster children's reading competence in their L1 (Diehr & Rymarczyk 2012). Also for motivational reasons, children should not be denied access to written text. Students who have learnt to read and write in the school language German expect to use this competence also in the foreign language and are very interested in reading and writing in English (Frisch 2015: 17). Especially picturebooks are very appealing to children. Reading a whole book in English by themselves provides children with a sense of achievement that can boost their confidence and motivation (Kolb 2013; Reckermann 2016). They can enjoy reading as a pleasurable and aesthetic experience. Studies into primary school learners' reading competences in the foreign language show that the competence standards defined by curricula are often surpassed (e.g., BIG-Kreis 2015: 40–44), which indicates that children are more often under- than overstrained.

The second disadvantage of denying primary school students access to independent reading is the chance to cater for different learning conditions, competence levels and interests. Instead of one story that is chosen by the teacher for the whole class, a selection of texts with a variety of topics and different language levels can be provided if children read by themselves. This allows for taking different dimensions of heterogeneity in the

classroom into account, for example students' diverse language learning experiences, learning styles and personal interests (Chilla & Vogt 2017).

The extensive reading approach puts independent reading and individual choice of texts at the centre. It can foster reading motivation, reading competence and general language competence (e.g. Biebricher 2008; Day & Bamford 1998; Krashen 2007, 2013). While most studies have been carried out with older learners, extensive reading is less popular in primary school. Proponents of this approach still claim that there is also potential for younger learners: "Extensive reading is appropriate at all stages of language learning, it is never too early […] to learn to read in a second language" (Day & Bamford 1998: xiv, see also Reckermann 2018). However, first experiences with this approach in primary school show that young learners often struggle with the opaque grapheme-phoneme correspondence of the English language. In a study on extensive reading with picturebooks in primary school (Kolb 2013) children had considerable difficulties: Frequently, they tried to read the texts aloud and then had trouble with the unknown pronunciation. This strategy significantly hindered their understanding of the text. Giving students an audio version of the text only partly solved the problem. While it helped them to understand the words and reduced their confusion through unknown spellings, the children very often lost track in the reading process and could not follow the text anymore. The multimedia and interactive features of story apps could alleviate this difficulty.

3. Story apps

Over the last few decades the rapid development of digital technologies has brought about new picturebook formats. Digital picturebooks for electronic devices (e.g. tablets, mobile phones, computers) are now produced, ranging from unaltered digital versions of texts that had originally been published in a paper-based format to story apps which feature auditory, performative and interactive elements (Al-Yaqout & Nikolajeva 2015; Hamer 2017; Sargeant 2015). A number of these features make story apps particularly interesting for foreign language learners. Learners can choose their preferred mode of reading (e.g. *read to me* or *read by myself*) and can navigate the reading process (e.g. by using thumbnail images to go

forward/backward, pause, re-read words or whole passages). Many story apps feature audio narration, that is children can have the text read to them while they read along. This is usually supported by text highlighting that uses colour to mark the word that is being read at the moment. With some story apps elements of the illustration are labelled when clicking on them, thereby providing vocabulary support. In addition, there are interactive features that ask readers to take part in the story, fulfil tasks, solve problems and even choose characters, settings or specific story paths. To this end, the technical possibilities of mobile devices are fully exploited. Touchscreen technology allows the users to influence the actions on the screen, the tablet computer or smartphone has to be shaken or tilted, users blow into the microphone and use the camera as a looking glass in which they see themselves as characters in the story. Thus, interaction between text and readers takes places not only on a cognitive, but also on a physical level (Manresa 2015: 108). There are varied possibilities to adapt the story app according to personal needs: written text can be shown or hidden, background noises can be switched on or off, the speed of audio narration and the language can be chosen (for these and further characteristics of story apps cf. Bircher 2012; Cahill & McGill-Franzen 2013; Stichnothe 2014; Turrión 2015; Yokota 2015).

First studies that tried to find out to what extent story apps can support reading comprehension have produced inconsistent results. Positive effects were found concerning reading motivation, vocabulary learning and comprehension support. Through readers' active engagement and their involvement in the story, story apps can increase reading motivation (Ciampa 2012; Ertem 2010). The multimodal presentation of the story can support comprehension. Animations, music and sounds can help readers to make meaning from the text as they can link (moving) pictures and auditory information with the content of the written text:

> In living books visual elements that are normally compressed into just one static illustration are instead split into several smaller portions, each representing one element of the narration. By synchronizing phases in the narration with portions of the picture there is a higher probability that connections will be made between words and non-verbal information (Bus, Verhallen & van der Kooy-Hofland 2009: 17).

However, this seems to be possible only if the different modes of information are congruent and are presented simultaneously to allow the readers to make connections between the different types of information. If this is not the case, the interactivity of story apps may disturb the reading process and the understanding of stories. Children then tend to focus their attention on trying out the hotspots (interactive buttons) that are irrelevant to the text, which can lead to cognitive overload and distraction from the content of the stories (Bus, Takacs & Kegel 2015: 92; Korat, Shamir & Segal-Drori 2015; Miller & Warschauer 2014; Smeets & Bus 2013, 2014; Takacs, Swart & Bus 2015: 3; Verhallen, Bus & de Jong 2006; Verhallen & Bus 2010). In general, cognitive load (Sweller 2005) seems to play an important role; if there are too many diverging pieces of information, the danger of cognitive overload arises, students cannot process the information anymore which hinders their understanding (Smeets & Bus 2013: 180).

The change between different modes of action is another challenge in this context. Manresa (2015) reports that children in her study perceived a tension between the actual story and the interactive elements. These were interpreted as a parallel game that was going on, making the students choose between reading a story and playing a game (Manresa 2015: 109). This tension is confirmed by Takacs et al. who argue for a close congruency between the plot and the interactive features:

> Story comprehension and playing with hotspots or games are two fundamentally different tasks, even when their content is related, and carrying out both requires task switching. […] the more closely related the story and the interactive additions are, the smaller the cognitive cost of switching between the two tasks is (Takacs et al. 2015: 4).

Hayles (2007) describes a similar challenge when she distinguishes between two types of attention: While deep attention, the ability to focus on an object intensively and persistently, is necessary to follow a narrative for a longer period of time, hyper attention, the ability to switch between tasks, is required during the reading process to cope with multiple modes of information. Hayles hypothesizes that "we are in the midst of a generational shift in cognitive styles" (2007: 187) and that young people need the challenge of both hyper and deep attention opportunities.

Concerning vocabulary acquisition, some studies have shown that the interactive features of story apps can have similar effects as reading aloud to children (Korat & Shamir 2007; Smeets & Bus 2014). Animations and sound effects support students in focusing on relevant details that are mentioned in the written text and to make a connection between words and images.

Whereas most studies on story apps have been conducted in L1 learning contexts, there is little evidence on their potential for foreign language learning.

4. Research design

The aim of the study presented here was to further explore the potential of story apps for independent reading. The qualitative research design follows a classroom action research approach (Burns 2010). A voluntary afternoon programme, the *English Book Club*, was established for third and fourth graders (children between the ages of 8 and 11) at a German primary school. The children had been learning English since Grade 1 with two 45-minute sessions each week. About 20 % of the children speak an L1 other than German at home. In an extensive reading setting the children were offered a variety of different paper-based picturebooks and story apps. They chose the stories they wanted to read and were encouraged to read with a partner to promote cooperative learning. For research purposes, the discussions between the students were helpful to identify comprehension strategies. Thus, independent reading did not necessarily mean individual reading.

Each picturebook or story app was accompanied by pre- and post-reading activities that focus on aspects of the content and language of the stories. These activities try to stimulate the children's use of reading strategies, by asking them, for example, to predict actions and events or to guess the meaning of unknown words from the context. Furthermore, the children were asked to note the main facts of the story (e.g. summary of the content, author, setting and characters) in a story map to check their comprehension of the plot. Last the students gave their opinion on the story in a rating grid.

The study focuses on the following research question: *What kind of support can the features of story apps provide for the learners' reading comprehension?*

In order to answer this question, a variety of data collection methods were employed. Video recordings were used to document how the children proceed when reading the story apps independently. As well as their discussions in pairs, the cameras also document the children's actions (their non-verbal behaviour) and contextual information (what is on the screen, and what the children do with the story apps). Since video recordings do not give explicit insights into the children's thoughts and their mental actions, the young learners were interviewed about their reading experiences and strategies retrospectively at the end of each school year. The children were asked to explain how they tried to make meaning from the text and encouraged to give reasons for the course of actions they took while reading the story apps. In addition, the learner texts provide valuable insights into the children's evaluation of the story apps and their level of comprehension. Two research cycles have been conducted so far.

The video recordings, the interviews and the children's worksheets were analyzed using Qualitative Content Analysis (Mayring 2014). While the interviews were analyzed after they had been transcribed completely, software for qualitative data analysis (MAXQDA) allowed for the categorization of the video data without prior transcription. This led to the identification of interesting sequences which were then transcribed and underwent a detailed sequential analysis. As a result, three major categories of support for the children's reading process emerged that will be presented in the next section.

5. How story apps can support the reading process

The data analysis identified providing comprehension support, sustaining reading motivation and supporting the development of reading strategies as affordances of story apps (Brunsmeier & Kolb 2017, 2018; Kolb & Brunsmeier 2018, 2019). In the following, these categories are further outlined and provide the basis for discussing the potential of story apps for individualized reading instruction.

5.1 *Providing comprehension support*

Compared to traditional picturebooks, the audio narration function plays a prominent role for the challenge to read an unknown text in the foreign language. It enables the children to read along and to link written and oral language. Therefore, the process could be characterized as reading-listening-comprehension. It is further supported by the text highlighting as well as by the voices of different speakers for different characters in the story. In the student interviews, one child describes how audio narration facilitates the reading task:

> *Wir lesen besser, da wir das hören, was die uns sagen, die von iPad drinnen, von die Buch* [We read better, because we hear what they tell us, the ones that are inside the iPad, in the book.] (Viktor 150715, l.58-59)[1]

Audio narration seems to be especially helpful if it is reader-activated, that is if the children can choose when to start being read to. This gives them control over their reading process and helps them to match written and spoken language at their own pace. That this matching is important for the children shows this comment in which the student criticizes that the written text disappeared too quickly from the screen:

> *Bei manchen Apps ist ja auch das bei der Schrift dann immer verschwunden und dann konnte ich halt nicht mehr lesen. (Mitschüler/in stimmt zu). Weil ich bin ja auch nicht so schnell.* [In some of the apps, the written text disappeared and I could not read the text again (classmate confirms), because I'm just not that fast.] (Luise, 150715, l. 39–41)

Compared to audio versions of paper-based picturebooks, the read-to-me function allows the children to adapt the speed of the reading process to their individual needs. They also very much appreciate being able to repeat individual words or sentences. One student commented:

> *Man kann, wenn man ein Wort nicht versteht, nicht verstanden hat, nochmal drauf drücken und dann sagt er es nochmal* [You can, if you did not understand a word, you can press on the word again and then he says it again.] (Leon, 150715, l. 76–77)

1 As the student interviews are in German, an English translation is provided in square brackets. The children's names have been pseudonymized.

Animations proved to be the second element of story apps that significantly supported text comprehension. Compared to paper-based picturebooks that can only represent actions in a static way, story apps offer dynamic illustrations. Some students referred to this aspect in the interviews:

> *Weil man sieht ja auch, ja der macht das jetzt.* [Because you actually see, yes, that's what he's doing now.] (Lena, 210616, l. 62)
> *Zum Beispiel steht dort 'Die Äste bewegten sich vom Baum' und das ist eben dann auf Englisch dort geschrieben und dann wird das im Bild dargestellt, wie der Ast so sich bewegt.* [It says, for example: 'The branches of the tree were moving' and that then is written in English and then the picture shows the branch moving.] (Colin, 150715, l. 112–113)

In addition, the animations frequently enable the children to guess unknown words from the context and support them in focusing their attention on important details in the text.

There are certain circumstances which proved to be particularly supportive for understanding animations and sounds. As with audio narration, the children found animations very helpful when they were *reader-activated*, as this student explains from his experience with the story app *The Three Little Pigs* (Nosy Crow & Bryan 2015a):

> *Es hat dann geholfen, dass man die Bilder auch so gut bewegen kann. Wenn man sie berührt dann haben sie zum Beispiel ihr Haus angefangen zu bauen.* [It was of great help that you could move the pictures so well. Once you touch the characters they start to build a house.] (Louis, 130415, l. 64–65)

Another aspect that could either support or hinder understanding was the timing of the animations. Animations that could be activated before the text of the slide had been read aloud in full often led the children to start these and focus their attention on the animations and actions rather than on reading and listening to the text. Moreover, it is very important that the different sources of information (words, pictures, movement, sound) correspond to each other. Many story apps feature reader-activated animations that do not contribute to the understanding of the story, as this student notes:

> LEON: *Bei* 'Three Little Pigs' *fand ich was geil. Habe ich was Cooles herausgefunden. Wenn man auf die Leute tippt und dann hochzieht, dann können die Saltos machen.* [I found out something wicked, really cool when I read the *Three Little Pigs*. When you tap on the characters and drag them, they can do somersaults.]

INTERVIEWER: *Aber hilft dir das beim Verstehen?* [Does this help you to understand the story?]
L: *Nee, aber fand ich cool.* [No, but it was cool.] (Leon, 150715, l. 104–108)

These interactive elements that distracted the students from the story's plot proved to be the major challenge. Especially if the reader-activated animations can be started before the text has been read out, many children tend to focus their attention on the opportunities to trigger animations and sounds. As a consequence, animations and sounds proved to be decisive criteria for selecting appropriate story apps since these features could both support or distract from the understanding of the story.

Furthermore, vocabulary aids supported comprehension especially on word level. Some story apps offer explicit vocabulary support by labeling pictures: written words appear next to objects and characters in pictures. This helps the reader derive meaning from the text and draw inferences. This kind of vocabulary support is usually reader-activated – the words appear if the reader taps on a specific place in the picture. It proved to be especially helpful for the children if the words show up right next to the relevant object and can therefore be unambiguously matched. It was also most supportive for text comprehension if the words shown actually appear in the text. Additional words provide further language input but could also distract the young learners from the story.

5.2. *Sustaining reading motivation*

The data showed that several factors seem to enhance and keep up children's reading motivation. Again, the animations play an important role. Some children report that they help them to focus on the story:

Weil man ist da auch aufmerksamer, weil da was passiert auf dem Bildschirm [*Because you are more attentive, because something is happening on the screen.*] (Simon, 190316, l. 79–82)

Especially the enhanced possibilities of interaction between readers and text lead to children's involvement. In some story apps, users are asked to complete tasks within the story and to solve problems for the protagonists, for example in *Jack and the Beanstalk* (Nosy Crow & Bryan 2015b) they have to find a key in the giant's castle to open a specific door. In *Snow White* (Nosy Crow & Bryan 2015c) they have to mix a potion for the wicked stepmother:

> *Damit man weiterkommt. Man muss erstmal so eine Aufgabe machen und wenn sie richtig gemacht wurde, ist da so ein Pfeil, auf den drückt man und dann kommt man halt weiter (…) Zum Beispiel bei* 'Nash Smasher'*, da musste man so eine Burg kaputt machen und dann musste man halt so eine Linie ziehen und dann hat er mit dem Hammer so eine Burg zerstört und das musste man öfter machen, bis die Burg kaputt gegangen ist*. [To proceed with the story we had to accomplish a task. If you complete it successfully, there is an arrow, which you have to press and then you can continue. For example, in the story app *Nash Smasher* we had to destroy a castle. We had to draw a line and then Nash smashes a castle with a hammer. We had to do that a number of times until the castle was completely destroyed.] (Sascha, 150715, l.87–89).

These tasks within the story fulfill two functions: first, they can provide additional incentives to understand the text because their completion is necessary to continue with the story ("damit man weiterkommt"). Furthermore, they check the readers' understanding of the story: users get immediate feedback on whether they have successfully completed the task which can enhance their motivation and support intensive engagement with the text (e.g. by prompting them to read passages again and to look for specific information). Second the children's reaction in the study show that this kind of involvement makes them feel like *players in the story*, *co-protagonists* (Hamer 2017: 69) who immediately experience the events instead of just observing them from outside. When two boys help the wolf in the *Three little pigs* (Nosy Crow & Bryan 2015a) to destroy the pigs' house by blowing into the tablet computer's microphone, one of them says: "Wir sind gemein" [We're mean] (Noah, V2 170615, 02:27). Not the wolf is being characterized as mean, but they themselves become protagonists in the story and identify with the villain's role. Mourão (2016) who has adapted Sipe's (2008) categories of response to literature for the EFL classroom describes this as a transparent response. The children react "as though they were living the story for real" (Mourão 2016: 35).

5.3 *Supporting the use of reading strategies*

Especially when reading texts in a foreign language, reading strategies can help to cope with comprehension problems (Grabe 2009; Paris, Wasik & Turner 1996). As the data of the study show, story apps seem to foster the use of specific reading strategies.

A noticeable strategy is making assumptions about the course of events in the story and revising them in the course of the reading process. When reading the story apps *Zoe's Green Planet* (Square Igloo, Tousnakhoff & Roussel 2013), two children discuss their ideas:

> MERYAM: Sie [Außerirdische] wollen auf die Erde fliegen? [*They (aliens) want to fly to the earth?]*
> JASEMIN: Mh, nee, also es ist ja, glaub ich, dass sein Raumschiff landet und fliegt nicht weg [*No, I think a space ship lands and doesn't fly away*].

After reading the first sentence of the story app, the students continue with their predictions:

> MERYAM: Ja und vielleicht, weil es rot ist, kommen da rote Männchen raus [and perhaps, because it's red, little red men will come out]. (Meryam, Malea, Jasemin 200616, l. 75–78)

Prediction is a very prominent strategy in the children's reading processes. When in the story app *Pete's Robot* (Heart-Drive Media 2016) a `heart-drive' is implanted into the robot, one child suspects for example: "Jetzt ist er wieder nett geworden, oder?" [Now he has turned friendly again, hasn't he?] (150715, observation documented in research diary). The student has come to the conclusion that a missing heart must have been the cause for the robot's destructive and unfriendly behavior, probably based on his prior knowledge that having a heart means being loving and caring.

Because the images in story apps are not static but change, the children's pleasure in making assumptions about the story is triggered. Another reason why story apps seem to lend themselves well to fostering predictions is that some of them actively ask children to speculate about elements of the story. They do so by hiding objects that have to be found, by showing only parts of objects and by asking children to choose between different options. One child describes the pleasure of these guessing games, referring to the picture of a shell that was not revealed in the beginning:

> Also ich fand das voll spannend, wo man nur die Augen mal so gesehen hat, weil dann konnte man auch kurz überlegen, was das jetzt sein könnte [*I found that really exciting, when you could only see the eyes, because then you could think about what that could be*] (Miriam 200616, l. 183–185)

That the children also very often revise their initial predictions can be seen in another sequence. In the pre-reading activity on the story app *Dino Boy* (Three Thumbs Up/ Thomas/ Jorgensen & Voigt 2011), a picture of a slide is presented along with the text *Dino Boy sees a big yellow slide through the window. He goes down the big yellow slide.* They are asked to speculate on the question *Where will it take him?* One group of students guesses that the slide will take the protagonist into a dinosaur world. Once they start reading the story, they first confirm their assumption: "Oh, Dinowelt – so wie wir es gesagt haben" [*Oh, dinosaur word – just as we said*]. However, as the story continues, they realise that they were mistaken because the slide does not take the boy to a dinosaur world: "Doch nicht ne Dinowelt" [*not a dinosaur world after all*] (Jonas und Philipp, 100615, 04:27–5:00).

Making predictions serves various purposes in the reading process: by ruling out specific events the readers limit the range of possibilities they could encounter in the story (Nuttal 2005: 13). It sets them on the right track so that they can follow the plot more easily even if they do not understand the whole text. Furthermore, speculating about the content of the story engages the readers with the text (ibid., p. 119). They are actively searching for clues that would prove or disprove their assumptions, giving the reading process a certain purpose.

6. Conclusion

The project presented here was initiated with the intention to enable children to independently read stories in the foreign language to cater for individual interests and competence levels. Based on the analysis of the different types of support story apps can provide, three aspects could be identified that can contribute to a learner-oriented approach (Schocker 2016) to reading in the primary EFL classroom: individual reading experiences, individual support and room for individual reactions to the text.

As far as *individual reading experiences* are concerned it became clear that due to the comprehensive support provided by story apps, children are in fact able to read stories on their own. The learner texts show that they can complete activities like reconstructing the storyline, matching pictures and sentences and fill in gap texts that require an understanding of the stories'

plot. Independent reading allows for individual choices in terms of content instead of having to make do with the teacher's choice. Furthermore, story apps cater for individual preferences in terms of accessing stories. Even though the linguistic level is necessary to understand the story, children can still start to approach the story app through the pictures and the actions which may facilitate access for reluctant readers. In some story apps children can determine the quantity of linguistic input by tapping on characters. At times, choices can be made regarding characters, the setting or the plot. However, the decision-making possibilities within the narrative structure are usually quite limited, the linear structure is rarely broken up and there are not really different strands of action (cf. Ritter 2013: 16).

Individual support is provided as, to some extent, story apps allow to cater for different competence levels. Some features that proved to support the children's understanding can be applied according to one's needs. Audio narration, for example, can usually be switched off, so that more advanced children do not necessarily have to use the read-aloud function. Animations, vocabulary help and tasks that users have to fulfill within the story can serve as individualized support especially if they are reader-activated. Only if students can decide by themselves whether and when they want to use animations, vocabulary help or tasks, this individualized support facilitates understanding. However, faced with the multiple offers of animations, games, noises and interactive features, it proved to be challenging for the students to not lose the focus on the actual story. Some story apps feature the text in different languages. For students with an L1 other than German, this could also be of help (cf. multilingual digital storybooks Bündgens-Kosten & Elsner 2014).

Room for individual reactions to the text as a third aspect when it comes to take learners' individual needs into account, is to give learners the possibility to show their personal reception of the text. Literary competence entails the ability to talk about reading experiences, analyse and interpret texts and express these interpretations in creative ways (Frisch & Alter 2018). Some story apps provide interesting opportunities to do so: children can, for example, record their own soundtrack to the story. They can either act out the original dialogues in an audio play or retell the story in their own words. Further examples of follow-up tasks would

be to create alternative endings by recording a soundtrack to screenshots from the app or drawings by the children that have been photographed or to develop their own animated scenes, for example to create another challenge for Jack in the story app *Jack and the Beanstalk* (Brunsmeier & Kolb 2016). In these tasks, the children develop texts that take up the story apps' multimodality and link visual, verbal and acoustic elements.

By providing comprehension support, helping to sustain reading motivation and developing reading strategies story apps can facilitate independent reading even at an early stage of the learning process. It has to be taken into consideration, however, that the interactivity of story apps is also a challenge since it can distract children from the story's plot. This is particularly the case if text, pictures and animations are not congruent and do not accommodate both deep and hyper attention in a productive way. In a classroom setting, it is therefore all the more important to develop suitable pre-, while- and post-reading activities that have the chance to impact the students' reading process. They should focus on the verbal level of the stories, give students feedback on their understanding of the plot, model the use of reading strategies and leave room for subjective reactions to the text.

Bibliography

Al-Yaqout, Ghada/ Nikolajeva, Maria (2015), Re-conceptualising picturebook theory in the digital age. *Nordic Journal of ChildLit Aesthetics* 6.[Online.https://www.tandfonline.com/doi/abs/10.3402/blft.v6.26971. 16.08.2021].

Biebricher, Christine (2008), *Lesen in der Fremdsprache. Eine Studie zu Effekten extensiven Lesens*. Tuebingen: Narr.

BIG-Kreis (2015), *Der Lernstand im Englischunterricht am Ende von Klasse 4. Ergebnisse der BIG-Studie.* München: Domino-Verlag.

Bircher, Katie (2012), What makes a good picture book app? *The Horn Book Magazine* 28, 72–78.

Bland, Janice (2013), *Children's Literature and Learner Empowerment.* London: Bloomsbury.

Bland, Janice/ Lütge, Christiane (eds.) (2013), *Children's Literature in Second Language Education*. London: Bloomsbury.

Brewster, Jean/ Ellis, Gail (2012), *Tell it again! The Storytelling Handbook for Primary English Language Teachers*. London: British Council.

Brunsmeier, Sonja/ Kolb, Annika (2016), Jack and the Beanstalk today – facing new challenges. Ein digitales Bilderbuch lesen und dazu eine Videosequenz entwickeln. *Der fremdsprachliche Unterricht Englisch* 144, 12–19.

Brunsmeier, Sonja/ Kolb, Annika (2017), Picturebooks go digital – the potential of story apps for the primary EFL classroom. *CLELE Journal* 5.1, 1–17.

Brunsmeier, Sonja/ Kolb, Annika (2018), 'I like the character, weil er so richtig funny ist' – Reading digital picture books in the primary EFL classroom. In: Buendgens-Kosten, Judith & Elsner, Daniela (eds.) (2018), *CALL in Multilingual Settings*. Bristol: Multilingual Matters, 41–58.

Bündgens-Kosten, Judith/ Elsner, Daniela (2014), Rezeptives Code-Switching ein- und mehrsprachiger Lerner/innen in multilingualen Settings. *Fremdsprachen Lehren und Lernen* 42.2, 56–73.

Burns, Anne (2010), *Doing Action Research in English Language Teaching. A Guide for Practitioners*. New York: Routledge.

Bus, Adriana/ Takacs, Zsofia/ Kegel, Cornelia (2015), Afforcances and limitations of electronic storybooks for young children's emergent literacy. *Developmental Review* 35, 79–97.

Bus, Adriana/ Verhallen, Marian/ van der Kooy-Hofland, Verna (2009), Digital picture storybooks. *Better: Evidence-based Education* 1, 16–17.

Cahill, Maria/ McGill-Franzen, Anne (2013), Selecting 'app'ealing and 'app'ropriate book apps for beginning readers. *The Reading Teacher* 67.1, 30–39.

Ciampa, Katia (2012), Electronic storybooks: A constructivist approach to improving reading motivation in Grade 1 students. *Canadian Journal of Edcuation* 35.4, 92–136.

Chilla, Solveig/ Vogt, Karin (2017), Englischunterricht in heterogenen Lerngruppen: eine interdisziplinäre Perspektive. In: Chilla, Solveig & Vogt, Karin (Hrsg.) (2017), *Heterogenität und Diversität im Englischunterricht*. Frankfurt a.M.: Peter Lang, 55–82.

Day, Richard/ Bamford, Julian (1998), *Extensive Reading in the Second Language Classroom*. Cambridge: Cambridge University Press.

Diehr, Bärbel/ Rymarczyk, Jutta (2012), Current developments in teaching English as a foreign language in primary schools. *Anglistik. International Journal of English Studies*, 23.1, 13–24.

Duscha, Michael (2007), *Der Einfluss der Schrift auf das Fremdsprachenlernen in der Grundschule. Dargestellt am Beispiel des Englischunterrichts in Niedersachsen.* Dissertation. [Online: https://publikationsserver.tu-braunschweig.de/receive/dbbs_mods_00021088. 16.08.2021].

Ertem, Ihsan Seyit (2010), The effect of electronic storybooks on struggling fourth-graders' reading comprehension. *Turkish Online Journal of Educational Technology* 9.4, 140–155.

Frisch, Stefanie (2015), Eckpunkte einer theoretisch und empirisch basierten Lesedidaktik für den Englischunterricht der Grundschule. In: Kötter, Markus & Rymarczyk, Jutta (Hrsg.) (2015), *Englischunterricht auf der Primarstufe. Neue Forschungen – weitere Entwicklungen*. Frankfurt: Peter Lang, 15–34.

Frisch, Stefanie & Alter, Grit (2018), Literarisches Lernen schon in der Grundschule? The Potential of Picturebooks. *Grundschule Englisch* 63, 2–5.

Grabe, William (2009), *Reading in a Second Language. Moving from Theory to Practice*. Cambridge: Cambridge University Press.

Hamer, Naomi (2017), The Design and Development of the Picture Book for Mobile and Interactive Platforms: "You get to BE Harold's Purple Crayon". In: Hamer, Naomi; Nodelman, Perry & Reimer, Mavis (eds.) (2017), *More Words about Pictures – Current Research on Picture Books and Visual/Verbal Texts for Young People*. New York: Routledge, 63–80.

Hayles, Katherine N. (2007), Hyper and deep attention: The generational divide in cognitive methods. *Profession* 2007, 187–199.

Hudson, Thom (2007), *Teaching Second Language Reading*. Oxford: Oxford University Press.

Kolb, Annika (2013), Extensive Reading of Picturebooks in Primary EFL. In: Bland, Janice & Lütge, Christiane (eds.) (2013), *Children's Literature in Language Education*. London: Bloomsbury, 33–43.

Kolb, Annika/ Brunsmeier, Sonja (2018), Story apps: The challenge of interactivity. In: Bland, Janice (ed.) (2018), *Teaching with Challenging*

Texts. English-Language Education with Children and Teenagers. London: Bloomsbury, 105–119.

Kolb, Annika/ Brunsmeier, Sonja (2019), Extensive reading in primary EFL – can story apps do the trick? In: Rokita-Jaśkow, Joanna & Ellis, Melanie (eds.) (2019), *Early Instructed Second Language Acquisition: Pathways to Competence*. Bristol: Multilingual Matters, 153–169.

Korat, Ofra/ Shamir, Adena (2007), Electronic books versus adults readers: Effects on children's emergent literacy as a function of social class. *Journal of Computer Assisted Learning* 23.3, 248–259.

Korat, Ofra/ Shamir, Adena/ Segal-Drori, Ora (2015), E-books as a support for young children's language and literacy: The case of hebrew-speaking children. *Early Child Development and Care* 184.7, 998–1016.

Krashen, Stephen (2007), Extensive reading in English as a foreign language by adolescents and young adults: A meta-analysis. *International Journal of Foreign Language Teaching* 3.2, 23–29.

Krashen, Stephen (2013), Free Reading: Still a Great Idea. In: Bland, Janice & Lütge, Christiane (eds.) (2013), *Children's Literature in Language Education*. London: Bloomsbury, 15–24.

Manresa, Mireia (2015), Traditional Readers and Electronic Literature. An Exploration of Perceptions and Readings of Digital Works. In: Manresa, Mireia & Real, Neus (eds.) (2015), 105–120.

Manresa, Mireia & Real, Neus (eds.) (2015), *Digital Literature for Children: Texts, Readers and Educational Practices*. Brüssel: Peter Lang.

Mayring, Philipp (2014), *Qualitative content analysis: Theoretical foundation, basic procedures and software solution*. [Online: http://nbn-resolving.de/urn:nbn:de:0168-ssoar-395173. 16.08.2021].

Miller, Elisabeth & Warschauer, Mark (2014), Young children and e-reading: Research to date and questions for the future. *Learning, Media and Technology* 39.3, 283–305.

Mourão, Sandie (2015), The potential of picturebooks with young learners. In: Bland, Janice (ed.) (2015), *Teaching English to Young Learners*. London: Bloomsbury, 199–218.

Mourão, Sandie (2016), Picturebooks in the primary EFL classroom: Authentic literature for an authentic response. *CLELE Journal* 4.1, 25–43.

Nuttal, Christine (2005), *Teaching Reading Skills in a Foreign Language*. Oxford: Macmillan.

Paris, Scott/ Wasik, Barbara/ Turner, Julianne (1996), The Development of Strategic Readers. In: Barr, Rebecca; Pearson, David; Kamil, Michael & Mosenthal, Peter (eds.) (2015), *Handbook of Reading Research*. New York: Longman, 609–640.

Reckermann, Julia (2016), I Can Read an English Book. 'The Queen's Knickers' Eigenständig Lesen. *Grundschulmagazin Englisch* 2, 24–26.

Reckermann, Julia (2018), *Reading Authentic English Picture Books in the Primary School EFL Classroom. A Study of Reading Comprehension, Reading Strategies and FL Development*. Frankfurt a. M.: Peter Lang.

Ritter, Michael (2013), Innovative Grenzgänge oder oberflächliche Effekthascherei? Tendenzen der Transformation literarischer Welten in Kinderbuch-Apps. [Online: https://docplayer.org/19660783-Innovative-grenzgaenge-oder-oberflaechliche-effekthascherei-tendenzen-der-transformation-literarischer-welten-in-kinderbuch-apps.html. 16.08.2021].

Rymarczyk, Jutta (2008), Früher oder Später? Zur Einführung des Schriftbildes in der Grundschule. In: Böttger, Heiner (ed.) (2008), *Fortschritte im Frühen Fremdsprachenlernen. Ausgewählte Tagungsbeiträge Nürnberg 2007*. München: Domino Verlag, 170–182.

Sargeant, Betty (2015), What is an Ebook? What is a Book App? And why should we care? An Analysis of Contemporary Digital Picture Books. *Children's Literature in Education*. [Online: http://link.springer.com/10.1007/s10583-015-9243-5. 16.08.2021].

Schocker, Marita (2016), Auf die Lerner kommt es an! Den Englischunterricht von den Lernenden aus denken. *Der fremdsprachliche Unterricht Englisch* 50.143, 2–7.

Sipe, Lawrence (2008), *Storytime. Young Children's Literary Understanding in the Classroom*. New York: Teachers College Press.

Smeets, Daisy & Bus, Adriana (2013), Picture Storybooks Go Digital: Pros and Cons. In: Neuman, Susan & Gambrell, Linda (eds.) (2013), *Quality Reading Instruction in the Age of Common Core Standards*. Newark, Del.: International Reading Association, 176–189.

Smeets, Daisy & Bus, Adriana (2014), The interactive animated e-book as a word learning device for kindergartners. *Applied Psycholinguistics* 36.4, 1–22.

Stichnothe, Hadassah (2014), Engineering stories? A narratological approach to children's book apps. *Nordic Journal of ChildLit Aesthetics*

5.1. [Online: https://www.tandfonline.com/doi/abs/10.3402/blft.v5.23602. 16.08.2021].

Sweller, John (2005), Implications of cognitive load theory for multimedia learning. In: Mayer, Richard (ed.) (2005), *The Cambridge Handbook of Multimedia Learning*. Cambridge: Cambridge University Press, 19–30.

Takacs, Zsofia/ Swart, Elise/ Bus, Adriana (2015), Benefits and pitfalls of multimedia and interactive features in technology-enhanced storybooks: A meta-analysis. *Review of Educational Research* 85.4, 698–739.

Turrión, Celia (2015), Electronic Literature for Children. Characterising Narrative Apps (2010–2014). In: Manresa, Mireia & Real, Neus (eds.) (2015), 87–102.

Verhallen, Marian/ Bus, Adriana/ de Jong, Maria (2006), The promise of multimedia stories for Kindergarten children at risk. *Journal of Educational Psychology* 98.2, 410–419.

Verhallen, Marian/ Bus, Adriana (2010), Low-income immigrant pupils learning vocabulary through digital picture storybooks. *Journal of Educational Psychology* 102.1, 54–61.

Yokota, Junko (2015), The Past, Present and Future of Digital Picturebooks for Children. In: Manresa, Mireia & Real, Neus (eds.) (2015), 73–86.

Story Apps cited in the text:

Crab Hill Press/ Doyle, Bill & Cummings, Troy (2010), *Nash Smasher*.

Heartdrive Media LLC (2016), *Pete's Robot*.

Nosy Crow/ Bryan, Ed (2015a), *The Three Little Pigs*.

Nosy Crow/ Bryan, Ed (2015b), *Jack and the Beanstalk*.

Nosy Crow/ Bryan, Ed (2015c), *Snow White*.

Square Igloo/ Tousnakhoff, Nathalie & Roussel, Matt (2013), *Zoe's Green Planet*.

Three Thumbs Up/ Thomas, Alex/ Jorgensen, Arnie & Voigt, M. (2011), *DinoBoy Adventures*.

Julia Reckermann and Karoline Wirbatz

EFL reading in CLIL and non-CLIL primary schools: A comparison of classroom reading activities, learners' preferences and actual reading comprehension competences

Abstract This chapter is based on a small-scale empirical study that examines English as a Foreign Language (EFL) reading comprehension competences in Content and Language Integrated Learning (CLIL) and non-CLIL primary schools. The focus of the study is four-fold. First, the study explores EFL reading preferences of CLIL and non-CLIL learners (total N = 75) and, second, provides an insight into the learners' EFL reading comprehension competences. Third, the study investigates a possible correlation between the learners' overall opinions towards EFL reading and their reading comprehension competence. Fourth, reading materials and activities that were used by the learners' teachers (N = 2 CLIL teachers and N = 2 non-CLIL teachers) are investigated. Core results indicate that more EFL reading takes place in the CLIL context, particularly concerning reading at text level. The results further show that the majority of the learners seem to like EFL reading, with only slight differences between the CLIL and the non-CLIL group. Yet, there is a considerable difference regarding EFL reading comprehension competences of the two groups, with the CLIL group outperforming the non-CLIL group significantly. Lastly, independent of the CLIL and non-CLIL context, there appears to be a positive correlation between the learners' opinions towards EFL reading and their actual reading comprehension competences.

1 Introduction

The role of reading in the English as a Foreign Language (EFL) classroom has changed significantly since the introduction of English at German primary schools in the 1960s. Changing its status from neglect to acceptance to a claim for more challenging tasks, reading nowadays is an accepted and fixed part of early EFL teaching and learning. As a result, EFL reading at primary school level has also moved more and more into the focus of

EFL research. Various rather small-scale studies (see, e.g., Frisch 2013; Jöckel 2016; Kolb 2013) have investigated different aspects of EFL reading in the primary school English classroom; i.e., reading instruction, reading strategies, reading comprehension and reading materials. However, only very few large-scale studies, e.g., the BIG study (BIG-Kreis 2015), Wilden/ Porsch & Ritter (2013) and the EVENING study (Engel/ Groot-Wilken & Thürmann 2009), were directly concerned with EFL reading competences so far. EFL reading in CLIL contexts has only scarcely been investigated (see studies by Reckermann 2018; Steinlen 2018, 2021; Zaunbauer/ Gebauer & Möller 2012).

Another aspect that has hardly been researched is young EFL learners' motivation regarding foreign language (FL) reading. Reading motivation is a complex phenomenon and can hardly be investigated with learners that only have a very limited command of the FL and possess only very basic reading competences. A large number of studies in other contexts, however, has shown that motivation plays a key role in learning and has documented the connection between reading motivation and reading competence (Henseler & Surkamp 2007: 2; Leubner/ Saupe & Richter 2016: 38–39; Marinak/ Gambrell & Mazzoni 2013: 121). Since investigating young EFL learners' motivation to read is difficult, if not impossible, this article focusses on the learners' opinions towards EFL reading and their EFL reading preferences, as part of their reading motivation, based on reading activities that are in accordance with current curricula.

The focus of the present study is therefore four-fold: First, to get an insight into classroom reading activities and to find out to what extent the materials and activities used by teachers in CLIL and non-CLIL programmes might differ from each other. Second, to explore young EFL learners' opinions towards EFL reading and their preferences regarding reading materials and activities, to draw conclusions about (fostering) their reading motivation, and find out about a possible difference among learners' preferences in CLIL and non-CLIL programmes. Third, to investigate young learners' reading comprehension competences in CLIL and non-CLIL programmes. Fourth, the study investigates a possible connection between the learners' overall opinions towards EFL reading and their reading comprehension competences.

With these foci, the study aims to provide a comprehensive picture of different, yet partly interdependent, aspects that play a role in EFL reading comprehension competences and which have not yet been investigated in CLIL contexts. Since previous research already indicates that CLIL learners have better EFL reading competences than non-CLIL learners (Zaunbauer et al. 2012; Steinlen 2018, 2021), this study also aims to uncover factors that might be central to this finding. There is reason to assume that particularly the aspects of reading motivation and reading activities as well as materials used in class positively influence EFL reading competences. The study aims to investigate whether this can – at least partially – explain the CLIL learners' advance in EFL reading comprehension competences. Certainly, however, more factors than the ones investigated in this study might also play a role (see, e.g., Steinlen [2018] for the role of learners' language background and [2021] for learners' social and cognitive background, or Wirbatz & Reckermann [2020] for the role of the learners' gender).

Section 2 of this chapter starts with an overview of the development of EFL reading in German primary schools. It takes into consideration the historical as well as the current debate evolving around that topic, as well as research findings, classroom practices and curricular demands. A focus will then be on the role of learners' reading motivation. In Section 2, research desiderata will be shown, on the basis of which the empirical study will then be detailed in Section 3. The design of the study will first be explained, followed by a presentation and discussion of the results. This discussion also entails implications for future teaching practices and research projects. Finally, a conclusion will be drawn.

2 EFL reading in German primary schools

2.1 *Historical, empirical and curricular developments*

The integration of written skills (i.e. reading and writing) into primary school EFL curricula has been a matter of debate since the first pilot projects of EFL programmes started in Germany in the 1970s and 1980s. Before the 1970s, the focus was solely on oral skills, i.e. speaking and listening. Written skills were for the most part neglected or completely ignored (see

Doyé 1993: 67; Mindt 2007: 12), with only very few researchers, among them Fay and Hellwig (1971), reporting about a careful integration of reading in primary school EFL classes.

With the beginning of the new millennium and with the introduction of foreign language teaching in all German primary schools[1] in 2006 (Schmid-Schönbein 2008: 18 f.), a debate which became known as the debate on early biliteracy prevailed in the discussion on EFL reading and writing with young learners (see Piske 2010). Arguments against early biliteracy included that learners might be overtaxed by reading and writing in English (Legutke/ Müller-Hartmann & Schocker-v. Ditfurth 2009: 57), that English literacy would negatively influence learners' first language (L1) development (Kierepka 2010: 93–94), that there is not enough time for literacy (Kierepka 2010: 91; Legutke et al. 2009: 57) and that too much emphasis might be put on written skills (Bleyhl 2007: 47; Legutke et al. 2009: 57). Arguments for early biliteracy, however, were stronger than the ones against it. They were mostly concerned with the learners' desire and interest to read and write in English (Diehr 2010: 53; Zaunbauer 2007: 46), the advantages of a simultaneous introduction of spoken and written language (Zaunbauer 2007: 46), the need to cater to different learning styles (Brown 2007: 129), the learners' natural encounter with written English in their everyday lives (Zaunbauer 2007: 46), the wish and need to prevent invented spelling (Diehr & Rymarczyk 2008: 9), and the supportive function of literacy for language awareness (Diehr 2010: 54) as well as for learner autonomy (Freudenau 2012). In addition, research indicated that reading and writing in English had no negative influence on the learners' L1 (literacy) development (Diehr & Rymarczyk 2008; Piske 2010; Rymarczyk & Musall 2010).

Researchers and teachers soon realized that EFL reading and writing should be part of teaching English to young learners. The results of large-scale studies, particularly the EVENING study (see Engel et al. 2009), also supported early biliteracy. The question was no longer 'if' English reading and writing should be part of the primary school EFL classroom, but 'how'

1 The majority of German primary schools offer English as the first foreign language, while a minority of primary schools, particularly those in the border region to France, start out with French.

EFL reading and writing could and should be taught. Research findings of the EVENING study (see Paulick & Groot-Wilken 2009) and the BIG study (see BIG-Kreis 2015) indicated that young EFL learners could easily cope with reading at word and sentence level. The studies also indicated that the reading test of the EVENING study, which was used in both studies (EVENING and BIG) to test reading at word and sentence level, was too easy for the learners since the results of both studies displayed ceiling effects (BIG-Kreis 2015: 68; Paulick & Groot-Wilken 2009: 189). Consequently, a demand for reading activities at text level rose. However, as outlined below, EFL reading at text level hardly takes place in German primary schools. First studies and practical projects, though, show very promising results (see, e.g., Frisch 2013; Kolb 2013; Reckermann 2018).

Especially CLIL schools might make a difference in that more EFL reading at text level is likely to take place, as the skill of foreign language reading is indispensable in (primary school) CLIL contexts (Bosenius 2009: 18; Burmeister & Piske 2008: 187; Steinlen 2021). Yet, no study looked at the scope of EFL reading activities in German CLIL primary schools to the present. Still, as indicated in the introduction, several empirical studies investigated EFL reading competences in the primary school CLIL context. In a study that examined reading fluency, reading comprehension and EFL vocabulary of primary school CLIL and non-CLIL learners, Zaunbauer et al. (2012) could show that the former group outperformed the latter in all three areas. With regard to EFL reading comprehension of CLIL and non-CLIL learners, Steinlen (2018, 2021) found similar results[2]. This advance of the CLIL learners is not necessarily surprising, because their contact time to the foreign language is a lot higher than that of non-CLIL learners. Looking only at Year 4 CLIL learners ($N = 11$), Reckermann (2018) found that they could read and understand, to varying extents, authentic English picture books and use a variety of different reading strategies to do so. Steinlen reports that primary school students in high intensity bilingual

2 It should be noted that the focus of Steinlen's study (2018, 2021) was not on EFL reading comprehension of CLIL and non-CLIL learners per se. The project examined the influence of gender, language background and social status on the EFL performance of German primary school learners attending different EFL programmes.

programs (50 % and 70 % of the teaching time is conducted in English) reached level A2/B1 for reading in English while level B1 is only expected of mainstream students at the end of Year 10 (2021: 125).

Together with the development of a gradual integration of reading skills in primary school EFL classrooms in Germany, curricular demands changed as well. Taking the curriculum of North-Rhine Westphalia (NRW), the federal state where the study at hand was conducted, as an example, the first curriculum for English in Years 3 and 4 from 2003 included rather easy reading activities mostly at the word and sentence level that should be based on familiar vocabulary (MSJK NRW 2003: 37). Based on research findings, particularly those of the EVENING study (see Engel 2009: 202), the curriculum was revised in 2008 and now sets rather ambitious aims for EFL reading outcomes. Reading (and writing) are ascribed a supportive function from the beginning of teaching English in Year 1 (MSW NRW 2008: 73) and reading activities should now take place at word, sentence as well as text level (ibid.: 78). Additionally, concerning reading materials, not only very simple texts but also authentic materials, such as stories, should now be used in the EFL classroom (ibid.: 74). Particularly in comparison to curricula of other federal states, these reading expectations placed on young EFL learners are quite high. In Saxony, for instance, young EFL learners only need to be able to recognise written words and chunks which they already know (SSK 2019: 6).

It does not come as a surprise that the demands outlined in the NRW curriculum are rarely put into practice: reading activities in German primary schools are often limited to reading single words or short sentences, but hardly include the use of stories, books or coherent texts (see BIG-Kreis 2015: 38; Kolb 2013: 33).

2.2 *The role of reading motivation and learners' preferences in reading comprehension processes*

Reading competence comprises of a number of different aspects, among them reading comprehension, reading aloud abilities (see, e.g., Diehr 2010; Diehr & Frisch 2010b), decoding competences in reading processes (see Frisch/ Breul/ Diehr/ Kastens & Becker in this volume), reading rate and reading fluency (see, e.g., Dehaene 2010), reading motivation (see

Henseler & Surkamp 2007) reading strategies (see Kolb 2013; Reckermann 2018) and literary learning through reading (see Alter & Frisch 2018). These are not independent of each other, but often in an interplay. The present study investigated reading comprehension as part of reading competence as well as the learners' preferences regarding reading materials and activities as part of their reading motivation.

In teaching contexts, reading is understood as a receptive skill (Waas 2014: 35), yet, it is by no means passive. Readers are actively involved in a number of different reading processes in order to understand and make sense of a text. This already hints at the (arguably) main outcome of reading: reading comprehension. Reading comprehension means that a reader is actively involved in processes that allow him or her to decode the meaning of a written piece of work (Steck 2009: 22). For more information on foreign language reading comprehension processes including a model of reading see Frisch et al. (in this volume; adapted from Diehr & Frisch 2010a: 27).

Reading motivation is tightly interwoven with reading comprehension, in that there is a strong link between reading motivation and reading achievement (Henseler & Surkamp 2007: 2; Leubner/ Saupe & Richter 2016: 38–39; Marinak et al. 2013: 121). Marinak et al. (2013: 121) outline "creating interest in reading" as one of the most important aspects of teaching reading in school. Learners should acquire the motivation to develop into competent, active and engaged readers. While Leubner et al. (2016: 38) use reading interest or reading enjoyment synonymously to reading motivation, Richter and Plath (2012: 21) distinguish between interest and motivation. According to them, interest is strongly connected to a specific reading object, e.g., a specific book, while motivation is rather a longer-lasting and overriding disposition that serves as a motive for purposeful reading (ibid.).

Motivation to read is a complex and multifaceted construct and consists of a number of different factors that influence the level of motivation to a greater or lesser extent. These factors include: individual reading self-concept (Marinak et al. 2013: 122), the role and value that an individual places on reading (Marinak et al. 2013: 122; Richter & Plath 2012: 42), interest in reading (Richter & Plath 2012: 22), time spent on reading

(ibid.: 42), enjoyment of reading (ibid.), attitude towards reading (ibid.), reading habits (ibid.), desire to communicate about reading (ibid.: 45) and expectations of success (Chiu 2018: 48). This list is by no means complete and does not include the way that some variables interact with each other and have a stronger or lesser influence on reading motivation.

Investigating the complex construct of FL reading motivation in empirical studies has hardly taken place for primary school EFL learners yet. Only a small number of empirical studies could show that for the most part primary school EFL learners are interested in and enjoy the reading of English texts in different contexts (Diehr & Frisch 2010b; Frisch 2013; Kolb 2013; Reckermann 2018). Only Frisch (2013: 108–186), whose study was concerned with EFL reading of phonic readers at text level, investigated the learners' reading motivation consisting of interest in particular reading material, self-assessment of reading competences, opinions towards reading English texts and sociocultural reading context in a pre- and post-design. The reason why empirical research hardly looked at young learners EFL reading motivation so far is probably that it is difficult to investigate at such an early stage of EFL learning. Richter and Plath (2012: 35–36) explain that reading motivation develops alongside with reading competence, and that only more or less competent readers possess reading motivation that is reliably measurable. Therefore, reading motivation can be supported at the same time as reading competence is gradually built up, but reading motivation can only be measured once a certain degree of competence has been reached and reading of texts takes place. This becomes even clearer when looking at common instruments that measure reading motivation, as for instance the Motivation to Read Profile (Marinak et al. 2013) or Richter and Plath's (2012) primary school learner questionnaire. Both consist of seemingly easy questions about reading, but these can only be answered when one actually spends time on reading, reads at text level and communicates about reading with others. All three aspects only take place to a very limited extent in German primary schools regarding reading in English.

Still, being aware that reading motivation plays a key role in the development of reading competences (Henseler & Surkamp 2007: 2; Leubner et al. 2016: 38), this article seeks to explore an aspect of reading motivation

that can realistically be measured with primary school EFL learners, and that is what the authors of this chapter refer to as the learners' EFL reading preferences. Preferences, in this case, are defined as what learners like or do not like concerning reading activities and reading materials in the EFL classroom. The learners are asked about their preferences in accordance with what they have likely experienced in class, taken into consideration the curriculum: reading aloud and reading silently, reading at word, sentence and text level, reading factual or fictional texts, and reading on the Internet. Additionally, one questionnaire item asks the learners about their overall opinion towards EFL reading. Their preferences and overall opinions certainly cannot cover the complex construct of reading motivation, but given the list of factors that are included in reading motivation, these two factors can be seen as being a small part of or at least as playing a role in reading motivation.

When investigating reading preferences, the question arises whether teachers can teach reading according to their learners' preferences or whether it is the curriculum that predisposes the way that reading is taught. Richter and Plath (2012: 22) argue in this respect that only because learners' interests and competences are taken into consideration when planning reading activities, e.g. the reading of a certain text, the school's educational mandate is not neglected. The authors postulate that a competent way of teaching literature includes the selection of a reading text as well as the consideration of the situation that the learners are in (ibid.). Therefore, this article does not suggest that learners' preferences should be the hidden curriculum for teaching reading, but since the learners' motivation, to a certain extent coined by their preferences, plays a key role in the development of reading competences, their preferences should be taken into consideration for the planning of reading activities in class.

2.3 *Summary, outlook and current desiderata*

To the present, a number of studies have been conducted that are concerned with young learners' EFL reading from different perspectives and in different contexts. Exploring classroom practices, the BIG study (2015), among others, could show that classroom practices do not always mirror curricular demands, particularly concerning reading at text level.

Therefore, it is worth taking a fresh look at classroom practices and see whether this gap still exists or whether time, experience and research results might have had an influence in that more teachers now deal with, e.g., reading activities at text level. Moreover, different contexts, particularly CLIL vs. non-CLIL contexts, probably make a difference regarding the development of reading competence. Possibly, practices that work in CLIL contexts could to a certain extent be adopted for non-CLIL contexts.

Reading motivation has hardly been researched in primary school EFL contexts so far. Still, Kolb (2013) as well as Diehr and Frisch (2010b) report that learners like the reading of picture books in general and are interested in reading English picture books. Frisch (2013: 180–186) found that the young learners of her study liked reading English texts, were interested in them, and that selected aspects of reading motivation improved over time through the increase of English reading competence. These studies can be seen as an indicator that learners show interest in more challenging reading tasks. The present study therefore seeks to explore the learners' preferences regarding EFL reading in more detail. Being aware that preferences, as an aspect of reading motivation, play a key role in developing substantial reading competences, studies which explore the young learners' preferences regarding EFL reading are desirable. Although their preferences cannot, by themselves, serve as a basis for teaching, they might shed light on current and future practices, in particular concerning the use of different texts and reading activities that go beyond word and sentence level.

Regarding young learners' EFL reading competences, a number of studies have been conducted up to the present in German primary schools. However, the two large-scale quantitative studies (BIG and EVENING study) showed ceiling effects and only tested EFL reading at word and sentence level. The studies concerned with reading at text level (Frisch 2013; Kolb 2013; Reckermann 2018) were small-scale, yet promising. This means that studies are needed which use an instrument that tests reading competences not only at the word and sentence, but also at the text level. Such an instrument should neither show ceiling effects, as the instrument used in the EVENING study repeatedly did, nor floor effects in that a more difficult test should still be manageable for EFL learners at the primary school level.

Regarding a comparison of CLIL and non-CLIL learners, Steinlen (2018, 2021) as well as Zaunbauer et al. (2012) could show that the CLIL learners' EFL reading competences exceed those of the non-CLIL learners. While this is hardly surprising, the question remains whether it is only CLIL that made the difference or whether other factors might play a role as well. Steinlen (2018, 2021) explored a possible influence of language background and social status, the role of gender has also been investigated (Steinlen 2018, 2021; Wirbatz & Reckermann 2020), but neither learners' preferences nor classroom practices regarding reading have been researched; yet, they might play a role.

3 The study

Based on previous research findings, current classroom practices, curricular demands and the research desiderata outlined above, the study at hand takes a closer look at the reading activities that take place in the primary school CLIL and non-CLIL EFL classroom. It also investigates the preferences of third graders in CLIL and non-CLIL programmes regarding those activities, as well as their actual reading comprehension competences. Lastly, a possible connection between reading comprehension competences and the learners' preferences is explored.

3.1 *Research design*

Research questions

This study aims to answer the following research questions:

1. How often do Year 3 EFL teachers use different reading activities and materials in CLIL and non-CLIL programmes?
2. What are young EFL learners' preferences in CLIL and non-CLIL programmes regarding reading activities and materials in the classroom?
3. What are young EFL learners' reading comprehension competences in CLIL and non-CLIL programmes?
4. Is there a connection between the learners' overall opinions towards EFL reading and their reading comprehension competences?

Setting and participants

The study was conducted at four primary schools located in the same region of NRW. Two of these schools have a CLIL programme in which around 10 lessons per week were taught in English or in English/German via team-teaching. The other two schools were regular schools with two lessons of English per week. All 75 participants (N = 42 CLIL learners and N = 33 non-CLIL learners) were attending the third grade at the time of data collection. Important background variables are shown in Table 1.

Tab. 1: Background information of the participant learners

Variable	**CLIL group (N = 42)**	**Non-CLIL group (N = 33)**	**Results of statistical analysis** (Are the differences between the CLIL and the non-CLIL group significant?)
Gender	Boys: N = 22 Girls: N = 20	Boys: N = 14 Girls: N = 19	X^2 (1, N = 75) = 0.73, p = 0.392
Age	M = 9.0 years SD = 3.4 months	M = 9.3 years SD = 6.64 months	t(72) = 3.14 p = 0.002**[a]
Migration background	Yes: N = 21 No: N = 21	Yes: N = 25 No: N = 8	X^2 (1, N = 75) = 5.17, p = 0.023*
Language background	Monolingual: N = 25 Bi-/multilingual: N = 17	Monolingual: N = 11 Bi-/multilingual: N = 21	X^2 (1, N = 75) = 4.6, p = 0.032*
Socio-economic status (SES) (on a scale from 1 to 9; 1 being the lowest and 9 the highest possible SES)	M = 5.74 SD = 1.98	M = 4.44 SD = 1.84	t(53.58) = -2.71 p = 0.009** (no variance homogeneity according to Levene-test)

a low significance* p<0.05; significant ** p<0.01; high significance*** p<0.001

One can see that there are slight differences between the two groups regarding all variables, mostly in favour of the CLIL group. A statistical analysis via Pearson's Chi square and t-tests revealed that these differences are statistically significant for all variables except for gender. This means that the CLIL group seems to be at an advantage. This advantage might have an influence on the CLIL group's reading comprehension results, which will be considered in the analysis and interpretation of the respective data. An influence on reading preferences is unlikely[3].

The learners' English teachers, which at the CLIL schools were also their CLIL teachers, also participated in the study (N = 2 CLIL teachers and N = 2 non-CLIL teachers)[4]. The two CLIL teachers both held a fully-fledged teacher degree in English and had 9 and 10 years of experience in teaching this subject. The non-CLIL teachers had 11 and 15 years of EFL teaching experience, and had gained their teaching qualification through a further training course (methodological qualification as well as a language test certifying C1 level).

Research instruments

The data for this study were collected through a teacher questionnaire, a learner questionnaire and a test which tested the learners' EFL reading comprehension competences.

In the teacher questionnaire, the teachers provided bio data as well as information about their EFL qualifications and teaching experience. They were also asked about their EFL classroom practices in regard to teaching

3 An influence of the variable SES on the complex construct of reading motivation cannot be excluded, but reading motivation as such was not investigated. An influence on reading preferences as investigated and defined for the purpose of this study seems unlikely.

4 The authors are aware of the fact that N = 4 is a very small sample size which is associated with various issues, such as low or no statistical power and no basis for generalisations. However, only four teachers could be questioned since conclusions were to be drawn in connection to the learners that took part in the study. The four teachers were the only ones who taught English in this group of subjects. Thus, particularly regarding the role of the teacher, this study can only carefully draw very context-bound conclusions and thereby possibly generate an hypothesis for further investigations. Larger confirmatory studies are still necessary to test the proposed hypothesis.

reading. That is, they were given a list of different reading activities and had to specify which of them they implemented 'often', 'sometimes' or '(almost) never' during their lessons[5]. The teacher questionnaire was designed in German and translated for the purpose of this chapter by the authors.

A learner questionnaire that the learners filled in during class while their teachers were present provided background information about the learners as well as their preferences and opinions regarding EFL reading activities. They were asked to rate different EFL reading activities by either ticking 'good', 'medium' or 'not so good'. The learner questionnaire was also designed in German and translated for this paper by the authors.

Additionally, the reading and writing part of the 'Cambridge young learners English movers' test' (University of Cambridge 2014) (YLE-test) was used to determine the learners' EFL reading comprehension competences at word, sentence and text level. For reasons of feasibility and practicality, only the reading and writing part of a sample paper of this test was used[6]. Since Reckermann (2018: 222–223) had already found that young EFL learners had difficulties to cope with English task instructions in this test, a German translation of the instructions was added. The scoring of the test was based on the official scoring rubric of the sample YLE-test (max. score 40 points), with one slight modification regarding the spelling of words: In accordance with the curriculum for EFL in primary schools in NRW (MSW NRW 2008: 82), incorrectly spelled words that were still recognisable were taken as correct. This was also done to not test the learners' writing and spelling competences, but to focus on the content of their answers and thus their reading abilities.

The test consisted of six different tasks that can be classified as shown in Table 2.

5 The difference between the three rating options is rather subjective, but still gives an impression of how often reading is dealt with in comparison to other activities. The rating is thus based on the teacher's individual classroom situation and not pre-defined by numbers.

6 This test mainly tests reading and not writing competences. In this test, writing does not go beyond the mere copying of words, which hardly provides a comprehensive picture of EFL writing competences. Therefore, the test was used only to measure reading comprehension competences, without excluding that writing was involved to a certain extent.

Tab. 2: Analysis of the tasks of the YLE-test used to examine the EFL reading comprehension competences in this study

Task (max. points)	Task description	Reading comprehension at…			Factual character	Fictional/story character	Dominant topic(s)/word field(s)
		Word level	Sentence level	Text level			
1 (6 p.)	A description of an item in one or two short sentences has to be understood. The correct item must be identified and copied from a selection of labelled pictures.	x*	X		X		Animals, body parts, food and drinks
2 (6 p.)	Six statements (one short sentence each) that describe a picture must be marked as right (yes) or wrong (no).		X			X	Colours, furniture, numbers
3 (6 p.)	In a conversation between two children, the correct short answer (one out of three) has to be identified for six different short questions.		X			X	Well-being, going home, clothes, drinks
4 (7 p.)	Six gaps in a short story of eleven, mostly short, sentences have to be filled with words from a selection of nine (nouns and verbs) that are visualised. Then, a suitable heading for the story must be chosen from three different suggestions.	x		X		X	Toys and books, free time activities, adventures
5 (10 p.)	A short story of three paragraphs with a picture per paragraph is given. After each paragraph, three (once four) sentences about the story must be supplemented with the correct word or words as found in the story. The gap in each sentence must be filled with one to three words.	x		X		X	Holidays, family, free time activities, weather, animals, days of the week, food
6 (5 p.)	A factual text about cats with ten short sentences is given. The text has five gaps that have to be filled correctly. For each gap, three different options (words) are provided.	x	X		X		Prepositions, verbs, adjectives; cats

*The small 'x' indicates that word-level reading was implicitly included, but not the focus of the task.

As can be seen, reading comprehension is tested at the text level in two tasks, which equals 33.3 %, and in four tasks at the sentence level, which equals 66.6 %. Reading at word level is implicitly part of four of the six tasks, although not their main focus. The word fields and topics covered in the tasks are in accordance with those found in the curriculum for primary school English in NRW (ibid.: 76).

As the test had already been used successfully with young learners at a German school with a bilingual programme at the beginning of a fourth grade (see Reckermann 2018), further piloting was deemed unnecessary. However, as will be seen later in this chapter, a piloting phase might have already indicated that the test was slightly too difficult for the learners of this study.

3.2 *Presentation and discussion of results*

The presentation and discussion of the results is divided into four subsections that correspond to the four research questions as outlined in Section 3.1.

3.2.1 Reading activities in CLIL and non-CLIL programmes

In the teacher questionnaire, the teachers were asked about the extent to which they use different reading activities and materials. Table 3 shows how often the teachers use reading activities at word, sentence and text level and whether they use silent reading for comprehension or reading aloud activities.

Tab. 3: Reading activities, silently or aloud, at different levels as used by the CLIL and non-CLIL teachers

Level	Activity	Answers					
		Often		Sometimes		(Almost) Never	
		CLIL teachers	Non-CLIL teachers	CLIL teachers	Non-CLIL teachers	CLIL teachers	Non-CLIL teachers
Word level	Learners **silently** read single English words for comprehension.	II	II				
	Learners read **aloud** single English words.	II	II				
Sentence level	Learners **silently** read short English sentences for comprehension.	I		I	II		
	Learners read **aloud** short English sentences.	I		I	II		
Text level	Learners **silently** read whole English texts for comprehension.			II			II
	Learners read **aloud** whole English texts.			I		I	II

Table 3 shows that there is a tendency towards more reading for comprehension in the CLIL classes, as this technique is used more frequently by the CLIL than the non-CLIL teachers. At word level, the activities used by all four teachers are the same and they are used equally often. At sentence and text level, however, the activities "reading silently" and "reading aloud" are used more frequently by the CLIL teachers. Interestingly, reading for comprehension at text level does (almost) never take place in the non-CLIL teachers' English classes, which corresponds with earlier research findings, but not necessarily with NRW's curricular demands (see Section

2.1). By contrast, the CLIL teachers' practices appear to comply with the curriculum, while the non-CLIL teachers should not only use the activities "word level reading" and "sentence level reading" but also reading aloud and silent reading activities at text level (MSW NRW 2008: 78).

Tab. 4: Reading materials and genres used by the CLIL and non-CLIL teachers

Reading materials/ genre	Activity	Answers					
		Often		Sometimes		(Almost) Never	
		CLIL teachers	Non-CLIL teachers	CLIL teachers	Non-CLIL teachers	CLIL teachers	Non-CLIL teachers
(Fictional) Stories	Learners read English stories (with support).			II			II
Factual texts	Learners read English factual texts (with support).			I		I	II
Internet/ computer	Learners search for and read English texts on the Internet/ computer (with support).					II	II

Table 4 shows that regarding reading materials and genres of texts, the CLIL classes sometimes read stories, while the non-CLIL classes (almost) never do so. Furthermore, the non-CLIL classes (almost) never read factual texts, while one CLIL class does so at least sometimes. It is interesting, however, that the other CLIL class (almost) never does so, as one would expect that in CLIL lessons factual texts are read in the foreign language as a means for gaining knowledge. Interestingly, the CLIL and the non-CLIL classes (almost) never read English texts on the computer or the Internet. Although the Internet arguably provides a wide range of English reading materials, it is (almost) never used in the classes that were part of this study. Similar to the results on the use of different reading activities as described above, the reading of different English text types generally happens more often in the CLIL than the non-CLIL classroom.

The findings presented in this section can, of course, not be transferred to other contexts or other classes. Although the findings for the non-CLIL classes are in accordance with what has been found out about reading

activities in classroom practice before, the number of teachers that participated in this study is too small to draw any generalising conclusions. Yet, the data shows that the CLIL teachers, in contrast to the non-CLIL teachers, use reading at text level at least sometimes. This is not surprising given the fact that they have a lot more lessons in English and thus more time to include reading activities. This finding could indicate that, given more time, the CLIL classes' practice could be transferred to regular EFL classes which could embed more reading activities than is currently done. The fact that they scarcely use reading activities beyond word and sentence level might be based on the time-issue elaborated on by Kierepka (2010: 93–94).

3.2.2 CLIL and non-CLIL learners' reading preferences

In the learner questionnaire, the learners were asked to rate reading in English in general as well as to rate different reading activities and materials/genres that are dealt with in the EFL classroom. Since the teachers claim to use some of the activities/materials in question (almost) never, we must assume that the children only rated those activities/materials that were actually used in class, even if only once, or did not rate the respective item at all. Results are given in percentages rather than absolute numbers, to allow an easier comparison between the CLIL and the non-CLIL group.

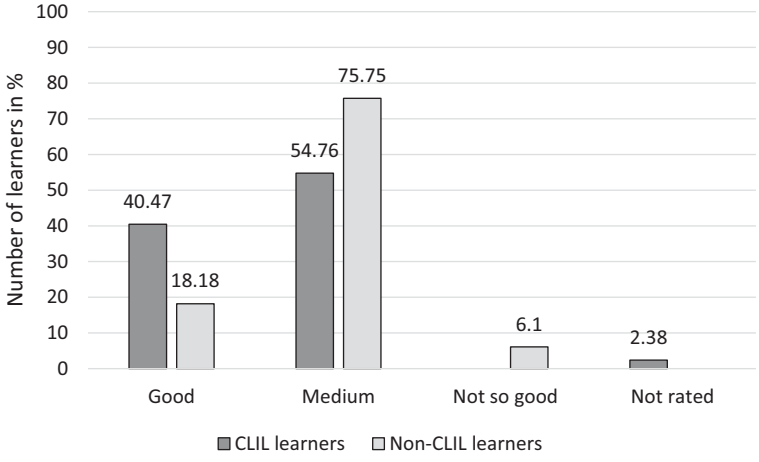


Fig. 1: *CLIL and non-CLIL learners' general opinions of reading in English (in %)*

Figure 1 shows how the learners rate reading in English in general.

As can be seen, about 40 % of the CLIL learners rate English reading generally 'good', while the remaining CLIL learners rate it 'medium'. One CLIL learner did not answer the question. In contrast, only about 18 % of the non-CLIL learners rate English reading in general as 'good', while about 75 % rate it 'medium' and about 6 % rate it 'not so good'. These results display a tendency in favour of the CLIL learners, whose opinions about reading in English are more positive. However, the overall result is positive in that only two of the non-CLIL learners (i.e., 6.1 %) state to not like reading in English. This clearly underlines the necessity and rightness of using English reading activities already with young EFL learners in primary school. All learners appear to have a generally positive opinion about EFL reading, regardless which programme they attend. This positive mindset concurs with the results of previous studies about young EFL learners' overall opinions on EFL reading (Diehr & Frisch 2010b; Frisch 2013; Kolb 2013; Reckermann 2018). Yet, in the light of the old debate on early biliteracy (see Section 2.1), it is interesting to know that the learners seem to like and appreciate the use of EFL reading activities in general.

Figure 2 shows the learners' preferences regarding reading silently and reading aloud.

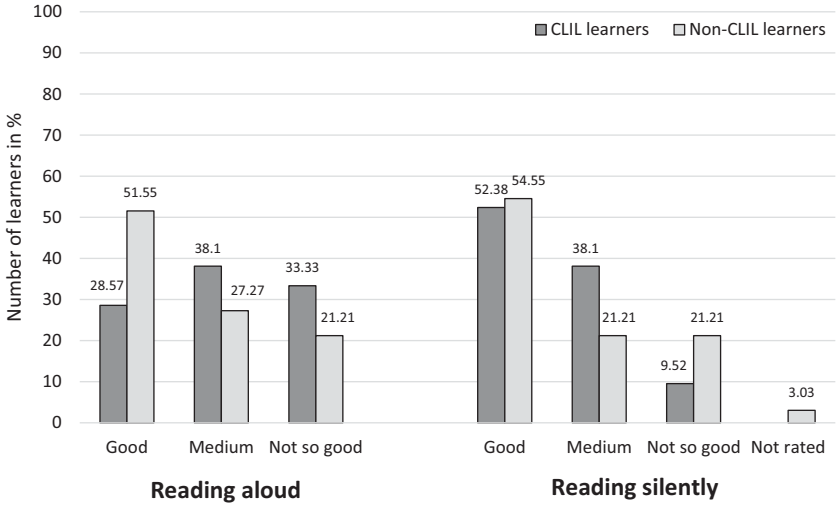


Fig. 2: *CLIL and non-CLIL learners' preferences regarding reading silently and reading aloud in English (in %)*

Looking at the non-CLIL learners, they seem to like and dislike reading aloud and reading silently equally much. About 50 % of them rate reading aloud as well as reading silently 'good', while about 20 %–25 % rate both activities as either 'medium' or 'not so good'. The CLIL learners, in contrast, seem to prefer reading silently to reading aloud. About 52 % of them rate reading silently as 'good', while only about 28 % consider reading aloud an activity they like doing. An equal number of about 38 % rate both, reading silently and reading aloud, 'medium'. Only about 9 % of the CLIL learners do not like reading silently at all, while about 33 % do not like reading aloud. This result shows the importance of including both reading silently and aloud activities in class, in order to cater to different learner preferences. Arguing that reading comprehension, which is usually the result of silent reading, is the desirable reading outcome, Diehr (2010) also makes a case for oral reading. She claims that carefully selected reading aloud activities guide learners in noticing and decoding the grapheme-phoneme correspondence of a word to associate the written word with its correct pronunciation, that reading aloud is the precursor to the important feature of sub-vocalisation in reading comprehension processes, and that reading aloud helps learners to gradually develop a second language inner voice (Diehr 2010: 59–61). Taking these well-grounded arguments into consideration one can easily see that both reading aloud and silent reading activities are needed in the development of foreign language (reading) competences. Why the CLIL learners seem to like reading aloud less than the non-CLIL learners and seem to favour silent reading more, however, remains up to speculation. Possibly, this relates back to the fact that the CLIL learners also read more at text level, which they might feel more comfortable with doing silently and for comprehension rather than aloud and/or for an audience.

The CLIL and non-CLIL learners' preferences regarding reading materials and genres used in the EFL classroom are shown in Figure 3

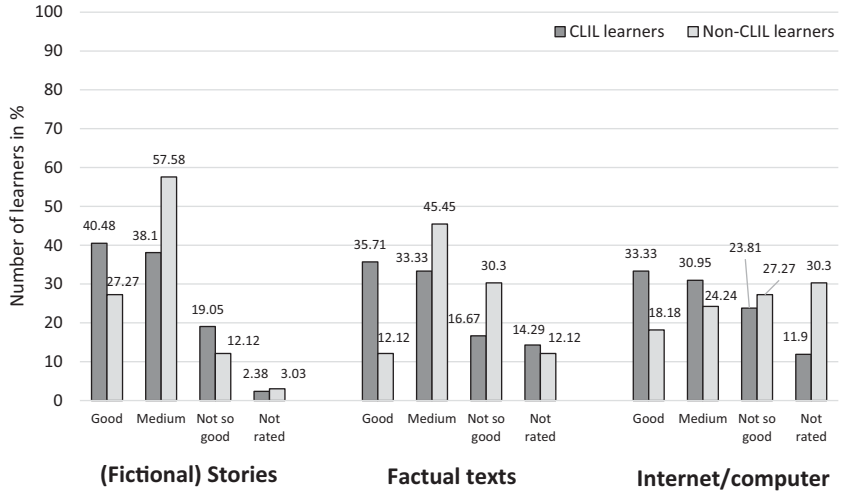


Fig. 3: *CLIL and non-CLIL learners' preferences regarding reading materials and genres in the EFL classroom (in %)*

The CLIL learners' preferences for the three different types of text (i.e., fictional stories, factual texts, and Internet/computer) seem balanced. About 35 %–40 % of them rate reading all three text types 'good', about 30 %–35 % rate all three 'medium' and about 15 %–20 % rate all three as 'not so good'. Overall, a large number of about 80 %, adding up those who rate 'good' and 'medium', thus seem to like reading different texts in English at least to a certain extent. An encouraging low number of only around 20 % of the CLIL learners seem to not like reading different texts in English. The fact that about 15 % do neither rate factual texts nor texts on the Internet mirrors what has been stated by the teachers, namely that these reading activities are not frequently used in class. Overall, the CLIL learners appear to equally like different types of texts and no clear preference towards fictional or factual texts is found.

In the non-CLIL group, the overall picture is also positive. That is, only 12 % of the non-CLIL learners rate the reading of stories 'not so good', while about 30 % rate the reading of factual texts or texts on the Internet negatively. The remaining 45 %–85 % rate the reading of the three different reading materials 'good' or 'medium'. In contrast to the CLIL group,

however, fewer non-CLIL learners rate the reading of all three text types positively. Only about 27 % of this group rate reading English stories as 'good', only a low number of about 12 % rate reading factual texts 'good' and only about 18 % rate reading English texts on the Internet as 'good'. Nevertheless, nearly half of the non-CLIL learners rate reading stories and factual texts at least 'medium'. The CLIL learners might like factual texts more than the non-CLIL learners due to the fact that they are possibly more used to reading them in their CLIL subjects, which at least to a certain extent make use of such texts (see Section 3.2.1). Similar to the CLIL learners, 30 % of the non-CLIL learners do not like reading on the Internet or the computer. This result appears to mirror classroom practices, where such reading activities (almost) never take place (see Section 3.2.1.). It is noteworthy that only 11 % of the learners in the CLIL group and 30 % in the non-CLIL group did not rate this item at all, since all their teachers claim to (almost) never apply activities like this. Possibly, they might have done it once, or they might have encountered a similar activity in another subject and thus rated the item accordingly. However, caution is needed in the interpretation of this result, since the data might not be fully reliable.

Based on the findings presented in this section, the following suggestions can be made regarding the use of different reading activities and materials in the EFL classroom. First, more reading should take place in the EFL classroom in order to meet the learners' overall positive opinions towards it. Second, both activities, i.e. reading aloud and silent reading, should be used in the regular EFL classroom, since the learners seem to enjoy both to a certain extent and since both have their justification in gradually developing foreign language competences. The CLIL learners' preference in favour of silent reading would be worth a follow-up investigation in order to determine possible reasons. Yet, knowing that reading aloud can also be very beneficial for foreign language competence development (Diehr 2010), this should not be neglected but rather dealt with in a motivating and encouraging way. Third, different kinds of text should be used for EFL reading activities. Since the non-CLIL learners seem to prefer (fictional) stories, they should be used in order to further foster their reading motivation. Factual texts seem more important for the CLIL than the non-CLIL context and the CLIL learners' mostly positive mindsets towards this text type indicate that they can and should continue to be used in this context.

Potentially, if factual texts were used more frequently in non-CLIL EFL classes, non-CLIL learners would get used to them more, develop confidence for reading them and thus enjoy reading them better. Practical suggestions of how this could be done are published by Glaser & Schrader (2019a; 2019b).

3.2.3 CLIL and non-CLIL learners' reading comprehension competences

To answer research question 3, the test scores of the two groups on the YLE-test were analysed and compared to each other. The overall mean score as well as the mean scores for each group are shown in Table 5, with further information provided in the box plot shown in Figure 4.

Tab. 5: *EFL reading comprehension competences (mean scores)*

	N	**Mean score and standard deviation in %**
Total	75	*M* = 40.08 % (*SD* = 21.32 %)
CLIL learners	42	*M* = 49.23 % (*SD* = 20.69 %)
Non-CLIL learners	33	*M* = 28.45 % (*SD* = 15.93 %)

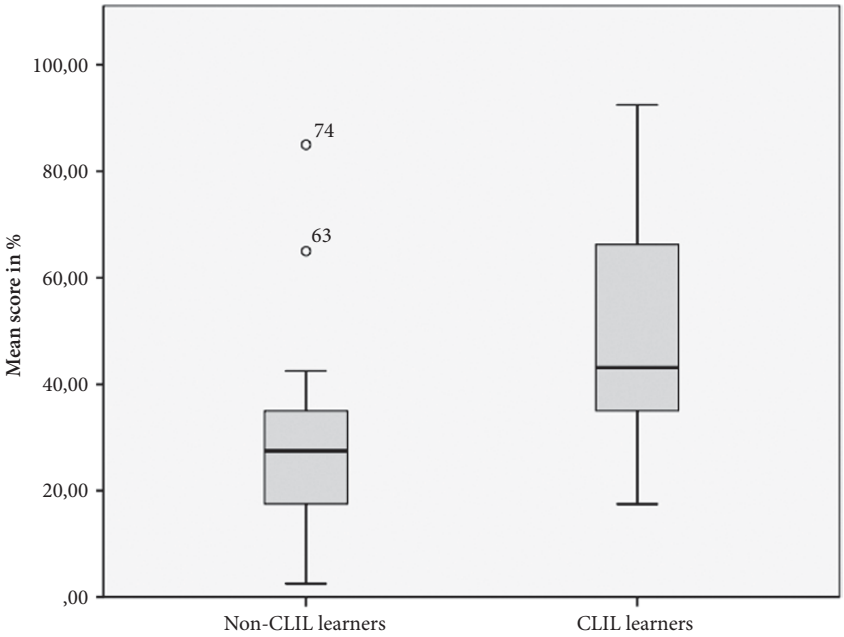


Fig. 4: *CLIL and non-CLIL learners' EFL reading comprehension competences*

As can be seen, the CLIL learners clearly outperform the non-CLIL learners in regard to their EFL reading comprehension competences, while at the same time there is also great variation within the two groups. The variation is larger within the CLIL group, whereas the non-CLIL group contains two outliers, at 65 % and 85 %, which clearly exceed the group's mean score of 28.45 %. A one-way ANOVA shows that the EFL reading comprehension competences of the two groups differ significantly, $F(1, 73) = 22.69$, $p = .000***$. This means that the CLIL learners' EFL reading comprehension is significantly better than that of the non-CLIL learners; a result that concurs with that of previous studies (Steinlen 2018, 2021; Zaunbauer et al. 2012).

An ANCOVA[7] was conducted to control for a possible influence of the groups´ differences in SES and age. The analysis shows that the SES does not have a statistically significant impact on the results, $F(1, 64) = 19.93$, $p = .000***$. The differences in SES between the groups do therefore not influence the overall result, so that the CLIL learners' better English reading comprehension is not based on their higher mean SES. The same holds true for the variable age, for which the ANCOVA could show that controlling for age does not influence the overall result either, $F(1, 72) = 15.46$, $p = .000***$. The other two non-parametric background variables (migration and language background) could not be controlled for via an ANCOVA. However, previous studies have shown that they might have an impact on the learners´ reading comprehension competences (see, e.g., Steinlen 2018, 2021; Wilden & Porsch 2015).

The CLIL group's better result is not surprising, though, in that they enjoyed many more lessons in English than the non-CLIL group and, thus, it is likely that more EFL reading activities took place during these lessons. This statement is backed up by what the teachers revealed about classroom reading activities (see Section 3.2.1). An influence of the variables age and SES, for which Table 1 showed a significant difference between the groups in favour of the CLIL group, could be controlled for and an influence of both on the learners' reading comprehension can be ruled out.

Looking at the model of foreign language reading (Frisch et al. in this volume), the CLIL learners' advance in EFL reading comprehension could be explained by a possible positive influence of CLIL on the factors 'linguistic knowledge' and 'world knowledge'. Because the CLIL group is taught more often in English than the non-CLIL group, the CLIL learners' mental lexicon most likely contains more English lexis (see, e.g., Zaunbauer et al. 2012). Grammatical awareness is likely to be influenced positively through more EFL lessons as well. Their linguistic knowledge is thus broader than that of the non-CLIL group. Furthermore, the CLIL learners might have more world knowledge, particularly topic-based knowledge,

7 The data on SES and age do not fulfil one of the two preconditions for an ANCOVA (no variance homogeneity [SES: $p = .01$; age: $p = .002$], homogeneity of regression slopes is given), yet there is no sound alternative statistical analysis which makes up for this lack.

because they deal with content topics extensively in their CLIL subjects. Looking at an extension of Frisch et al.'s (in this volume) model as found in Reckermann (2018: 106), two further factors are in constant interplay with reading processes: text knowledge and reading strategies. Again, the CLIL learners might have more general text knowledge since they are more often confronted with English texts in the context of their CLIL classes. In addition, they might have already developed more strategies to cope with the foreign language, including reading strategies, because they probably find themselves in situations where they need coping strategies frequently in the scope of their CLIL lessons. All four factors are likely to be influenced by CLIL in a positive way and then have a positive influence on the CLIL learners' EFL reading comprehension competences.

At a more general level, also the qualification for teaching EFL of the CLIL teachers, which is higher than that of the non-CLIL teachers (see Section 3.1), might have a positive impact on the four factors mentioned in the previous paragraph and thus on the CLIL group's better performance. This statement, however, should be treated cautiously. Further research is needed that takes a closer look at teachers' actual EFL classroom practices to see whether there is a difference between (a) those with a fully-fledged education and those with only a further training programme, and (b) those with an additional training programme for CLIL and those without. Furthermore, it seems worthwhile to investigate differences between teachers who hold a degree for both, the language and the content subject, and those who hold a degree for only one subject (Rymarczyk & Yearwood 2016). Last but not least, the question of whether this is a foreign language degree or a content degree might be interesting (ibid.), especially if one is interested in researching teaching reading.

The overall test results show a floor effect which indicates that the YLE-test that was used was too difficult for the learners. That is, not just for the non-CLIL learners but also for the CLIL learners, who on average scored less than 50 %, the tasks did not allow all learners to display their actual EFL reading competences. As already mentioned above, the test had not been piloted since Reckermann (2018) had used it successfully in her study with CLIL learners at the very beginning of their fourth grade. In retrospect, this decision was wrong, and a repeated piloting might have avoided the floor effect.

Looking at the analysis of the test in Table 2, one can see that the test is mostly concerned with reading at sentence and text level. Although reading at word level is implicitly part of most of the tasks, it is not tested itself. In contrast, the test results used in the EVENING study (Engel et al. 2009) and BIG study (BIG 2015), which mostly tested reading competence at word and sentence level, showed ceiling effects and was thus too easy (see Section 2.1). For future studies, a reading test for English reading comprehension competences that neither shows floor nor ceiling effects should possibly contain tasks at word, sentence and text level. A test like this, however, still needs to be developed. Yet, the test presented by Frisch et al. (in this volume) goes into this direction as it investigates reading comprehension at text level while at the same time it explores the reader's abilities at the level of basic reading and decoding processes.

Since the analysis of the teacher questionnaire revealed that the CLIL and non-CLIL teachers use reading activities at word, sentence and text level to various extents (see Section 3.2.1), the learners' scores for the test tasks were examined separately. Special attention was given to possible differences between tasks at sentence level and tasks at text level. This might also reveal whether the aforementioned floor effect can be traced back to one particular task type.

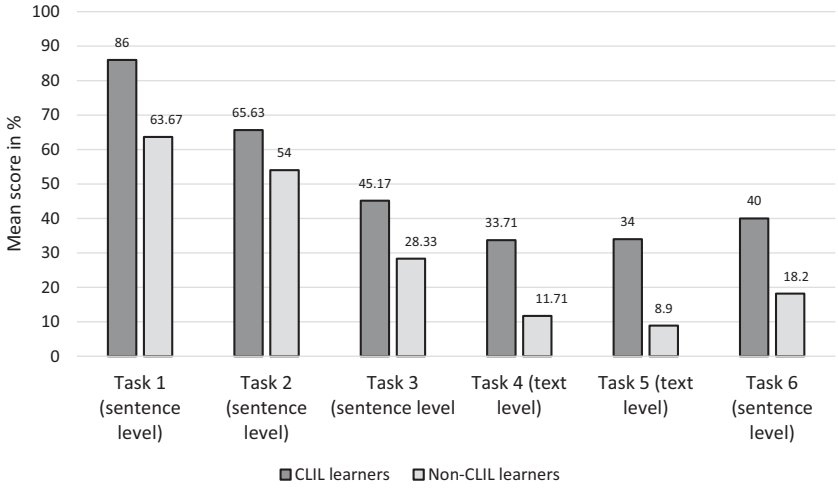


Fig. 5: *CLIL and non-CLIL learners' mean scores on the six different tasks of the YLE-test (in %)*

Figure 5 shows that the CLIL learners score higher than the non-CLIL learners on all six tasks. The largest difference is between the scores for Task 1 with about 33 % and the smallest is for Task 2 with about 11 %. The differences in mean score for the tasks at text level (Task 4 & 5) are 22 % and 26 %, which are also the greatest differences besides the ones for Task 1. One might have expected that the difference is greater for the tasks at text level and lower for the tasks at sentence level, but this is not the case. Figure 5 further reveals that the mean scores become lower in both groups from task to task, with the only exception of Task 6. Although the learners were given freedom of choice regarding the order, the majority of them might have started with Task 1 and then proceeded gradually from task to task. This approach has probably led to a loss in concentration and possibly also in motivation, resulting in the loss in points from task to task.

In a next step, the mean score of the two tasks at text level (Task 4 & 5) was analysed and compared to that of the tasks at sentence level (Task 1, 2, 3 & 6). The results are displayed in Table 6 and additional information is provided in the box plots shown in Figures 6 and 7.

Tab. 6: *EFL reading comprehension competences at sentence and text level (mean scores)*

		N	Mean score and standard deviation in %
Sentence level	Total	75	*M* = 52.58 % (*SD* = 20.70 %)
	CLIL learners	42	*M* = 60.87 % (*SD* = 18.25 %)
	Non-CLIL learners	33	*M* = 42.03 % (*SD* = 18.95 %)
Text level	Total	75	*M* = 23.18 % (*SD* = 26.29 %)
	CLIL learners	42	*M* = 33.47 % (*SD* = 27.92 %)
	Non-CLIL learners	33	*M* = 10.07 % (*SD* = 16.85 %)

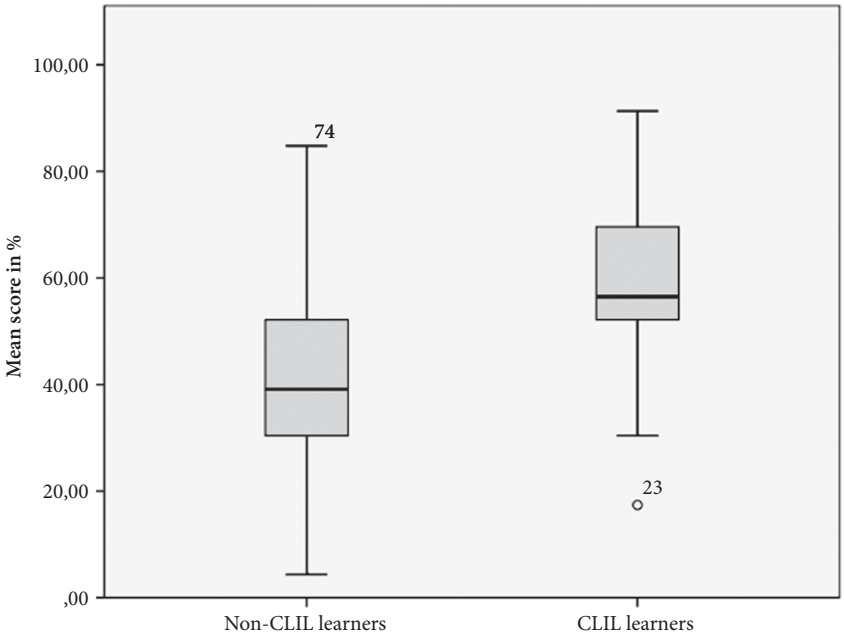


Fig. 6: *CLIL and non-CLIL learners' EFL reading comprehension competences at sentence level (Task 1, 2, 3, 6)*

As can be seen, the difference between the two groups is about 18 % in mean scores for the tasks at sentence level and about 23 % for the tasks at text level, both in favour of the CLIL learners. With respect to the tasks at sentence level, there is also variation within the two groups, as well as one outlier in the CLIL group at 17.39 %, which clearly falls below the group's mean score of 60.87 %. With respect to the tasks at sentence level, the variability is larger within the CLIL group. The data of the non-CLIL group contains the same two outliers that already showed in the data set regarding the general EFL reading competences (see Fig. 4), again clearly exceeding the group's mean score (10.07 %) at 41.18 % and 85.29 %.

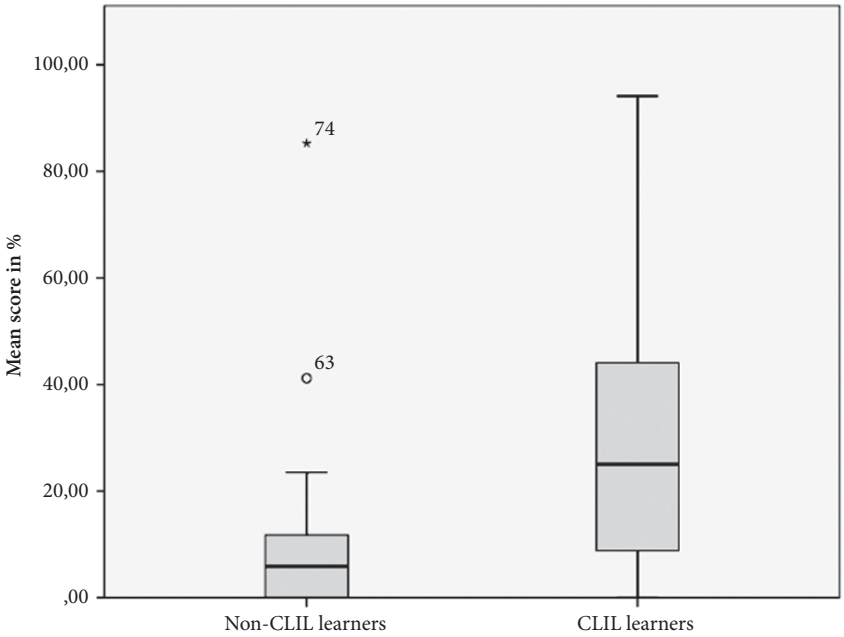


Fig. 7: *CLIL and non-CLIL learners' EFL reading comprehension competences at text level (Task 4 & 5)*

A one-way ANOVA shows that the EFL reading comprehension competences at sentence level, as well as at text level differ significantly for the different English programmes; sentence level: $F(1, 73) = 19.05$, $p = .000***$ and text level: $F(1, 73) = 18.00$, $p = .000***$. The differences in reading comprehension competences between the CLIL and the non-CLIL group are hence not only based on the students' performance on the tasks at text level, but the results for the tasks at sentence level differ just as much. Yet, the results indicate that the floor effect can be attributed to the tasks at text level. Both learner groups score significantly better, i.e. by about 30 %, on the tasks at sentence level than on those at text level, indicating that both groups have clearly more difficulties with the tasks at text level. This is further backed up by the fact that the non-CLIL learners only reach about 10 % on average on the two tasks at text level, which again shows that these tasks were clearly too difficult for them. Although

the CLIL learners outperform the non-CLIL learners on Task 4 and 5 as well, they also only score about 33.47 % on average, implying that they also had difficulties with these two tasks. However, a future reading test should still contain text level tasks in order to avoid the ceiling effect of a test that only includes word and sentence level tasks.

3.2.4 About a connection between the learners' overall opinions towards EFL reading and their reading comprehension competences

To assess the relationship between the learners´ overall opinions towards EFL reading and their reading comprehension competences, a Pearson product moment correlation coefficient (PPMCC) was computed. The PPMCC shows that there is an overall moderate positive correlation[8] between the two variables, $r = .436$, $N = 75$, $p < .001$***. That is, a more positive opinion towards EFL reading is correlated with increases in the learners' reading test scores. This finding is supported when looking at the two learner groups separately; CLIL learners: $r = .439$, $N = 42$, $p = .004$** and non-CLIL learners: $r = .366$, $N = 33$, $p = .036$*. Overall, it appears that the learners who generally enjoy EFL reading are more likely to have better results in EFL reading comprehension, regardless if they attended a CLIL or a non-CLIL programme.

This result underlines what has been stated for the impact of motivation on reading achievement and competences in Section 2.2. The more a learner is motivated to read, or as investigated in this study has a positive opinion about EFL reading, the better his/her actual reading comprehension competences.

4 Conclusion

This study examined EFL reading in CLIL and non-CLIL primary schools. It focused on CLIL and non-CLIL teachers' use of reading activities and materials, CLIL and non-CLIL learners' opinions and preferences towards EFL reading, the learners' reading comprehension competences and

8 Effect size after Cohen (1988): $|r| = .10$ (small), $|r| = .30$ (moderate), $|r| = .50$ (large)

a possible connection between their opinions of EFL reading and their reading comprehension competences.

Core results showed that, first, reading at text level is sometimes used in CLIL classes, but (almost) never takes place in non-CLIL classrooms, while reading at word and sentence level takes place similarly often. The use of fictional stories and factual texts takes place comparably often, while the Internet/computer is hardly used for EFL reading activities in both groups. Second, both learner groups have overall positive opinions towards EFL reading and show a balanced preference of reading aloud in contrast to reading silently. Their preferences regarding stories, factual texts, and the Internet/computer as reading materials also seem rather balanced. Third, the CLIL learners' EFL reading comprehension competences significantly exceed those of the non-CLIL learners, both in their overall mean scores as well as in the mean scores divided by tasks at sentence and text level. The YLE-test, however, showed floor effects and proved to be too difficult for both groups, particularly regarding the tasks at sentence level. Fourth, overall positive opinions towards EFL reading appears to have a positive impact on the learners' EFL reading competences, regardless of their CLIL or non-CLIL affiliation.

This confirms what previous studies have already found, namely that the CLIL programme makes a positive difference concerning the learners' EFL reading comprehension competences. However, what feature exactly of the CLIL programme is responsible for this difference could not be shown. It might be the differences in teaching practices, the differences in education of the teachers, the frequency with which different EFL reading activities are used, or, and this is likely, a combination of all these factors. The CLIL group's higher SES and younger age do not influence the results. More so, the more positive a learner's opinion towards overall reading, the better his/her reading comprehension, but this is independent of CLIL or non-CLIL teaching.

With respect to classroom practices, it would be desirable to generally include more EFL reading activities at text level. Ideally, reading at text level should not only be implemented in CLIL programmes, but in regular primary school EFL programmes as well. The CLIL and non-CLIL learners' overall positive opinions towards EFL reading should be sustained by using interesting, meaningful and challenging but manageable

EFL reading tasks, based on different reading materials and text types. In addition, the potential of the Internet as a source of differentiated EFL reading materials and activities should be investigated. Having underpinned the impact of an overall positive opinion towards EFL reading as part of reading motivation, primary school EFL teachers should pay careful attention to their learners' EFL reading motivation in order to lay a sound basis for more demanding reading activities at the secondary school level. Future research should, among other aspects, be concerned with finding a reading test that is suitable to assess young learners' EFL reading (comprehension) competences without displaying ceiling or floor effects.

Bibliography

Alter, Grit/ Frisch, Stefanie (2018), Literarisches Lernen schon in der Grundschule? The potential of picture books. *Grundschule Englisch* 63, 2–5.

BIG-Kreis der Stiftung LERNEN (2015), *Der Lernstand im Englischunterricht am Ende von Klasse 4. Ergebnisse der BIG-Studie*. München: Domino Verlag.

Bleyhl, Werner (2007), Schrift im fremdsprachlichen Anfangsunterricht – ein zweischneidiges Schwert. *Take Off!* 1, 47.

Bosenius, Petra (2009), Content and language integrated learning. A model for multiliteracy? In: Ditze, Stephan-Alexander & Halbach, Ana (eds.) (2009), *Bilingualer Sachfachunterricht (CLIL) im Kontext von Sprache, Kultur und Multiliteralität*. Frankfurt a.M.: Peter Lang, 15–26.

Brown, H. Douglas (2007), *Principles of Language Learning and Teaching* (5th edn.). White Plains: Pearson Longman.

Burmeister, Petra/ Piske, Thorsten (2008), Schriftlichkeit im fremdsprachlichen Sachfachunterricht an der Grundschule. In: Böttger, Heiner (ed.) (2008), *Fortschritte im frühen Fremdsprachenlernen. Ausgewählte Tagungsbeiträge Nürnberg 2007*. München: Domino, 183–193.

Chiu, Ming M. (2018), Contextual influences on girls' and boys' motivation and reading achievement: Family, schoolmates, and country. In: Garcia, Pelusa Orellana & Lind, Paula Baldwin (eds.) (2018), *Reading Achievement and Motivation in Boys and Girls. Field Studies and Methodological Approaches*. Cham: Springer Nature, 49–63.

Cohen, Jacob (1988), *Statistical Power Analysis for the Behavioral Sciences* (2nd edn.). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Dehaene, Stanislas (2010), *Lesen. Die größte Erfindung der Menschheit und was dabei in unseren Köpfen passiert* (trans. Reuter, Helmut). München: Knaus.

Diehr, Bärbel (2010), Research into reading in the primary school. A fresh look at the use of written English with young learners of English as a foreign language. In: Diehr, Bärbel & Rymarczyk, Jutta (eds.) (2010), *Researching Literacy in a Foreign Language among Primary School Learners. Forschung zum Schrifterwerb in der Fremdsprache bei Grundschülern*. Frankfurt a.M.: Peter Lang, 51–68.

Diehr, Bärbel/ Frisch, Stefanie (2010a), A roadmap to reading. Bewusstmachende Verfahren im Umgang mit der englischen Schriftsprache. *Grundschule* 9, 26–28.

Diehr, Bärbel/ Frisch, Stefanie (2010b), Lesen sie doch? Fragen an die LiPs-Studie (Lesen im Englischunterricht auf der Primarstufe). *IMIS Beiträge – Institut für Migrationsforschung und Interkulturelle Studien* 37, 143–163.

Diehr, Bärbel/ Rymarczyk, Jutta (2008), „Ich weiß es, weil ich es so spreche". Zur Basis von Lese- und Schreibversuchen in Klasse 1 und 2. *Grundschulmagazin Englisch* 6: 1, 6–8.

Doyé, Peter (1993), Fremdsprachenerziehung in der Grundschule. *Zeitschrift für Fremdsprachenforschung* 4: 1, 48–90.

Ehlers, Swantje (2007), Lesetheorien, Lesekompetenz und Narrative. In: Bredella, Lothar & Hallet, Wolfgang (eds.) (2007) *Literaturunterricht, Kompetenzen und Bildung*. Trier: Wissenschaftlicher Verlag Trier, 107–126.

Engel, Gaby (2009), EVENING – Konsequenzen für die Weiterentwicklung des Englischunterrichts in der Grundschule. In: Engel, Gaby et al., 197–215.

Engel, Gaby/ Groot-Wilken, Bernd/ Thürmann, Eike (eds.) (2009), *Englisch in der Primarstufe – Chancen und Herausforderungen. Evaluation und Erfahrungen aus der Praxis*. Berlin: Cornelsen.

Fay, Gesine/ Hellwig, Karlheinz (1971), *Englischunterricht in der Grundschule*. Hannover: Schroedel.

Freudenau, Tanja (2012), Multimedia dictionaries in the primary school. In: Lennon, Paul (ed.) (2012), *Learner Autonomy in the English Classroom. Empirical Studies and Ideas for Teachers*. Frankfurt a.M.: Peter Lang, 95–122.

Frisch, Stefanie (2013), *Lesen im Englischunterricht der Grundschule. Eine Vergleichsstudie zur Wirksamkeit zweier Leselehrverfahren*. Tübingen: Narr.

Frisch, Stefanie/ Breul, Carsten/ Diehr, Bärbel/ Kastens, Claudia/ Becker, Annette (in this volume), Developing and assessing reading comprehension in primary EFL learners.

Glaser, Karen/ Schrader, Anne (2019a), Learning to read, reading to learn. Fremdsprachliche Sachtexte erschließen. *Grundschule Englisch, 68*, 30–32.

Glaser, Karen/ Schrader, Anne (2019b), What will the weather be like in Sydney? Sachtexten Schritt für Schritt Informationen entnehmen. *Grundschule Englisch, 68*, 26–29.

Henseler, Roswitha/ Surkamp, Carola (2007), Leselust statt Lesefrust. Lesemotivation in der Fremdsprache Englisch fördern. *Der fremdsprachliche Unterricht Englisch, 41*, 2–11.

Jöckel, Alicia (2016), Lesen und Schreiben im Englischunterricht der Grundschule. In: Böttger, Heiner & Schlüter, Norbert (eds.) (2016), *Fortschritte im frühen Fremdsprachenlernen. Tagungsband zur 4. FFF-Konferenz*. Braunschweig: Westermann, 134–142.

Kierepka, Adelheid (2010), Historischer Blick auf den Fremdsprachenunterricht in der Grundschule. *IMIS Beiträge – Institut für Migrationsforschung und Interkulturelle Studien* 37, 83–97.

Kolb, Annika (2013), Extensive reading of picture books in primary EFL. In: Bland, Janice & Lütge, Christiane (eds.) (2013), *Children's Literature in Second Language Education*. London: Bloomsbury, 33–43.

Legutke, Michael K./ Müller-Hartmann, Andreas/ Schocker-v. Ditfurth, Marita (2009), *Teaching English in the Primary School*. Stuttgart: Klett.

Leubner, Martin/ Saupe, Anja/ Richter, Matthias (2016), *Literaturdidaktik* (3rd edn.). Berlin: de Gruyter.

Marinak, Barbara A./ Gambrell, Linda B./ Mazzoni, Susan A. (2013), *Maximizing motivation for literacy learning. Grades K-6*. New York: The Guilford Press.

Mindt, Dieter (2007), Fortschritte im frühen Fremdsprachenlernen von 2000 bis heute und Perspektiven für die Zukunft. In: Böttger, Heiner (ed.) (2007), *Fortschritte im frühen Fremdsprachenlernen. Ausgewählte Tagungsbeiträge Nürnberg 2007*. München: Domino, 12–22.

MSJK – Ministerium für Schule, Jugend und Kinder des Landes Nordrhein-Westfalen (2003), *Richtlinien und Lehrpläne zur Erprobung für die Grundschule in Nordrhein-Westfalen. Englisch*. Frechen: Ritterbach.

MSW NRW – Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalens (2008), *Richtlinien und Lehrpläne für die Grundschule in Nordrhein-Westfalen: Deutsch, Sachunterricht, Mathematik, Englisch, Musik, Kunst, Sport, Evangelische Religionslehre, Katholische Religionslehre*. Frechen: Ritterbach.

Paulick, Christian/ Groot-Wilken, Bernd (2009), Rezeptive Fähigkeiten und Fertigkeiten am Ende der 4. Klasse unter besonderer Berücksichtigung der sprachlichen Schülerbiographien. In: Engel, Gaby et al., 179–196.

Piske, Thorsten (2010), Positive and negative effects of exposure to L2 orthographic input in the early phases of foreign language learning. A review. In: Diehr, Bärbel & Rymarczyk, Jutta (eds.) (2010), *Researching Literacy in a Foreign Language among Primary School Learners. Forschung zum Schrifterwerb in der Fremdsprache bei Grundschülern*. Frankfurt a.M.: Peter Lang, 37–50.

Reckermann, Julia (2018), *Reading Authentic English Picture Books in the Primary School EFL Classroom. A Study of Reading Comprehension, Reading Strategies and FL Development*. Berlin: Peter Lang.

Richter, Karin/ Plath, Monika (2012), *Lesemotivation in der Grundschule. Empirische Befunde und Modelle für den Unterricht* (3rd edn.). Weinheim: Beltz Juventa.

Rymarczyk, Jutta/ Musall, Annika (2010), Reading skills of first graders who learn to read and write in German and English. In: Diehr, Bärbel & Rymarczyk, Jutta (eds.) (2010), *Researching Literacy in a Foreign Language among Primary School Learners. Forschung zum Schrifterwerb in der Fremdsprache bei Grundschülern*. Frankfurt a.M.: Peter Lang, 69–88.

Rymarczyk Jutta/ Yearwood, Tanyasha (2016), 1 Subject + 1 Subject = 1 Subject -Teacher? Teacher Beliefs on Cross-Curricular Collaboration in Forming the Design of CLIL Degree Courses. In: *Language in Different*

Contexts. Research papers. Lithuanian University of Educational Sciences, Faculty of Philosophy, 260–268.

Schmid-Schönbein, Gisela (2008), *Didaktik und Methodik für den Englischunterricht*. Berlin: Cornelsen.

SSK – Sächsisches Staatsministerium für Kultus (2019), *Lehrplan Grundschule Englisch*. [Online: http://lpdb.schule-sachsen.de/lpdb/web/downloads/2427_lp_gs_englisch_2019.pdf?v2 17.08.2021].

Steck, Andrea (2009), *Förderung des Leseverstehens in der Grundschule. Fortbildungsbausteine für Lehrkräfte*. Baltmannsweiler: Schneider Verlag Hohengehren.

Steinlen, Anja K. (2018), Grundschulen mit verschiedenen Englischprogrammen: Geschlecht und Mehrsprachigkeit auf dem Prüfstand. *Zeitschrift für Fremdsprachenforschung* 29: 1, 3–26.

Steinlen, Anja K. (2021), *English in Elementary Schools. Research and Implications on Minority and Majority Language Children's Reading and Writing Skills in Regular and Bilingual Programs*. Tübingen: Narr Francke Attempto.

University of Cambridge (ed.) (2014), *Cambridge English. Young Learners: Young learners English tests (YLE). Sample papers. Starters Movers Flyers*. Cambridge: Cambridge University Press.Waas, Ludwig (2014), Auf dem Weg zu kommunikativen Fähigkeiten. Songs, rhymes and then? *Grundschulmagazin Englisch* 12: 1, 35–36.

Wilden, Eva/ Porsch, Raphaela (2015), Die Hör- und Leseverstehensleistungen von Kindern im Fach Englisch am Ende der Grundschulzeit unter besonderer Berücksichtigung lebensweltlicher Ein- und Mehrsprachigkeit. In: Kötter, Markus & Rymarczyk, Jutta (eds.), *Englischunterricht auf der Primarstufe: Neue Forschungen – weitere Entwicklungen*. Frankfurt am Main: Peter Lang, 9–42.

Wilden, Eva/ Porsch, Raphaela/ Ritter, Markus (2013), Je früher desto besser? Frühbeginnender Englischunterricht ab Klasse 1 oder 3 und seine Auswirkungen auf das Hör- und Leseverstehen. *Zeitschrift für Fremdsprachenforschung* 24: 2, 171–201.

Wirbatz, Karoline/ Reckermann, Julia (2020), Ein Vergleich des Leseverständnisses im Englischen und Deutschen von Mädchen und Jungen der dritten Klasse in unterschiedlichen Englischprogrammen. In: Böttger, Heiner; Festmann, Julia & Müller, Tanja (eds.), *Language*

Education and Acquisition Research. Focusing Early Language Learning. Bad Heilbrunn: Klinkhardt, 117–136.

Zaunbauer, Anna C. M. (2007), Lesen und Schreiben in der Fremdsprache – von Anfang an. *Take Off!* 1, 46.

Zaunbauer, Anna C./ Gebauer, Saskia, K./ Möller, Jens (2012), Englischleistungen immersiv unterrichteter Schülerinnen und Schüler. *Unterrichtswissenschaft: Zeitschrift für Lernforschung* 40, 315–333.

Karen Glaser

Scaffolding creative writing in the primary EFL classroom: Exploring the role of picture dictionaries and composition guidelines in the creation of *Elfchen* poems

Abstract This paper presents an analysis of 43 foreign language (L2) English *Elfchen* poems created by two groups of German 4th-graders in their second year of English as a foreign language (EFL) instruction. The children were familiar with this text type from their first language (L1) German lessons but had never encountered this in English before. Both groups used picture dictionaries to look up words, but they differed with regard to the instructional scaffolding they were given. While Group A (n = 17) was provided with two sample poems plus a revision of the required number of words on each line, Group B (n = 26) additionally received information on the required word classes and the type of content to be expressed in each line as well as word provision beyond the dictionary. The poems were analyzed for adherence to the composition guidelines, word classes, type-token-ratio, and spelling accuracy. The study gives insights into how young learners at the pre-A1 level master the task of composing such a poem and what kind of scaffolding appears conducive in the process.

1. Introduction: Literacy in primary English education and research (in Germany)

Literacy refers to both the receptive skill of decoding written symbols to interpret their meaning (reading) and the productive ability to encode thoughts, intentions and meaning in such a written medium (writing). In a foreign language (L2), those cognitive processes usually pose greater challenges than in the first language (L1), as the decoding and encoding of meaning is closely connected to the lexical repertoire as well as pronunciation, orthography, and grammar. Accordingly, literacy is especially challenging at the lower L2 proficiency stages, which is where young L2 learners are usually situated. Although L2 literacy has been defined as a central component of communicative competence for German secondary

education (KMK 2003: 8), it takes a backseat in many German curricula for primary English language teaching (PELT), many of which are still rooted in the Primacy of Orality Principle. This is a vestige from the so-called literacy debate (*Schriftlichkeitsdebatte*) that surrounded the introduction of PELT as an obligatory school subject in 2004 and which was heavily marked by the fear of overtaxing the children if they were subjected to the written code in English. It was argued that the orthographic system of English with its rather intransparent deep orthography would confuse the children in their concomittant acquisition of the more transparent orthography of German with a comparatively high degree of regular phoneme-grapheme correspondences (PGCs) (e.g. Bleyhl 1999). As this narrow focus on orthography shows, the debate viewed literacy almost exclusively on the word level, excluding communicative language use on more textual planes (reading story books, composing messages etc.). In other words, the discussion was limited to considerations of the "feedforward" of going from spelling to pronunciation for reading and the "feedback" of going from word pronunciation to orthographic encoding for writing (Aro & Wimmer 2003). Although these are important subskills of L2 literacy development, they fail to take into consideration that written language development encompasses more than word spelling (Burmeister 2010; Reckermann 2018), and that the written mode serves a range of supporting functions in L2 acquisition, also with young learners, as pioneering studies such as Reichart-Wallrabenstein (2004) or Duscha (2007) have shown for the German PELT context.

Still, many of the current PELT curricula bear the marks of this reluctance to systematically address literacy skills in beginning English as a Foreign Language (EFL) instruction. The state of Saxony, where the present research project was carried out, has one of the most conservative curricula in that respect, which in essence dates back to 2004 and covers literacy skills in an exclusively recognitional and reproductive fashion. This stands in quite some contrast to the research into literacy development of young learners of English (YLEs) over the past years, which has shown rather unanimously that the learners cope with the L2 written code well, even if it is introduced as early as first grade (Rymarczyk & Musall 2010) or if English is learned as the second additional language besides German (Rymarczyk 2010). More recent research on reading in

second and fourth grades (Diehr & Frisch 2010; Frisch 2013; Reckermann 2018) also suggests that the children handle the written L2 code well, enjoy reading, and benefit from it for their overall language abilities. The EVENING study conducted in North Rhine-Westphalia (NRW) with 1500 learners at the end of grade 4 (Engel, Groot-Wilken & Thürmann 2009) showed that fewer than 4 % of the children struggled with the English reading task, suggesting that the children's competences had been largely underestimated. A replication of the EVENING reading test in a study in Switzerland (Husfeldt & Bader Lehman 2009), in the GanzIn study with over 6,500 children (Wilden, Porsch & Ritter 2013), and in the German-wide BIG study (BIG-Kreis 2015) yielded similar results.

Yet, despite the overall positive outcome of these studies, this research overview also indicates a research lacuna, as almost all of these studies focus on reading and thus only on the receptive aspect of literacy. The productive skill of writing has, however, been hardly addressed in PELT research, and consequently next to no empirical findings exist about writing development and teaching in the early stages of EFL acquisition, let alone a theory of literacy acquisition for foreign language beginners at primary level (Burwitz-Melzer 2010). The occasional vivid discussions about writing in foreign language learning that have been carried out in the past (e.g., Hufeisen 2005; Kupetz 1997) all refer to older and thus cognitively more mature learners. The occasional forays of international research into the writing skills of YLEs also mostly address learners in grades higher than the fourth (e.g. Griva, Tsakiridou & Nihoritou 2009 on the writing strategies employed by 6th-grade EFL learners in Greece).

Among the few existing studies into PELT writing is Rymarczyk (2008), who reports about a small-scale study with learners who had received English instruction from grade 1 and who were asked at the end of grade 2 to write down the English words they know. The spellings employed by the children, whose instruction had not systematically addressed the written form, are suggestive of L1 transfer of German PGCs, including noun capitalization, which is a regular feature of German spelling (all nouns are to be capitalized). Rymarczyk interprets the children's repeated invented spellings as indicators of internal rules which the children had developed for themselves based on the only writing system they had been systematically introduced to, viz. that of their L1. Glaser and Schrader

(2019) observed similar phenomena in Saxon primary schools: noun capitalization, application of German PGCs (such as <kat> for *cat*) as well as omission of silent letters (e.g., <ston> for *stone*). The BIG study of learners at the end of Grade 4 (BIG-Kreis 2015) featured a writing section which asked the children to label pictures with model words (copying) and without (writing from memory), and to write down a color, weekday, and a clothing item of their choice without model. In addition, the children were asked to complete sentence prompts. The findings for spelling are in line with Rymarczyk's (2008) and Glaser and Schrader's (2019) observations of L1 transfer; in addition, they showed that some children had started to develop initial hypotheses about PGCs in English (e.g., that <oo> can represent [u:]) although they did not necessarily apply them with orthographic correctness (e.g., <shoos> for *shoes*).

The main focus of all of these studies was, however, writing on the word level, which is a gap that the present study attempts to address. In analyzing the creation of short poems (*Elfchens*) by 4th-grade PELT students in Saxony on the backdrop of the scaffolding the learners received, the study adds insights into the written L2 production of YLEs and into the instructional support that is potentially helpful in the process.

2. Creative writing in German PELT and *Elfchen* poems

Given that German PELT instruction does not place a major focus on the skill of writing, it is not surprising that *creative* writing plays an even more marginal role. Merely two of the 16 federal curricula (Baden-Wuerttemberg, NRW) mention creative writing in the form of poems, and one (Bremen) contains the term "creative copying", i.e. the creation of short texts on the basis of model texts. For the project's Saxon context, a poem creation thus constituted an innovative format, a view corroborated by the fact that neither the EFL teacher nor the leaners had dealt with *Elfchens* in English before. Overall, this situation is rather lamentable, as this type of poem is rather frequently used in primary L1 German instruction; accordingly, EFL instruction could easily build on this familiarity and benefit from positive transfer and reinforcement effects.

The name *Elfchen* is a play on words in German, as these poems consist of 11 words (German: *elf*). The diminutive ending *-chen* turns the name

into 'small eleven' and at the same time makes it synonymous with 'little fairy'. Each *Elfchen* features a strict structure of five lines, on which the 11 words are distributed in a 1-2-3-4-1 pattern. In addition to the just-mentioned reinforcement effects, those formal features hold the potential to instill a first awareness of the differences in the written forms of the L1 and the L2, such as the existence of full and contracted forms in English (e.g., *I'm* vs. *I am*), which yield different numbers of words and can thus be intentionally manipulated to fulfill the poem criteria, or the difference in the spelling of compound nouns, which in German are normally spelled as one word, not two.

In contrast to this structural criterion, the semantic requirements for the five lines are not set in stone and can thus vary between schools and teachers, both in L1 and L2 lessons, making them flexible and easily adjustable to the needs, interests and skills of the learners. One possible set of guidelines is provided by Willing & Geldschläger (n.d.): line 1: noun delineating the poem topic; line 2: action (or non-action) of the word from line 1; line 3: location or characterization of the word; line 4: our impression of the word; line 5: a one-word summary of the poem. Quite differently, Steitz-Kallenbach (1998: 33) proposes the following sequence: a color – the topic – a characteristic – an event or activity – a punchline summary. With regard to the L2 young beginning learners' classroom, Glaser (2018: 168) suggests that it might be judicious to refrain semantic requirements, at least initially. Overall, this flexibility makes *Elfchens* so suitable for primary school and is also a reason why they deserve greater attention in PELT: They can, for instance, be used to practice previously introduced vocabulary, phrases or syntactic patterns in a very focused way, but they can also be used in a more open fashion to invite brainstorming, language exploration, and self-expression, holding great potential for differentiating instruction. Lastly, as *Elfchen* poems feature only 11 words, they are suitable for beginning learners in that they are in line with the CEFR descriptors for writing skills at the most basic proficiency level, A1, which feature the creation of "simple isolated phrases and sentences" (Council of Europe 2001: 61).

3. Scaffolding writing

The term *scaffolding* was transferred into the educational context by Wood, Bruner and Ross (1976) as a metaphor for any kind of supportive instructional technique used by educators to help a child reach higher levels of understanding or performance and thus of independence in learning. Like a physical scaffold in construction work, it is meant to bridge the gap between what a child can achieve unaided and with assistance. Also like a physical scaffold, instructional scaffolds are supposed to be temporal in nature, to be removed when no longer needed. With regard to foreign language teaching, a great deal of scaffolding is connected to the successful use of the target language, both receptively and productively. Genetsch et al. (2012: 105 f.) distinguish between *input scaffolding* (support for comprehension), *output scaffolding* (support for production), and *process scaffolding* (support for task completion). In addition, Brush and Saye (2002) distinguish between *hard* and *soft scaffolding*, with hard scaffolding denoting all the materials developed and techniques planned before the instruction (including curricular guidelines), and soft scaffolding referring to the teacher's spontaneous (re)actions during the lesson.

The study presented here analyzes the effect of scaffolding on the written products of young learners of English in a German primary school with a major focus on differences in hard, i.e. planned, scaffolding, including both process scaffolding (support for completing the task 'poem creation') and output scaffolding (support for written L2 production). As further detailed below (cf. Section 6), the two learner groups examined here differed with regard to the provision of unknown words by the teacher, but both groups were comparable in that every learner was provided with a copy of a picture dictionary aiming at young German learners of English published by DUDEN (Müller-Wolfangel, Pardall & Prusse 2012). Picture dictionaries are considered a vital resource of lexical material for young beginning learners as they help the learners to "quickly attach meaning to new terms" via the visual semantization inherent in the picture or photographs (Chujo & Nishigaki 2005: 36). The DUDEN dictionary was chosen as it provides PELT relevant vocabulary grouped according to topics and has a clear layout with unambiguous and appealing pictures, which was considered an important prerequisite for the children's independent, individual work. Each topic spans a

double-page captioned with the topic heading, which was considered conducive to the poem task as it means that words were grouped thematically. In addition to the visual meaning provision, the dictionary presents the written form (spelling) of the respective words. The link between meaning and written form is one of the most basic aspects of vocabulary depth (Nation 2001), and, together with the pronunciation-meaning connection, an important component of the meaning-spelling-pronunciation triangle that forms the most basic level of word knowledge (Glaser 2018: 157).

4. Research aims

The study presents an analysis of beginning YLEs' creative writing of *Elfchen* poems by analyzing the written products of two 4th-grade classes on the backdrop of the different kinds of scaffolding the groups received. More specifically, the analysis targeted the adherence to the composition guidelines and the learners' spelling mastery, addressing the following research questions (RQs):

(1) Poem composition:

   (a) How do the YLEs adhere to the composition criteria presented to them during the instruction, and to what extent are the features of the model poems reflected in the learners' work?
   (b) Did the different kinds of scaffolds (models and composition criteria) lead to any noticeable differences in the poems created by the two learner groups?

(2) Spelling: To what extent do the poems exhibit spelling difficulties, and are these different between the groups?

5. Participants

Forty-three young learners of English participated in the study and composed one *Elfchen* poem each. These learners came from two 4th-grade classes (n_A = 17; n_B = 26) from two consecutive school years in a German primary school and had been receiving their EFL education from the same English teacher. All children were 9–10 years old and had been taught EFL since Grade 3 with two 45-minute lessons per week. The school is a state school located in Saxony, and the EFL instruction was based on the

Saxon curriculum (Sächsisches Staatsministerium für Kultus 2004/2009). Parental consent was obtained for all participants.

The proficiency of the YLEs was pre-A1 bordering on A1. Proficiency testing was not possible for organizational reasons, but an evaluation by the English teacher was obtained for the two groups using the more detailed ACTFL proficiency descriptors. This evaluation[1] is provided in Table 1 along with demographic and linguistic background information. As can be gleaned from this table, Group B was slightly stronger than Group A in the oral skills (listening and speaking); for the written skills of reading and writing the groups were, however, comparable at the most basic level of 'novice-low.'

Tab. 1: Learner Background Information

	Group A	**Group B**
N	17 (out of a class of 24[a])	26 (intact class of 26)
School year	4	4
Age	9–10	9–10
Proficiency (ACTFL) (EFL teacher's evaluation)	Listening: novice-mid Speaking: novice-low, for some novice-mid Reading: novice-low Writing: novice-low	Listening: novice-high Speaking: novice-mid, for some novice-high Reading: novice-low Writing: novice-low
L1s	German: 15 bilingual Mandarin-German: 1 bilingual Italian-German: 1	German: 24 Mandarin Chinese: 1 Polish: 1

a Only the data by children who attended all three intervention sessions and who obtained written parental consent were included in the analysis. In this group, seven children did not fulfill these criteria.

1 Note that this is a very broad characterization of entire classes, which fails to depict any individual variation between the learners. Also note that evaluation is based on the teacher's impressions rather than objective test results. The author acknowledges these methodological limitations.

6. Pedagogic intervention

For each group the intervention per se was provided in two lessons that took place several days apart. In a subsequent third lesson, the children were individually taken to a separate room to audio-record the oral presentation of their poems. In the first lesson, the intervention started after a short warm-up and continued until the end of the lesson, spanning approximately 40 minutes. In the second lesson, approximately 35 minutes were used for the intervention. In each group, the intervention and the data collection were carried out by a graduating PELT teacher trainee in close collaboration with the children's EFL teacher, who was present during both intervention lessons and taught the class in the third while the trainee conducted the recordings. The YLEs were familiar with the text type *Elfchen* from their L1 German instruction but had never encountered it in their EFL lessons, neither receptively nor productively. The EFL teacher had never used this format in any of her English classes either prior to the first round (Group A), and was asked not to introduce it to Group B before the second intervention round in order to keep this starting condition equal between the two groups. Likewise, neither the learners nor the EFL teacher had worked with picture dictionaries (or any other dictionaries) in their EFL lessons before the intervention.

The intervention in the two groups was identical in terms of the overall sequence of the two lessons and the subsequent recording. In the first lesson, the children were introduced to the task (composing an *Elfchen* in English), including a review of the 11-word requirement and 5-line structure from their German lessons. They were then provided with the concrete instructions and requirements for the English poems (more details below) and subsequently introduced to the picture dictionary. In the remainder of the first lesson, the students started to work on their drafts on a sheet that contained five lines of varying length to implicitly scaffold the poem structure. The second lesson was devoted to finishing the drafts, copying the poems onto a new sheet of paper featuring the same lines, and practicing reading the poems out loud.

Apart from these identical steps, the instruction in Groups A and B differed with regard to the scaffolding provided to the learners, rooted in the intention to improve the instruction from one year to the next. Those

differences concerned (a) the provision of models and composition criteria, and (b) the language provision by the teacher, both of which are now explained in some more detail.

6.1 *Model provision and composition criteria*

In the first group, the poem was modeled for the learners by means of two sample poems which had been created by the regular EFL teacher: (1) *summer* | *I like* | *to go swimming* | *when it is hot* | *fun;* (2) *Easter* | *I love* | *hunting Easter eggs* | *in our big garden* | *surprise.* In either model, the first line contained the topic, while the second through fourth lines formed a sentence. This latter feature was, however, not pointed out explicitly to the children, nor did the children receive any instructions as to word classes to be use or content to be expressed on the lines. The only explicit requirement was to use the first line for the poem's topic.

After the first round the teacher reported that the learners had handled the *Elfchen* creation with much greater ease than expected, and expressed the wish to give more detailed requirements in the second round. Emulating *Elfchens* in which each line constitutes one unit of meaning, she chose the pattern 'topic – descriptive characterization by means of adjectives – event/activity – personal impression – summary'. The model provision in Group B was done by means of the poster reproduced in Figure 1, which not only featured a written reminder of the number of words to be produced on each line (on the left), but also (on the right) the word classes to be used in lines 1, 2, and 5, and the contents to be expressed in lines 3 (What is happening with the noun?) and 4 (What can you feel, see, think?). Accordingly, the model provision in the two groups differed in that the model in Group B did not contain a sentence spanning lines 2–4, and in that the model provision in the second round was more explicit than that used in the first round, featuring metalinguistic composition guidelines.

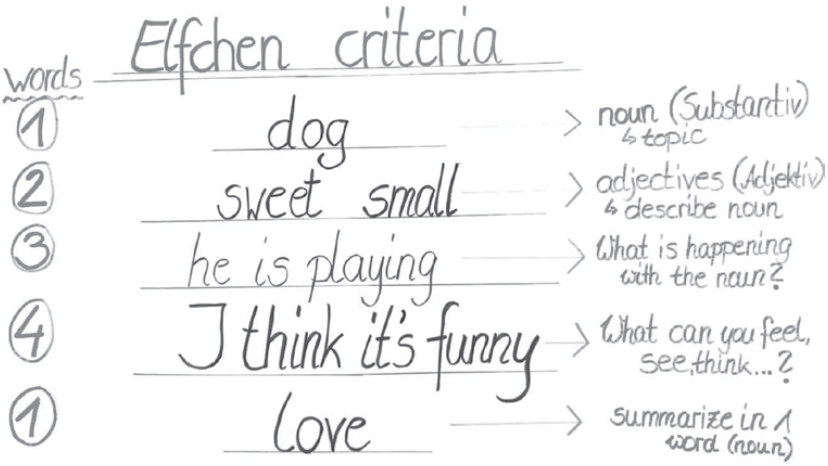


Fig. 1: Model Elfchen including composition criteria used in Group B

6.2 *Provision of words in addition to dictionary*

For the first intervention round, it had been decided not to provide the learners with words in addition to those they already knew and those they found in the dictionary. This was rooted in the interest to see how the YLEs worked with picture dictionaries. However, this restriction was relaxed somewhat during the Group A lessons when some major weaknesses of the dictionary became obvious, in order not to demotivate the students, and it was completely abandoned in the second round. Rather, the learners in Group B were systematically supported in their word search by collecting all requested words visibly on a big *Words you need* poster on the board via L1-L2 pairing so that the learners had constant access to both the meaning and the spelling. Pronunciation was provided by the trainee teacher or the EFL teacher. The poster was hung up again at the beginning of the second lesson and continued to have an overview of all the words enquired about by Group B.

7. Data and analysis procedures

The results are based on 43 *Elfchens*, which were analyzed in their written form; with illegible passages the recordings were consulted for hints as to what the child had intended to express.

To derive answers for the first research question (1a – adherence to models and composition requirements), the poems were analyzed with regard to:

(i) number of words produced (target: eleven)
(ii) word distribution on the five lines (target: 1-2-3-4-1)
(iii) for Group A: the extent to which the word on the first line denotes the topic of the poem (explicit requirement), and the extent to which the children composed sentences spanning lines 2–4 (implicitly modeled)
(iv) for Group B: the extent to which the children adhered to the explicit word class and content requirements stipulated in the composition criteria (cf. Fig. 1)

To answer research question 1b (differences between the groups resulting from scaffolding differences), the poems were analyzed with regard to:

(v) function-to-content word ratio (FW-CW ratio) to detect possible effects of the word class requirements in Group B that favored content words
(vi) type-token-ratio (TTR) to shed light on lexical variation

To answer the second research question (spelling mastery), the poems were analyzed for

(vii) spelling accuracy (via a rating scale from 0 to 5, see below), and
(viii) error types (open-ended coding, see below).

Spelling accuracy (vii) was measured via a six-point rating scale ranging from zero to five. It was informed by the scale which Tangel and Blachman (1992) had developed for aural dictation with L1 English learners and adapted to fit the study's L2 production context (e.g. by including transfer phenomena) as follows:

0 – random string of letters, unintelligible both in the poem and the recording
1 – word is only identifiable based on the audio recording (e.g., <nich> for *nice*)
2 – word is recognizable but has incorrect spelling at the phoneme level (e.g., <betiful> for *beautiful*)
3 – the spelling represents the phonemes correctly but features minor spelling errors (e.g., <wonderfull>)
4 – the spelling is correct except for capitalization or apostrophe errors (e.g., <Head>, <I'>)
5 – the spelling is correct

Error types (viii) were coded in an open, data-driven fashion informed by Rymarczyk's (2008) and Glaser and Schrader's (2019) coding schemes. In its final form, it featured the following categories: capitalization error, phonetic approximation, double consonant, similar L2 word, apostrophe error, <th>-related error, silent -e drop, and other German spelling conventions.

8. Results and discussion

8.1 *Research question 1: Poem composition*

The first research question concerned possible differences in the structural make-up of the poems. In the following, the results are presented with regard to (a) adherence to the composition criteria and models, and (b) lexical differences between the two groups.

(a) Adherence to the composition criteria and models

As for (i) the number of words (target: eleven), 39 of the 43 poems featured exactly 11 words. In Group A, one student produced 10 words, leaving the last line empty. In Group B, also one student (out of 26) wrote 10 words, producing only three words on line 4 (*I like duck*) instead of the required four. Two children in Group B wrote 12 words, with one filling the last line with 2 words (*ridden riding*), and the other producing 4 on line 3 (*I like the mouse*) instead of the required 3. In terms of (ii) word distribution on the five lines (target: 1-2-3-4-1), the only deviations were those just reported and thus a result of the word count deviation. Accordingly, with

merely four children deviating from the target by only one single word either way, the data suggests that the formal requirements of creating an *Elfchen* in L2 English are well manageable for young learners.

The analysis then proceeded to (iii) the composition criteria for Group A, viz. the explicit requirement to indicate the topic on the first line, and the implicit modelling of a sentence spanning lines 2–4. The explicit requirement was met by all 17 children, who either wrote the topic heading used by the dictionary (n = 11; *animals* (7), *pets* (2), *jobs* (1), *fruit* (1)) or a representative (hyponym) of a category (n = 6; *monkey, turtle, dolphins, dog, soccer, swimming*). The implicitly presented criterion of a sentence spanning lines 2–4 was not visible in any one of the poems; however, in 13 of the 17 poems, the writers reproduced the pattern from lines 2 and 3, starting with *I like* (8), *I love* (4) or *I have* (1) on line 2 and presenting a direct object on line 3 involving at least the first word. Ten of these thirteen poems feature a new sentence on line four (e.g. *the pig eats gras* [sic]), making this pattern the most frequently employed structure in Group A. The four children who had not used line 2 for a sentence beginning chose to express an adjective-based characterization instead (e.g. *so sweet*), but three of these also produced a sentence on line 4 similar to the other learners.

Taken together, those findings suggest that of the implicitly presented three-line sentence pattern, the initial two lines featuring a subject + predicate + direct object structure were reproduced by the learners. Per contrast, no poem featured the added adverbial phrase expressed in the fourth lines of both sample poems. Only one poem in Group A featured an adverbial phrase, in this case embedded in a noun phrase (*swimming with a monofin*). This poem was entirely error-free and had been created without any help from the dictionary or teacher-provided vocabulary, which suggests that the author was probably among the more advanced learners in this group. Accordingly, the fact that this student was the only one who produced an adverbial phrase despite the prominence of adverbial phrases in the model poems invites a few assumptions, which seem worth exploring in future studies: First, adverbial phrases might emerge at a later developmental stage than that of most of the participants, which in turn might be linked to constraints in processability of this syntactic pattern in the model (Pienemann 2003). Alternatively, the learners might have noticed it but did not possess the linguistic means to express it, as the EFL

and the trainee teacher reported that many learners in Group A inquired about prepositions, which the dictionary did not provide. No information was collected as to which part of the poem the learners required these for, but it is conceivable that some learners might have produced adverbials in the form of prepositional phrases if they had been assisted with the respective prepositions. A third potential explanation of the lack of adverbial phrases concerns the saliency of the requirement: As it was merely implicitly modeled rather than explicitly communicated, it might not have reached the learners' attention at the level of noticing (Schmidt 2001).

The poem requirements for Group B (analysis step iv) were first analyzed with regard to the word class requirements set for line 1 (a noun delineating the topic), line 2 (two adjectives describing the noun), and line 5 (noun summarizing the poem). Table 2 shows the rate of adherence for each of these lines.

Tab. 2: Adherence to formal criteria in Group B

Line	Requirement	Number of poems fulfilling requirement	Alternatives employed
1	Noun delineating the topic	26 (100 %)	n/a
2	Two adjectives describing the noun	22 (85 %)	n+adj (2), adj+n (1), *be*+adj (1)
5	Noun summarizing the poem	22 (85 %)	verb (2), adj (2)

Similar to Group A, the noun requirement for the first line was unproblematic, with six children listing the category (*pets* (3), *jobs* (2), *food* (1)) and 20 a hyponym (various animals (19), *strawberries* (1)). The requirements for line 2 (two adjectives) were also met with comparative ease, with only three deviations. In two cases, the children repeated the noun from the first line and added an adjective (*penguin smoll* [sic], *kitten small*); in the third case, the student produced an adjective + noun (*sweet hair*), and in the fourth case, the child connected the adjective to the topic noun via the copula (*is grin*[2] [sic]). Despite deviating on a formal level (word class), however, these children managed to provide some form of

2 <grin> = *green*.

description. Finally, the results for line 5 also indicate that the learners handled the explicit metalinguistic composition criteria rather well, with again only four children not providing a noun but producing an adjectival (*sweet, short*) or verbal line (*run; ridden riding*) instead.

A closer look at the kinds of nouns chosen by the learners reveal, however, that the majority of the 22 final nouns cannot be said to constitute some sort of a summary of the poem but rather the hypernym (8) or hyponym (1) of the topic noun (e.g. *elephant – animal*) or merely a repetition of a noun previously used in the poem (5). What is more, only two of the 22 nouns were abstract in nature, and these were a reproduction of the word *love* used in the sample poem. An *Elfchen* summary is, however, usually carried out on a more abstract plane describing impressions, thoughts, feelings or emotions, and the learners might simply not have had the abstract noun repertoire to do so, as PELT typically focuses on concrete noun vocabulary. This observation is supported by the poems by the Group A students, who had not been given word class restrictions for line 5 and who chose adjectives in 14 out of 17 cases (the other three being *love* (2) and *fun*), such as *funny, nice, cool*, or *wonderful* etc. – despite the fact that both model poems used nouns in line 5. It might thus be worthwhile for future *Elfchen* creation activities to change the noun requirement in line 5 in favor of adjectives or to abandon a word class requirement for line 5 altogether.

The remaining requirements for Group B concerned the two longest lines 3 and 4 and were purely semantic in nature (What is happening with the noun? – What can you feel, see, think…?). For line 3, the analysis revealed that only six children produced some sort of 'action' of the noun, with four of these constituting the 'x *is* V+*ing*' pattern presented in the model. Per contrast, the vast majority of 18 children produced a characteristic of the noun, either in a sentential 'x *is* adj' form (12, e.g. *cat is sweet, penguin is funny*), as a noun phrase (3, e.g. *king of animals*) or as a row of adjectives (3, e.g. *small big clever*). As for line 4, most children (21) produced a clause starting with *I like, I love* or *I find* and thus fulfilled the semantic requirement for this line. The remaining five poems used this line to present more characteristics of the topic noun. It can thus

be concluded that the requirement in line 3, which asked the learners to write about an action performed by the noun, proved to be most difficult for the learners to realize. Two reasons are conceivable for this observation: First, the learners' limited repertoire of action verbs, which are necessary to express actions and events. Second, the prevalence of the 'x *is* y' structure might have been caused by the model poem, which featured the present progressive structure *he is playing*. It cannot be assumed that the children were aware of the different syntactic status of *is* (copula vs. auxiliary); in addition, they may not yet have been able to perceive *playing* as a verbal component, especially given that German does not have a structural equivalent of the progressive. In combination, both reasons might hint at the lexical and grammatical challenges which the production of action descriptions using the present progressive poses for young beginning learners of English.

Taken together, the findings from both groups suggest that the learners generally handled the composition requirements quite well, especially in terms of the explicit formal requirements (word count, word distribution, word classes). Deviations occurred but remained minimal. In Group A, the learners implemented the implicitly provided sentential structure of lines 2 and 3 (subject + predicate + direct object), but failed to produce adverbial phrases in line 4. Whether this was due to the implicit nature of this requirement or the learners' restricted L2 proficiency cannot be judged on the basis of the data at hand but appears to be a worthwhile endeavor for future research. In Group B, the learners showed the greatest deviation in line 3, which required an action but was mostly realized by characterizations of the topic noun. In line 5, they exhibited a tendency to use concrete rather than abstract nouns. Both findings might be indicative of the learners' limited grammatical (line 3) and lexical (lines 3 and 5) repertoire, suggesting that these requirements might be more adequate for more proficient learners. A possible adaption for beginners might be to relax the noun requirement line 5 and to require line 3 contents that match the learners' linguistic repertoire more closely, e.g. location (Where is x? / Where does x live?).

(b) Lexical differences potentially caused by differences in scaffolding

The differences in poem composition were first analyzed by means of (v) the function word-to-content word (FW-CW) ratio, as the more explicit metalinguistic guidelines in Group B referred exclusively to content words (nouns, adjectives) and might thus have prompted a higher use of these, whereas the sentential models in Group A might have triggered a greater use of function words needed for clausal syntax (e.g. pronouns, articles etc.). The FW-CW ratio was calculated for each poem as the share of function words in the overall word number; those percentages were then compared using exploratory data analysis.

For Group A, the average use of function words was 30.2 %, i.e., roughly one third. In Group B, this average was considerably lower at 19.9 %, i.e. about one fifth, a difference which reached statistical significance ($p<.001$). This is quite remarkable given the small sample size, suggesting that the different composition guidelines did indeed have the hypothesized effect on the learners' choice of word categories. This conclusion is further supported by the scatter analysis of the FW percentages, which takes intra-sample variation and outliers into consideration. Figure 2 provides the respective boxplots, which clearly show that the groups cluster around different medians (Group A: M = 27.3; Group B: M = 18.2), while at the same time displaying comparable dispersion patterns, both in terms of overall range (Group A: 21.82; Group B: 28.03; no outliers) and interquartile range (IQR = 9.09 for both groups). What is more, the learners in Group B neither inquired about function words as frequently as the Group A students (as indicated by the *Words you need* poster, which contained only two function words: *on* and *or*), nor produced them from memory, as the poems attest. This indicates that they did not perceive the same need to employ function words as the learners in the first group. It can thus be concluded from the data that the way in which the *Elfchen* poems were explicitly and implicitly scaffolded in terms of composition criteria had an observable effect on the types of words (content or function) produced by the learners.

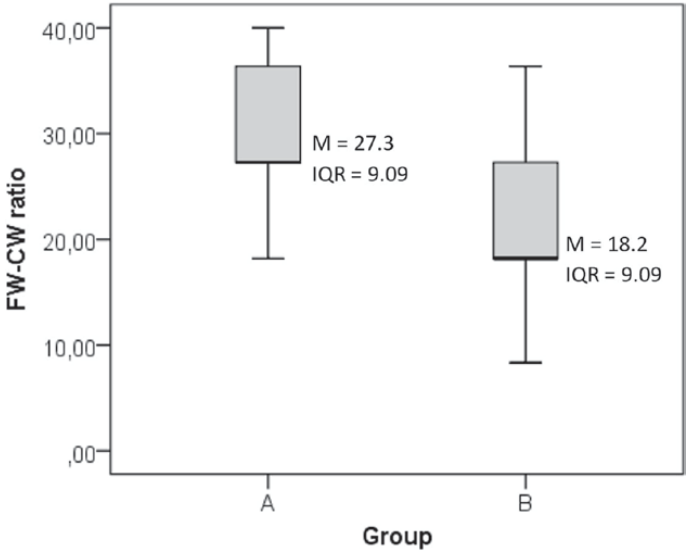


Fig. 2: Boxplots for function-to-content word (FW-CW) ratios

The second measure employed to analyze potential differences in poem composition was (vi) type-token ratio (TTR), i.e., the ratio of different lexemes (types) compared to the overall number of words produced (tokens). This was done based on the observation that many learners in Group B had used the same nouns repetitively in their poem, which suggests a low degree of lexical variation. The TTR analysis also revealed systematic differences between the groups. Those differences were in the expected direction, i.e., the Group A poems exhibited a higher TTR average ($\bar{x}$ = .91) than Group B ($\bar{x}$ = .84), which suggests that the learners in the first group used more different vocabulary items in their poems. Again, the difference was statistically significant (p = .007), which supports the assumption of a systematic rather than random difference in the lexical make-up of the poems. The scatter analysis in Figure 3 further corroborates this conclusion: A quarter of the poems in Group A used 11 different words in the poem, as indicated by the fourth quartile at 1.0. Per contrast, in Group B this was the exception rather than the norm (only two poems in this group exhibited a TTR of 1.0).

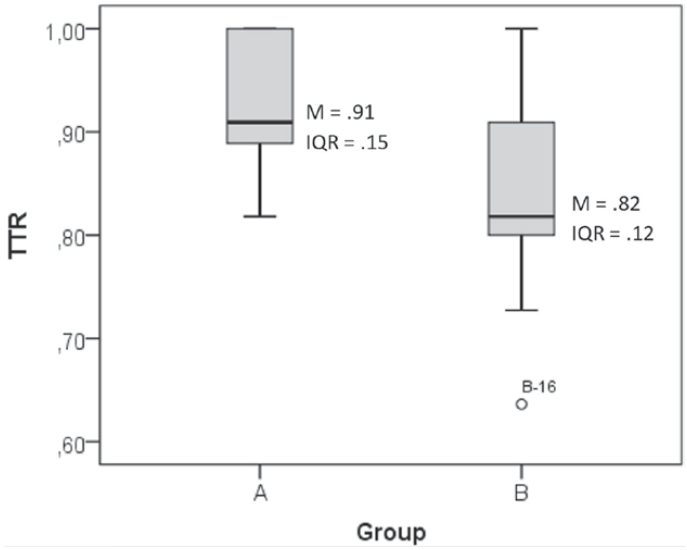


Fig. 3: Boxplots for type-token ratios (TTRs)

The conclusion that this difference was chiefly caused by the different composition criteria is further corroborated by L2 proficiency, which under equal circumstances would have caused the reverse pattern: According to the EFL teacher's evaluation, the second group was stronger orally than the first, both receptively and productively (cf. Tab. 1), which is usually interwoven with greater lexical skills (Kim et al. 2011: 517). In fact, some of the poems in Group B contain vocabulary that is not normally expected at primary level, such as *wildness*, *fluffy, pond*, or *shy*, none of which were contained in the dictionary nor provided by the teacher and thus must have been known to the students. The Group A poems do not contain such more advanced items. With a stronger lexical repertoire, the Group B learners were actually the more likely candidates for a higher, not a lower, TTR, which suggests that the restrictive word class requirements in Group B not only led to a lower use of function words (as indicated by the FW-CW ratio) but also prompted a higher degree of word repetition (as indicated by the TTR), mostly on the noun level. Accordingly, the stricter guidelines in Group B appeared to have worked against rather than in

favor of linguistic creativity, both on the level of syntactical and of lexical variation.

8.2 *Research question 2: Spelling mastery*

The second research question was devoted to spelling mastery. In light of the frequent word repetitions in Group B, all identical words within a poem (i.e., identical grammatical form and identical spelling) were removed from the analysis to avoid data distortion from correct self-copying. Same-type tokens featuring a different grammatical form (e.g. singular–plural) or different spellings (e.g., <dolphin> and <dolfin>) were, however, kept as these pose different spelling demands or indicate varying levels of spelling mastery, respectively. The remaining words were then analyzed by means of the above 6-point scale; a word featuring several issues was rated for the most severe error. Subsequently, the average spelling accuracy of each poem (SpellAcc) was determined as the arithmetic mean of all word ratings in that poem.

The analysis showed that the groups also differed systematically with regard to their spelling accuracy (p = .012). This time, however, the pattern was inverted: With a SpellAcc of 4.83, Group B outperformed Group A (4.48). The boxplots provided in Figure 4 illustrate the systematic nature of the differences between the groups. They further attest to the high spelling performance in the Group B learners, whose fourth quartile is identical with the highest possible value of 5.00 and whose interquartile range is only .27 (total range: .60). In Group A, the learners were more heterogeneous in their spelling, as indicated by the greater interquartile range of .70 (total range: 1.91), and only three learners achieved the perfect score.

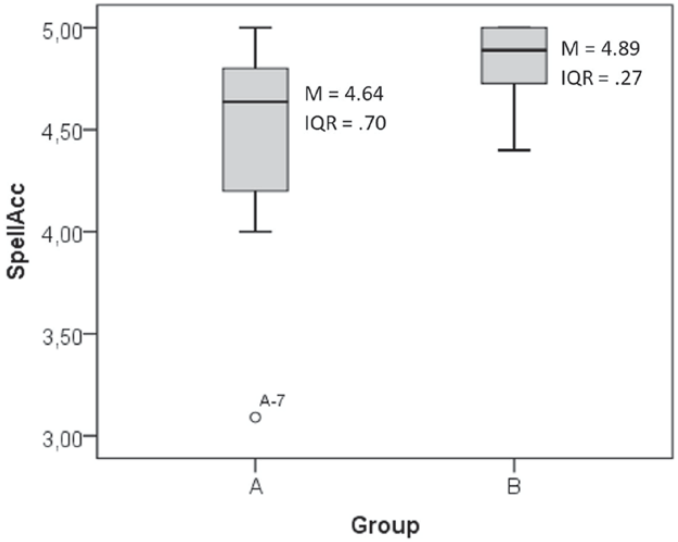


Fig. 4: Boxplots for Spelling Accuracy (SpellAcc)

A look at the error types helps understand the specific features of English spelling the learners struggled with. In Group A, out of the 172 words which were subjected to the spelling analysis, 38 were misspelled (22.1 %). In Group B, this concerned 28 of 248 words (11.3 %). The kinds of errors the groups exhibited are listed in Table 3 in ascending order of frequency.

Tab. 3: Types of spelling errors in the poems

Error type	Capitalization		Phonetic approximation		Double consonant		Similar L2 word		Apostrophe error		<th>-related error		Silent -e drop		German spelling		Incomprehensible		Misc		Sums
Example	<Head>		<faverit>		<funy>		<me> for *my*		<I'>		<lithle>		<hav>		<wunder voll>		<han>				
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*
Group A	12	30.0	5	12.5	6	10.5	4	10.0	2	5.0	3	7.5	1	2.5	1	2.5	3	7.5	3	7.5	40
Group B	19	61.3	2	6.5	0	0.0	2	6.5	2	6.5	0	0.0	2	6.5	2	6.5	0	0.0	2	6.5	31
All	31	43.7	7	9.9	6	8.5	6	8.5	4	5.6	3	4.2	3	4.2	3	4.2	3	4.2	5	7.0	71

As the table clearly shows, the biggest challenge the learners in both groups faced were capitalization errors. 24 of the recorded 31 instances concerned nouns and can thus be attributed to L1 transfer from German. The remaining instances were three adjectives (*big, strong, favorite*), the article *the* (three times), and the verb *like*. Although capitalization is a comparatively minor error as it is not related to insufficient letter knowledge and does not hamper word recognition, the results still suggest that learners would profit from capitalization instruction (esp. on the L1-L2 differences in noun capitalization) as it is still an aspect of correct orthography. The table further shows that this error type accounted for a much greater share of spelling issues in Group B than in Group A, suggesting that the Group A students struggled more with deep-structure issues of letter choice and sequencing, as indicated by the rather higher values for e.g. phonetic approximation, double consonant spellings (over- and underproduction) or *th*-related errors. In addition, all instances of words that were incomprehensible both in the poem and in the recording were produced in Group A. All in all, these more qualitative observations lend further support to the quantitative finding that the spelling accuracy in Group B was considerably higher than in Group A.

Two reasons might account for this difference. The first reason concerns a difference in the treatment, viz. the fact that the learners in Group B were provided with unknown words and their spelling, whereas this was not done for Group A. The second concerns the difference in proficiency level between the two groups, especially the more advanced lexical skills in Group B reported above. To investigate those factors further, the misspelled words were compared to the words scaffolded by the dictionary, the model poems, and those provided by the teacher (to keep up motivation in Group A, the teacher eventually decided to provide three items: *fur, sweet* and *very*). It soon transpired that the capitalizations of the topic nouns at the hypernym level might be attributable to the dictionary, whose section headings feature capital first letters and thus might explain the capitalization of the words *animals*, *pets*, *fruit*, and *jobs*. It does, however, neither explain the noun capitalization at the hyponym level (e.g., for *lion, turtle, watermelon*, or *policeman*) as these words are all presented with lower-case initial letters, nor all the instances of non-noun capitalization. Also, no teacher-provided word featured capitalization. Leaving

capitalization issues aside, of the remaining identifiable[3] 24 misspelled words in Group A, 21 were neither provided in the dictionary nor by the teacher, while three were (<tutle>, <verry>, <vervie>). In Group B, of the 11 remaining misspelled words, eight were scaffolded either by the dictionary (<policman>, <dolfin>, <mous>, <grin> for *green*) or the *Words you need* poster (2x <smoll>, <strang> for *strong*, <street> for *sweet*), which leaves only three non-scaffolded words (<I'>, <*for puppies*>, <wather> for *water*). We thus have instances of non-uptake of scaffolded words in both groups, but the high error rate of non-scaffolded words in Group A suggests that their low spelling accuracy can be attributed – at least partially – to the non-provision of words. In this sense, the learners in the first group were inadvertently put at a double disadvantage: Not only were they less proficient, but they also received a lower amount of lexical (and thus orthographic) scaffolding.

9. Summary and conclusion

The study presented here analyzed 43 *Elfchen* poems composed by young, beginning EFL learners in two 4th-grade classes in a German primary school, investigating the effects of different ways of scaffolding (models, composition criteria and word provision) on the adherence to the composition guidelines, on the poem's lexical make-up, and on the children's spelling proficiency. The analysis for RQ1 (adherence to composition requirements) showed that the learners handled the word count and line distribution requirements well. In addition, all learners followed the word class requirement for line 1, and the majority of learners in Group B also fulfilled the word class requirements for lines 2 and 5, suggesting that YLEs in year four of primary school can handle such metalinguistic requirements well, despite their beginning stage of L2 proficiency. The learners in Group A also succeeded, for the most part, in composing sentential patterns of the subject + predicate + direct object type as modeled in the sample poems on lines 2 and 3, but they failed to produce adverbial phrases on line 4 (or anywhere else), which was tentatively explained with their limited proficiency and the non-explicitness and thus non-saliency of

3 The three incomprehensible words are excluded here.

this requirement. In Group B, the learners handled the line 4 requirement of reporting about their impressions well by using *I like, I love*, or *I find* structures, but they struggled with the line 3 requirement of composing action statements, producing mainly characteristics instead. In addition, although fulfilling the structural criterion of producing a noun on line 5, the Group B children struggled to express summary-like, abstract concepts here and resorted to word repetition and hypernym/hyponym use instead. Taken together, the findings for RQ 1a thus suggest an interplay of proficiency and saliency of requirements and allow the following implications when working with *Elfchen* poems in PELT:

(i) The results indicate that *Elfchen* poems are a suitable and manageable text type for primary EFL learners. Accordingly, this kind of writing activity seems highly suitable for inclusion into young learners' EFL lessons.

(ii) Explicit (content) word class requirements are manageable, as are requirements to produce characteristics and personal impressions of the *I like*-type. Per contrast, the production of adverbial phrases appeared challenging, either due to the learners' limited proficiency or the non-saliency of the implicit requirement. The latter might be remedied by making this requirement explicit; the former by an increased focus on function words in PELT. This would not only benefit their *Elfchen* writing but also their overall L2 communicative ability (Mindt & Schlüter 2007: 78).

(iii) The production of abstract nouns as a summary in line 5 seems to be challenging for beginning YLEs; a relaxation of this requirement might help address this issue.

(iv) The production of action statements appears difficult as well, although it cannot be said with certainty whether these difficulties derived from lexical (action verbs) or grammatical challenges (structural similarity of the present progressive and the copula + adjective sequence). In any case, this finding suggests that learners might profit from an increased teaching focus in this area, to equip them with the means to fulfill this poem requirement and to further their general L2 skills.

The results for RQ2 showed that the differing requirements caused systematic differences in the lexical make-up of the poems, with the poems written in Group A featuring a higher share of function words as well as greater lexical diversity as indicated by a higher type-token ratio. On the one hand, those results were expected in light of the different composition guidelines, more specifically in light of the strict word class requirements in Group B as opposed to the absence of these in Group A (paired with the implicit model provision of sentential structures). On the other hand, these results were somewhat surprising in that they overrode proficiency effects. Accordingly, they point to the considerable influence of composition requirements, which in this case turned out to be more of a restriction than a true scaffold of creative writing. It might thus be advisable to refrain from word class requirements, especially as the learners' L2 proficiency increases. This is an important implication for the *process scaffolding* of creative poem writing tasks which is situated at the macro-level of *hard scaffolding*.

The final part of the analysis was devoted to the learners' spelling mastery. Two implications for the PELT classroom can be derived from the results: First, the capitalization of section headings can exacerbate the L1 transfer of noun capitalization in German YLEs. Second, creative writing activities of young learners require flexible and extensive lexical scaffolding, especially at the initial stages of proficiency. Firstly, this provides them with the necessary lexical support in expressing their intended thoughts, which is crucial for any writing task but especially for creative writing. Secondly, it offers the kind of spelling support which is necessary at this level. For instance, the CEFR descriptors of orthographic control at the A1 level are geared towards "copying" rather than "spelling", the latter of which is only reserved for the learners' "address, nationality and other personal details" (Council of Europe 2001: 118). Accordingly, these findings provide implications for the *output scaffolding* of creative poem writing tasks and concern both *hard* and *soft scaffolding* as lexical scaffolding can both be planned (e.g. by providing dictionaries) and spontaneously given by the teacher in line with the learners' individual needs.

Naturally, this study comes with a number of limitations. The first and most obvious of these is the small sample size, on the basis of which the

results and conclusions need to be treated with due caution. The second concerns the initial restriction on word provision in Group A, which was only relaxed after some students showed strong signs of demotivation and task abandonment. A closer documentation of the words the children had inquired about before the relaxation would have allowed more detailed insights into the children's intended word production.

Despite these shortcomings, the study contributes to the emerging research into young learners' foreign language literacy skills with an analysis of authentic writing samples created by YLEs. By linking the differences in scaffolding to the features of the final products, the study adopted a classroom-based approach which aims to inform L2 teaching on the grounds of evidence-based findings. Needless to say, more research is needed in this direction, analyzing bigger cohorts as well as a greater variety of proficiency levels and L1s, controlling and documenting the parameters of the intervention in even greater detail. As the present analysis has shown, putting a greater focusing on the emerging literacy of young learners, both in terms of research and teaching, is a worthwhile endeavor.

Bibliography

Aro, Mikko/ Wimmer, Heinz (2003), Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics* 24: 4, 621–635.

BIG-Kreis (2015), *Der Lernstand im Englischunterricht am Ende von Klasse 4. Ergebnisse der BIG-Studie*. München: Domino.

Bleyhl, Werner (1999), Empfehlungen zur Verwendung des Schriftlichen im Fremdsprachenerwerb auf der Grundschule. *PRIMAR* 8: 22, 45–52.

Brush, Thomas A./ Saye, John W. (2002), A summary of research exploring hard and soft scaffolding for teachers and students using a multimedia supported learning environment. *The Journal of Interactive Online Learning* 1: 2, 1–12.

Burmeister, Petra (2010), *Did you know that 15 difrent Fish arts in the Kiel Canal live? On foreign language writing in partial immersion primary school classrooms. In: Diehr, Bärbel & Rymarczyk, Jutta (eds.) (2010), *Researching Literacy in a Foreign Language among Primary School Learners*. Frankfurt/M.: Peter Lang, 131–146.

Burwitz-Melzer, Eva (2010), Lesen und Schreiben im Englischunterricht der Primarschule und des Übergangs (Klasse 3–6). In: Appel, Joachim/ Doff, Sabine/ Rymarczyk, Jutta & Thaler, Engelbert (eds.) (2010), *Foreign Language Teaching. History, Theory, Methods*. Berlin: Langenscheidt, 101–122.

Chujo, Kiyomi/ Nishigaki, Chikako (2005), Creating e-learning material to teach essential vocabulary for young EFL learners. In: Anthony, Laurence/ Fujita, Shinichi & Harada, Yasunari (eds.) (2005), *Proceedings of IWLeL 2004: An Interactive Workshop on Language e-Learning*. Tokyo: Waseda University, 35–44.

Council of Europe (2001), *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.

Diehr, Bärbel/ Frisch, Stefanie (2010), Lesen sie doch? – Fragen an die LiPs-Studie (Lesen im Englischunterricht auf der Primarstufe). *IMIS-Beiträge* 37, 143–163.

Duscha, Michael (2007), *Der Einfluss der Schrift auf das Fremdsprachenlernen in der Grundschule. Dargestellt am Beispiel des Englischunterrichts in Niedersachsen*. Doctoral dissertation, Technische Universität Carolo-Wilhelmina zu Braunschweig. [Online: https://publikationsserver.tu-braunschweig.de/receive/dbbs_mods_00021088. 24.08.2021].

Engel, Gaby/ Groot-Wilken, Bernd/ Thürmann, Elke (eds.) (2009), *Englisch in der Primarstufe – Chancen und Herausforderungen: Evaluation und Erfahrungen aus der Praxis*. Berlin: Cornelsen.

Frisch, Stefanie (2013), *Lesen im Englischunterricht der Grundschule. Eine Vergleichsstudie zur Wirksamkeit zweier Lehrverfahren*. Tübingen: Narr.

Genetsch, Martin/ Hallet, Wolfgang/ Surkamp, Carola/ Weisshaar, Harald (2012), Cheap Children. Individualisierung und Differenzierung mit einer Kompetenzaufgabe für Klasse 7 zu Kinderarbeit und Kinderrechten. In: Hallet, Wolfgang & Krämer, Ulrich (eds.) (2012), *Kompetenzaufgaben im Englischunterricht. Grundlagen und Unterrichtsbeispiele*. Seelze: Klett, 98–112.

Glaser, Karen (2018), Digitaler Mehrwert im Englischunterricht der Grundschule: Wortschatzerwerb mit dem TING-Hörstift.

In: Dausend, Henriette & Brandt, Birgit (eds.) (2018), *Lernen digital: Fachliche Lernprozesse im Elementar- und Primarbereich anregen*. Münster: Waxmann, 151–178.

Glaser, Karen/ Schrader, Anne (2019), The role of reading and writing in primary EFL curricula: An exploratory study in two Saxon primary schools. In: Falkenhagen, Charlotte/ Funk, Hermann/ Reinfried, Marcus & Volkmann, Laurenz (eds.) (2019), *Sprachen lernen integriert – global, regional, lokal. Dokumentation zum 27. Fremdsprachenkongress der DGFF, Jena, 27. - 30.09.2017*. Baltmannsweiler: Schneider Verlag Hohengehren, 388–402.

Griva, Eleni/ Tsakiridou, Helen/ Nihoritou, Ioanna (2009), A study of FL composing process and writing strategies employed by young learners. In: Nikolov, Marianne (ed.) (2009), *Early Learning of Modern Foreign Languages. Processes and Outcomes*. Bristol: Multilingual Matters, 132–148.

Hufeisen, Britta (2005), Schreiben in der Fremdsprache: Lehren und lernen in Gruppen mit unterschiedlichen Ausgangssprachen. In: Müller-Hartmann, Andreas & Schocker-von Ditfurth, Marita (eds.) (2005), *Aufgabenorientierung im Fremdsprachenunterricht. Task-Based Language Learning and Teaching: Festschrift für Michael K. Legutke*. Tübingen: Narr, 147–160.

Husfeldt, Vera/ Bader Lehman, Ursula (2009), *Englisch an der Primarschule. Lernstandserhebung im Kanton Aargau*. [Online: https://www.sprachenunterricht.ch/sites/default/files/E_Bericht_090108_0.pdf. 24.08.2021].

Kim, Young-Suk/ Al Otaiba, Stephanie/ Puranik, Cynthia/ Folsom, Jessica S./ Greulich, Luana/ Wagner, Richard K. (2011), Componential skills of beginning writing: An exploratory study. *Learning and Individual Differences* 21: 5, 517–525.

KMK [Kultusministerkonferenz] (2003), *Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Mittleren Schulanschluss*. [Online:https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2003/2003_12_04-BS-erste-Fremdsprache.pdf. 24.08.2021].

Kupetz, Rita (ed.) (1997), *Vom gelenkten und freien Schreiben im Fremdsprachenunterricht*. Frankfurt am Main: Europäischer Verlag der Wissenschaften.

Mindt, Dieter/ Schlüter, Norbert (2007), *Ergebnisorientierter Englischunterricht für die Klassen 3 und 4*. Berlin: Cornelsen Scriptor.

Müller-Wolfangel, Ute/ Pardall, Cornelia/ Prusse, Daniela (2012), *Mein erstes sprechendes Bildwörterbuch Englisch*. Mannheim: Dudenverlag.

Nation, Ian S. (2001), *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Pienemann, Manfred (2003), Language processing capacity. In: Daughty, Catherine & Long, Michael H. (eds.) (2003), *The Handbook of Second Language Acquisition*. Malden: Blackwell, 679–714.

Reckermann, Julia (2018), *Reading Authentic English Picture Books in the Primary School EFL Classroom*. Berlin: Peter Lang.

Reichart-Wallrabenstein, Mareike (2004), *Kinder und Schrift im Englischunterricht der Grundschule: Eine theorie- und empiriegeleitete Studie zur Diskussion um die Integration von Schriftlichkeit*. Berlin: Dissertation.de.

Rymarczyk, Jutta (2008), Früher oder später? Zur Einführung des Schriftbilds in der Grundschule. In: Böttger, Heiner (ed.) (2008), *Fortschritte im frühen Fremdsprachenlernern. Ausgewählte Tagungsbeiträge Nürnberg 2007*. München: Domino, 170–182.

Rymarczyk, Jutta (2010), Früher Schriftspracherwerb in der ersten Fremdsprache Englisch bei Kindern mit Migrationshintergrund. *Forum Sprache* 4, 60–78.

Rymarczyk, Jutta/ Musall, Annika (2010), Reading skills of first graders who learn to read and write in German and Englisch. In: Diehr, Bärbel & Rymarczyk, Jutta (eds.) (2010), *Researching Literacy in a Foreign Language among Primary School Learners*. Frankfurt/M.: Peter Lang, 69–88.

Sächsisches Staatsministerium für Kultus (2004/2009), *Lehrplan Grundschule Englisch*. [Online: http://lpdb.schule-sachsen.de/lpdb/web/downloads/2427_lp_gs_englisch_2019.pdf?v2. 24.08.2021].

Schmidt, Richard W. (2001), Attention. In: Robinson, Peter (ed.) (2001), *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press, 3–32.

Steitz-Kallenbach, Jörg (1998). Sich selbst in Sprache erfahren und finden. (Auto-) Biographisches und projektives Schreiben in der Sekundarstufe 1. *Praxis Deutsch,* 25: 152, 31–37.

Tangel, Darlene M./ Blachman, Benita A. (1992), Effect of phoneme awareness instruction on kindergarten children's invented spelling. *Journal of Reading Behavior* 24: 2, 233–261.

Wilden, Eva/ Porsch, Raphaela/ Ritter, Markus (2013), Je früher desto besser? Frühbeginnender Englischunterricht ab Klasse 1 oder 3 und seine Auswirkungen auf das Hör- und Leseverstehen. *Zeitschrift für Fremdsprachenforschung* 24: 2, 171–201.

Willing, Rebekka/ Geldschläger, Jakob (n.d.), *Elfchen*. [Online: https://wortwuchs.net/ elfchen/. 24.08.2021].

Wood, David/ Bruner, Jerome S./ Ross, Gail (1976), The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry* 17: 2, 89–100.

List of authors

Dr. des. Malgorzata Barras, University of Fribourg, Rue de Rome 1, 1700 Fribourg, Switzerland. Dept. of Multilingualism and Foreign Language Education.
malgorzata.barras@unifr.ch.
Malgorzata Barras is a researcher and teacher trainer at the University of Fribourg and at the Institute of Multilingualism in Fribourg, Switzerland. She specializes in language testing and strategy use in foreign language assessment, second language teaching, and qualitative research methodology. She also teaches in various advanced programs of professional teacher development across Switzerland.

Annette Becker, M.A., Karlsruhe University of Education, Bismarckstr. 10, 76133 Karlsruhe, Germany. Institute of Multilingualism
annette.becker@ph-karlsruhe.de.
Her current research interests include aspects of Second Language Acquisition, TEFL, Critical Discourse Analysis, Media Discourse and Political Discourse Analysis.

Prof. Dr. Carsten Breul, University of Wuppertal, Gaußstr. 20, 42119 Wuppertal, Germany. Dept. of English and American Studies.
breul@uni-wuppertal.de.
Carsten Breul's area of research is English linguistics and contrastive English-German linguistics. His main focus is on topics in the domain of semantics, pragmatics and information structure.

Prof. Dr. Bärbel Diehr, University of Wuppertal, Gaussstraße 20, 42119 Wuppertal, Germany. Dept. of English and American Studies.
diehr@uni-wuppertal.de.
Since her retirement in 2020 her research interests have focussed on critical discourse analysis and information literacy with a view to TEFL (Teaching English as a Foreign Language) as well as on bilingual education programmes and CLIL.

Prof. Dr. Stefanie Frisch, University of Wuppertal, Gaußstr. 20, 42119 Wuppertal, Germany. Dept. of English and American Studies.

frisch@uni-wuppertal.de.
Her research interests focus on teaching English as a foreign language to young learners. Her main focus is on the development of basic oracy and literacy skills, literary learning and bilingual education in primary school. She is also involved in EFL teacher education projects.

Jun.-Prof. Dr. Karen Glaser, University of Leipzig, Marschnerstr. 31, 04109 Leipzig, Germany. Dept. of Primary and Pre-Primary Education.
karen.glaser@uni-leipzig.de.
Her current research centers around Teaching English as a Foreign Language to young leaners in primary school with a special focus on literacy acquisition. In addition, she investigates the development of pragmatic skills in foreign language learners of all age groups and proficiency levels, and language teacher education with a focus on classroom interaction as well as the use of digital media.

Katharina Karges, University of Fribourg, Rue de Rome 1, 1700 Fribourg, Switzerland. Dept. of Multilingualism and Foreign Language Education.
katharina.karges@unifr.ch.
Her research interests focus on foreign language assessment, especially computer-based and large-scale assessments. She is also interested in questions of test validity and computer-assisted corpus linguistics and has participated in the development of several large-scale assessments. In her PhD project, she investigates item difficulty in a multilingual foreign language reading assessment.

Prof. Dr. Claudia Kastens, Wuppertal University, Gaußstr. 20, 42119 Wuppertal, School of Education, Dept. of Primary Education.
kastens@uni-wuppertal.de.
Her main research interests are on the development of motivation and learning of young learners and the role of instructional quality in various subjects. Currently she expanded this research in the development of motivation, well-being and learning of student teachers (Referendare) and the role of feedback and perceived support from educators.

Prof. Dr. Stefan D. Keller, School of Education, University of Applied Sciences and Arts Northwestern Switzerland, Hofackerstr. 30, CH-4132 Muttenz, Institute of Secondary Education.
stefan.keller@fhnw.ch.

His research interests include foreign language writing, feedback and assessment of language competences. A particular research interest is in the use of technology for assessing student writing and generating process-oriented feedback ("automated essay evaluation"). He often works in interdisciplinary projects with educational psychologists and computer linguists.

Prof. Dr. Annika Kolb, Freiburg University of Education, Kunzenweg 21, 79117 Freiburg, Germany. Dept. of English as a Foreign Language.
annika.kolb@ph-freiburg.de.
Her research interests include teaching English in the primary school and continuity in primary and secondary foreign language education. Furthermore, she researches the use of picturebooks and story apps in the foreign language classroom as well as foreign language teacher education within the blended-learning MA programme E-LINGO *Teaching English to Young Learners*.

Peter Lenz, Institute of Multilingualism, University of Fribourg and HEP Fribourg, Murtengasse 24, CH-1700 Fribourg, Switzerland.
Peter.Lenz@unifr.ch
Until 2020, Peter Lenz was a senior researcher, project leader, and teacher trainer in the fields of multilingualism and foreign-language education. He specializes in language testing, assessment, and curriculum development, and is the acting president of EALTA. In recent years, Peter participated in several large-scale studies making major contributions as a planner, task developer, consultant, and/or data analyst.

Dr. Carolyn Leslie, Nova University Lisbon, School of Social Sciences and Humanities, Avenida de Berna, 26C, 1069-061, Lisbon, Portugal. Modern Languages, Cultures and Literatures Dept.
cleslie@fcsh.unl.pt.
Her research interests focus on language learning through peer oral interaction in the primary English classroom, in particular the role of interaction patterns, individual differences and engagement. She also investigates approaches to assessment in primary language education and the professional development of teachers of English in the first cycle of compulsory education.

Dr. Gee Macrory, formerly Principal Lecturer, is currently a Visiting Research Fellow in the Faculty of Education at the Manchester Metropolitan University, UK.
g.macrory@mmu.ac.uk
Gee Macrory has extensive experience of language teacher education, in secondary and primary modern foreign languages, bilingualism and early language learning. Her most recent research is on the experiences of student teachers learning to teach foreign language literacy in primary and secondary classrooms, and young children's perceptions of foreign language orthographies.

Dr. Estela Monteiro, PhD in Sciences of Education, Minho University, 4700-057 Braga, Portugal.
epmonteiro@gmail.com.
Her current research interests focus on pedagogical supervision, teacher performance evaluation, and teacher development. In addition to this, she investigates the causes behind early school dropout, academic underachievement and approaches/practices used in assessment of primary English education in Portugal.

Dr. Maria Alfredo Moreira, University of Minho, Institute of Education, Campus de Gualtar, 4700-057 Braga, Portugal. Dept. of Integrated Studies of Literacy, Didactics, and Supervision.
malfredo@ie.uminho.pt.
Her current research interests focus on teacher development through narrative inquiry and action research with a focus on social justice. She is also interested in second and foreign language education, particularly culturally relevant pedagogies.

Dr. Sandie Mourão, Nova University Lisbon, Avenida de Berna 26C, 1069-061 Lisbon, Portugal. Centre for English, Translation and Anglo-Portuguese Studies.
sandiemourao@fcsh.unl.pt.
Her research focuses on early language learning (English as a Foreign Language), picturebooks in language education, developmentally appropriate practices and approaches to assessment. Her current research investigates teacher education for intercultural citizenship through picturebooks.

Jun.-Prof. Dr. Julia Reckermann, University of Münster, Johannisstr. 12-20, 48143 Münster, Germany. Dept. of English Language Education.
julia.reckermann@uni-muenster.de.
Her research interests focus on teaching English to young learners, particularly foreign language reading competences, as well as differentiation and individualisation in heterogeneous/inclusive classrooms. She additionally focuses on teachers' diagnostic competences and the teaching of English as a global/international language. Further interests circle around different aspects of CLIL.

Prof. Dr. Jutta Rymarczyk, Heidelberg University of Education, Keplerstraße 87, 69120 Heidelberg, Germany. Dept. of English as a Foreign Language.
rymarczy@ph-heidelberg.de.
Her current research interests focus on early biliteracy development and diagnostics (English and German). In addition to this, she investigates immersion programmes at primary and secondary level (*Bilingualer Sachfachunterricht/ CLIL*) and the role of visual literacy in this context. This is closely connected to the use of works of fine arts in teaching foreign languages, in the classroom and in museums.

Dr. des. Ruth Trüb, School of Education, University of Applied Sciences and Arts Northwestern Switzerland, Bahnhofstrasse 6, CH-5210 Windisch, Institute of Primary Education.
ruth.trueb@fhnw.ch.
She is a specialist in the teaching and learning of English in primary schools and her research interests include foreign language writing and teachers' assessment competencies. Her main concern is research that aims at having a direct impact on everyday classroom practice.

Dr. Karoline Wirbatz, TU Dortmund, Emil-Figge-Straße 50, 44221 Dortmund. Dept. of Applied Linguistics and ELT.
karoline.wirbatz@tu-dortmund.de.
Her research interests center around first, second & third language acquisition, bilingual education and classroom SLA. In addition to this, her work addresses the professional development of student-teachers.

Inquiries in Language Learning
Forschungen zu Psycholinguistik und Fremdsprachendidaktik

Edited by / herausgegeben von Christiane Bongartz / Jutta Rymarczyk

Vol. / Bd. 1 Bärbel Diehr /Jutta Rymarczyk (eds. / Hrsg.): Researching Literacy in a Foreign Language among Primary School Learners. Forschung zum Schrifterwerb in der Fremdsprache bei Grundschülern. 2010.

Vol. / Bd. 2 Alexander Heimes: *Psycholinguistic Thought Meets Sociocultural Theory*. Die integrativen Zusammenhänge von Fachmethodik und Fremdsprachenlernen im bilingualen (Geschichts-)Unterricht. 2011.

Vol. / Bd. 3 Markus Kötter /Jutta Rymarczyk (Hrsg.): Fremdsprachenunterricht in der Grundschule. Forschungsergebnisse und Vorschläge zu seiner weiteren Entwicklung. 2011.

Vol. / Bd. 4 Nicole Ogasa: Gefühle und Lernen im Fremdsprachenunterricht. Der Einfluss von Gefühlen auf das Lernen. 2011.

Vol. / Bd. 5 Henning Rossa: Mentale Prozesse beim Hörverstehen in der Fremdsprache. Eine Studie zur Validität der Messung sprachlicher Kompetenzen. 2012.

Vol. / Bd. 6 Udo Ohm / Christiane Bongartz (Hrsg.): Soziokulturelle und psycholinguistische Untersuchungen zum Zweitspracherwerb. Ansätze zur Verbindung zweier Forschungsparadigmen. 2012.

Vol. / Bd. 7 Bärbel Diehr / Lars Schmelter (Hrsg.): Bilingualen Unterricht weiterdenken. Programme, Positionen, Perspektiven. 2012.

Vol. / Bd. 8 Jutta Rymarczyk (ed.): Foreign Language Learning Outside School. Places to See, Learn and Enjoy. 2013.

Vol. / Bd. 9 Sabrina Bechler: Bilinguale Module in der Grundschule. Integriertes Inhalts- und Sprachlernen im Fächerverbund Mensch, Natur und Kultur. 2014.

Vol. / Bd. 10 Michał B. Paradowski (ed.): Teaching Languages off the Beaten Track. 2014.

Vol. / Bd. 11 Christiane Schöneberger: Grammar Growth in Child Second Language German. Investigating DP Development in an Immersion Setting. 2015.

Vol. / Bd. 12 Stéfanie Witzigmann / Jutta Rymarczyk (Hrsg.): Mehrsprachigkeit als Chance. Herausforderungen und Potentiale individueller und gesellschaftlicher Mehrsprachigkeit. 2015.

Vol. / Bd. 13 Martina Weitz: Die Rolle des L2-Inputs in bilingualen Kindergärten. 2015.

Vol. / Bd. 14 Markus Kötter / Jutta Rymarczyk (Hrsg.): Englischunterricht auf der Primarstufe. Neue Forschungen – weitere Entwicklungen. 2015.

Vol. / Bd. 15 Sylvie Méron-Minuth / Senem Özkul (Hrsg.): Fremde Sprachen lehren und lernen. Aktuelle Fragen und Forschungsaufgaben. 2015.

Vol. / Bd. 16 Sviatlana Karpava: Vulnerable Domains for Cross-Linguistic Influence in L2 Acquisition of Greek. 2015.

Vol. / Bd. 17 Christiane M. Bongartz / Andreas Rohde (Hrsg.): Inklusion im Englischunterricht. 2015.

Vol. / Bd. 18 Bärbel Diehr / Angelika Preisfeld / Lars Schmelter (Hrsg.): Bilingualen Unterricht weiterentwickeln und erforschen. 2016.

Vol. / Bd. 19 Jan Springob: Inklusiver Englischunterricht am Gymnasium. Evidenz aus der Schulpraxis im Spiegel von Spracherwerbstheorie und Fremdsprachendidaktik. 2017.

Vol. / Bd. 20 Alexander Kautzsch: The Attainment of an English Accent. British and American Features in Advanced German Learners. 2017.

Vol. / Bd. 21 Jaewon Nielbock-Yoon: A Case Study on Trilingual Siblings' Code Switching. Focus on Minority Language Development. 2017.

Vol. / Bd. 22 Stefanie Rottschäfer: Accent and Identity in Learner Varieties of English. A Study with German and French University Students in an English as a Lingua Franca Setting. 2018.

Vol. / Bd. 23 Almut Ketzer-Nöltge: Satzverarbeitung in der Erst- und Zweitsprache Deutsch. Gedächtnisleistung und kognitive Belastung. 2019.

Vol. / Bd. 24 Eva M. Knopp: From Bilingual to Biliterate: Secondary Discourse Abilities in Bilingual Children's Story Telling. Evidence from Greek Heritage Language Speakers in Germany and the United States. 2019.

Vol. / Bd. 25 Katharina Hirt: Genustransfer durch «Thinking for Speaking». Kognitive Muster und ihre Bedeutung für den DaF-Unterricht. 2019.

Vol. / Bd. 26 Holger Limberg / Karen Glaser (Hrsg.): Pragmatische Kompetenzen im schulischen Fremdsprachenunterricht. 2020.

Vol. / Bd. 27 Christiane Bongartz / Jacopo Torregrossa (eds.): What's in a Narrative? Variation in Storytelling at the Interface Between Language and Literacy. 2020.

Vol. / Bd. 28 Maria Andreou: Language and Cognitive Aspects of Child Bilingualism. Research Observations and Classroom Applications. 2021.

Vol. / Bd. 29 Sandra Bellet: Gesamtsprachliche Bildung und früher Englischunterricht. Professionalisierung von Lehramtsstudierenden im Kontext von Mehrsprachigkeit. 2022.

Vol. / Bd. 30 Stefanie Frisch / Jutta Rymarczyk (eds.): Current Research into Young Foreign Language Learners' Literacy Skills. 2021.

www.peterlang.com