



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Comparing and Evaluating Language Learning Mobile
Apps for Secondary Level EFL Learners“

verfasst von / submitted by

Raphael Anton Zitzenbacher, BEd

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Education (MEd)

Wien, 2023 / Vienna 2023

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Betreut von / Supervisor:

UA 199 507 511 02

Masterstudium Lehramt Sek (AB)

UF Englisch

UF Geschichte und Politische Bildung

Univ.-Prof. Mag. Dr. Christiane Dalton-Puffer

Acknowledgements

First and foremost, I want to express my appreciation to my thesis supervisor, Univ.-Prof. Mag. Dr. Christiane Dalton-Puffer, for her continuous support and expert advice throughout this entire research process.

Furthermore, conducting this project would not have been possible without the contributions of the peer reviewers. My sincere gratitude goes to Katharina, Julia, Victoria, Christian, and Christoph for their valuable and constructive feedback.

I would also like to extend my heartfelt thanks to my family for their boundless love, encouragement, and constant support throughout my educational pursuits.

My appreciation further extends to all my friends and my teacher colleagues. You have cheered me on and kept me sane during both highs and lows.

Finally, I would like to thank Paddy for proofreading this thesis. I am beyond grateful for your love and support.

Table of contents

1. Introduction	1
Part I: THEORETICAL FRAMEWORK	3
2. Mobile learning	3
2.1. Defining mobile learning	3
2.2. Benefits of mobile learning	5
2.3. Challenges of mobile learning	6
3. Mobile assisted language learning (MALL).....	8
3.1. Defining MALL	8
3.2. Benefits of MALL	9
3.3. Challenges of MALL	10
4. Language learning apps	11
4.1. A taxonomy of language learning apps	11
4.2. Evaluation rubrics for (language) learning apps	16
Part II: EMPIRICAL PART	25
5. Methodology	25
5.1. Research interests and research design	25
5.2. Developing an evaluation rubric for language learning apps	26
5.3. First draft	26
5.4. Peer review	28
5.5. Final draft	29
5.6. Interpreting the evaluation result	64
6. Analysis and Evaluation of language learning apps	66
6.1. Duolingo	66
6.1.1. Educational value	67
6.1.2. Design.....	70
6.1.3. Engagement	75
6.1.4. Language learning	78
6.1.5. Conclusion	91
6.2. Mondly	91
6.2.1. Educational value	92
6.2.2. Design.....	95
6.2.3. Engagement	99
6.2.4. Language learning	102

6.2.5. Conclusion	115
7. Discussion of evaluation results	117
7.1. Language learning apps in the EFL classroom	117
8. References	120
9. Appendix.....	133
9.1. Abstract English	133
9.2. Abstract German	134
9.3. Evaluation Rubric for LLAs for Secondary Level Learners.....	135
9.4. Transcripts of the think-aloud protocols and follow-up interviews	143

List of abbreviations

CALL	Computer assisted language learning
EFL	English as a Foreign Language
ML	Mobile learning
MALL	Mobile assisted language learning
LLA	Language learning app
L1	First language
TL	Target language

List of figures

Figure 1 Rosell-Aguilar's taxonomy of apps for language learning (2017: 249).....	12
Figure 2 Cherner, Dix & Lee's taxonomy of apps for language learning (2014).....	13
Figure 3 Examples of immediate feedback in Duolingo.....	68
Figure 4 Duolingo for Schools management layer for teachers.....	69
Figure 5 Elements contributing to ease of use and navigation in Duolingo	72
Figure 6 Tiers in the learning path of Duolingo's English course	73
Figure 7 Example of the Duolingo learning path for unit 1.....	74
Figure 8 Examples of diversity in Duolingo characters	75
Figure 9 Elements of gamification in Duolingo	77
Figure 10 Examples of Duolingo's translation activities.....	81
Figure 11 Examples of flawed translation activities.....	82
Figure 12 Examples of Duolingo's listening and speaking activities	83
Figure 13 Examples of Duolingo's language in use and dialogue activities	85
Figure 14 Examples of explicit grammar teaching methods in Duolingo.....	86
Figure 15 Examples of comprehension tasks in Duolingo's story activities	86
Figure 16 Examples of implicit statistical learning in Duolingo.....	88
Figure 17 Examples of immediate feedback in Mondly.....	93
Figure 18 Elements contributing to ease of use and navigation in Mondly.....	96
Figure 19 Examples of units accessible through Mondly's content map	98
Figure 20 Examples of representation of diverse populations in Mondly	99
Figure 21 Elements of gamification in Mondly	102
Figure 22 Examples of Mondly's translation activities and verb lists.....	107
Figure 23 Examples of Mondly's language in use activities	108
Figure 24 Examples of Mondly's vocabulary and speaking activities.....	109
Figure 25 Examples of Mondly's vocabulary lessons and conversation lessons	110
Figure 26 Mondly's chatbot	111
Figure 27 Lessons and conversation practice in Mondly AR.....	113

List of tables

Table 1 Dimension (A1) <i>Thinking Skills</i> (Lee & Cherner 2015, adapted)	29
Table 2 Dimension (A2) <i>21st Century Skills</i> (Lee & Cherner 2015, adapted).....	30
Table 3 Dimension (A3) <i>Value of Errors</i> (Lee & Cherner 2015, adapted).....	31
Table 4 Dimension (A4) <i>Feedback to Teacher</i> (Lee & Cherner 2015, adapted)	33
Table 5 Dimension (A5) <i>Cooperative Learning</i> (Lee & Cherner 2015, adapted)	34
Table 6 Dimension (A6) <i>Accomodation of Individual Differences</i> (Lee & Cherner 2015, adapted).....	35
Table 7 Dimension (B1) <i>Ability to Save Progress</i> (Lee & Cherner 2015, adapted)	37
Table 8 Dimension (B2) <i>Platform Integration</i> (Lee & Cherner 2015, adapted).....	38
Table 9 Dimension (B3) <i>Screen Design</i> (Lee & Cherner 2015, adapted)	39
Table 10 Dimension (B4) <i>Ease of Use</i> (Lee & Cherner 2015, adapted)	40
Table 11 Dimension (B5) <i>Navigation</i> (Lee & Cherner 2015, adapted).....	41
Table 12 Dimension (B6) <i>Information Presentation</i> (Lee & Cherner 2015, adapted)	42
Table 13 Dimension (B7) <i>Media Integration</i> (Lee & Cherner 2015, adapted)	43
Table 14 Dimension (B8) <i>Cultural Sensitivity</i> (Lee & Cherner 2015, adapted).....	44
Table 15 Dimension (C1) <i>Learner Control</i> (Lee & Cherner 2015, adapted).....	46
Table 16 Dimension (C2) <i>Interactivity</i> (Lee & Cherner 2015, adapted)	48
Table 17 Dimension (C3) <i>Pace</i> (Lee & Cherner 2015, adapted)	49
Table 18 Dimension (C4) <i>Personal Preferences</i> (Lee & Cherner 2015, adapted)	50
Table 19 Dimension (C5) <i>Interest</i> (Lee & Cherner 2015, adapted)	51
Table 20 Dimension (C6) <i>Aesthetics</i> (Lee & Cherner 2015, adapted)	52
Table 21 Dimension (C7) <i>Gamification</i>	54
Table 22 Dimension (D1) <i>Quality of Input</i>	55
Table 23 Dimension (D2) <i>Opportunity to Produce Output</i>	58
Table 24 Dimension (D3) <i>Level of Language Learning Material</i>	59
Table 25 Dimension (D4) <i>Skills</i>	61
Table 26 Dimension (D5) <i>Exposure to the Target Language</i>	62
Table 27 Types of activities in Duolingo.....	80
Table 28 Evaluation results of Duolingo	89
Table 29 Types of activities in Mondly	105
Table 30 Evaluation results of Mondly.....	115

1. Introduction

"How absurd that our students tuck their cell phones, BlackBerrys, iPads, and iPods into their backpacks when they enter a classroom and pull out a tattered textbook." – Eli Broad

With the rapid advancement of technology in recent decades, there has been a notable shift of interest towards digital platforms and mobile applications in language learning (Burston 2015: 4). This interest in new technologies is paralleled by the growing smartphone ownership among children and teenagers. A survey conducted in Germany in 2021 revealed that 86% of ten to twelve-year-olds, 95% of thirteen to fifteen-year-olds, and 96% of sixteen to eighteen-year-olds possess a smartphone (Statista 2022a). Similar percentages of smartphone ownership can be anticipated among Austrian children and teenagers. For these young learners the emergence of language learning apps has provided unprecedented opportunities to engage in language learning independent of spatial, temporal, and educational constraints. Notably, the COVID-19 pandemic further propelled the prominence of digital platforms and apps in language education, as educators worldwide sought alternative means to deliver instruction in the context of distance teaching and learning. Although the restrictions imposed by COVID-19 have now eased, mobile language learning apps continue to occupy a central role in the field of language education.

Upon browsing the app stores of the two major operating systems, namely Android (Google) and iOS (Apple), one is overwhelmed by the plethora of language learning apps readily available for download. In 2022, both app stores offered about 455,000 educational apps (Wylie 2023), with a considerable portion of this selection consisting of language learning apps. This great number of available apps presents language educators and learners with a delicate issue, as they are tasked with the challenge of discerning which language learning apps are of high quality and provide genuine benefits for language acquisition. As Kukulska-Hulme (2013:3) pointedly observes

[t]here is no shortage of language-learning resources that can be accessed on the Web and downloaded to mobile devices, including innumerable podcasts and mobile apps. In this scenario of great abundance, finding one's way to the most appropriate, high-quality materials and resources to suit personal language-learning requirements is difficult.

This issue serves as the foundation of the present thesis, which tries to address the gap in research regarding the evaluation of language learning apps. While numerous studies in

recent years have explored the effects of implementing language learning apps in language education (e.g. Ajisoko 2020, Armas Pesantez et al. 2019, Castañeda & Cho 2016, Jiang et al. 2021, Loewen et al. 2019, and others), only one study evaluated language learning apps on the basis of a specifically devised evaluation rubric (Chen 2016). However, as a result of the ever-changing and fast developing nature of language learning apps, Chen's evaluation has become outdated. Furthermore, the introduced evaluation rubric shows some weaknesses in terms of extent and precision. Throughout the past decade, several other evaluation rubrics have emerged; however, as discussed in section 4.2., these rubrics are either aimed at all types of educational apps or tailored towards specific types of language learning apps.

Therefore, this thesis has two main objectives. Firstly, it aims to develop an evaluation rubric specifically designed for language learning apps, which can facilitate a systematic and comprehensive evaluation. Secondly, it seeks to apply this rubric in the evaluation of two popular language learning apps, namely Duolingo and Mondly, to provide insights into their design and mechanics. Lastly, these findings will be utilised to examine potential implementations and constraints associated with the integration of language learning apps in the EFL classroom.

The present thesis is divided into two main parts and structured as follows. The first main part establishes the theoretical background. It is composed of three sections focusing on mobile learning, mobile assisted language learning, and language learning apps respectively. First, a conceptualisation of mobile learning (ML) and mobile assisted language learning (MALL) will be attempted, followed by a comparison of the possibilities and challenges for second/foreign language learning. Subsequently, language learning apps will be discussed. In particular, different taxonomies for categorising language learning apps will be introduced. This will be followed by a discussion and comparison of several evaluation rubrics for (language) learning apps. The second main part of this thesis, the empirical part, will build on these theoretical underpinnings. First, the research interest and research design will be introduced. This will be followed by a section devoted to the development of the evaluation rubric for language learning apps. Within this section, the final draft of the evaluation rubric and the underlying theory will be discussed in detail. Following this, an analysis and evaluation of two language learning apps, Duolingo and Mondly, will be conducted. In the concluding section of part II, implications for implementing LLAs in the EFL classroom will be outlined.

Part I: THEORETICAL FRAMEWORK

2. Mobile learning

2.1. Defining mobile learning

Mobile learning (ML) is a complex concept. With the rise of mobile technology, "a community of practice" has emerged, actively involved "in the development and delivery of mobile learning", thereby gaining "experience and expertise" in this domain (Traxler 2007: 1). However, although mobile devices play an increasingly significant role in educational contexts and are almost ubiquitous in today's society, mobile learning is not a defined concept and is still evolving amongst its growing community. Kukulska-Hulme's and Shield's argument that "[m]obile learning is undergoing rapid evolution" (2008: 273) is still true today and is accompanied by a discourse that tries to yield an agreed-upon definition of the concept.

The definitions of mobile learning and the emphasised aspects of mobility vary depending on the researchers' individual perspectives. Winters, who states the chameleon-like character of the concept, argues for "a re-conceptualisation of the precise nature of mobile learning" (2006: 5). He summarises the perspectives that researchers tend to apply when they write and talk about mobile learning into four categories.

Firstly, the technocentric perspective, which predominates the literature, defines mobile learning as a form of learning that is delivered mainly through mobile electronic devices, such as smartphones or tablets (Winters 2006: 5). Such constrained definitions are rejected by researchers such as McQuiggan et al. (2015: 8), who claim that mobile learning cannot simply be equated with the use of physical mobile devices for learning purposes. Moreover, Traxler suggests that definitions should "look at the underlying learner experience and ask how mobile learning differs from other forms of education, especially other forms of e-Learning" (2007: 4).

The second perspective understands mobile learning "as an extension of e-learning" (Winters 2006: 5). This leads to all-encompassing definitions that do not consider the particular nature of mobile learning. Furthermore, approaching the concept from this viewpoint categorises mobile learning as a subset of e-learning, solely adding the factor of portability (Winters 2006: 5).

The third perspective perceives mobile learning as an augmentation of formal education in a classroom setting, since it breaks the boundaries of a face-to-face lecture. However, as formal education can also be conducted in the form of distance learning, highlighted by the COVID-19 pandemic, mobile learning should not only be characterised by its opposition to a traditional classroom setting, but to all forms of traditional learning (Winters 2006: 5-6).

Finally, researchers including Sharples, Taylor, and O'Malley, who in their early work propose a technocentric perspective, now turn away from the portability of the device and towards the mobility of the learner. This learner-centered perspective results in definitions that understand mobile learning in two ways: it allows mobile learning to be defined as learning that happens wherever the learner is (not site-specific) as well as learning that uses mobile technologies (Winters 2006: 6).

In contrast to the four categories introduced by Winters, an alternative approach to conceptualising mobile learning is provided by El-Hussein and Cronje, who define mobile learning according to three aspects of mobility: the mobility of technology, the mobility of learning, and the mobility of the learner (2010: 16-20).

Given the complexity of the topic and the plethora of wide-ranging definitions, for the purpose of this study, the concept of mobile learning shall be defined as follows:

Mobile learning refers to formal or informal learning that can occur anywhere at any time mediated and enhanced through mobile handheld devices, such as smartphones or tablets.

This working definition combines multiple elements of the various approaches to defining mobile learning. Firstly, it does not view mobile learning as conflicting with learning in the traditional classroom setting. Instead, mobile learning can happen within the boundaries of the language classroom in a more formal setting as well as outside of the language classroom in a presumably more informal settings. Secondly, it is learner-centered, as it emphasises the learners' ability to learn without being confined to spatial locations and temporal boundaries. Finally, it also covers the technocentric perspective on mobile learning, as the definition highlights how mobile handheld devices play a crucial role in mediating and improving the (mobile) learning experience. By incorporating these different elements, the definition seeks to synthesise varying perspectives in order to arrive at a relatively precise conceptualisation of mobile learning that remains adaptable to future technological developments.

2.2. Benefits of mobile learning

Mobile learning claims to provide numerous benefits for individual learners and educational institutions. One of the most striking advantages of mobile learning is the potential to learn irrespective of temporal boundaries or spatial location. Mobile learning can happen anywhere and at any time, thus breaking the spatial walls and the constraining time frame of traditional formal education in a classroom (McQuiggan et al. 2015: 10, El-Hussein & Cronje 2010: 18, Chinnery 2006: 13). This independence enables learners "to exploit short amounts of time and space for learning" (Traxler 2007: 8). Additionally, mobile learning is able "to support context-specific and immediate learning" and can connect learners even across huge distances (e.g. during distance learning) (Traxler 2007: 7).

Furthermore, mobile devices provide the opportunity of leaving the four walls of the classroom and exploring digital learning spaces, such as "virtual museums, online classes, and simulated experiences" (McQuiggan et al. 2015: 11). This sometimes brings learning beyond the control of the teacher, which might be perceived as a threat, but could also be used as a starting point to revitalise and rethink current teaching and learning practices (Kukulski-Hulme 2013: 4-5). In addition to being effectively used to enhance more traditional classroom settings, mobile devices can also facilitate alternative educational setups, such as flipped classrooms or blended learning environments (McQuiggan et al. 2015: 12).

Moreover, mobile learning allows for a more personalised learning experience (McQuiggan et al. 2015: 12, El-Hussein & Cronje 2010: 19, Traxler 2007: 7). Traxler, for example, defines personalised learning as "learning that recognises diversity, difference, and individuality in the ways that learning is developed, delivered, and supported" (2007: 7). As far as social, cognitive and physical variation amongst learners and their preferences with regard to learning styles and approaches are concerned, mobile learning also considers these factors (Traxler 2007: 7). From the perspective of the teacher, personalised mobile learning environments can enable them to track the development and progress of their learners more easily, which helps them to provide well-adjusted materials and assign homework and coursework that caters to the learner's individual needs (McQuiggan et al. 2015: 12).

Alongside the personalised aspect of learning, mobile learning can increase learners' motivation through "high levels of engagement [...], personalization and autonomy"

(McQuiggan et al. 2015: 12). By engaging in mobile learning, learners might experience a sense of independence and freedom (El-Hussein & Cronje 2010: 19) while they engage in interesting, relevant, and authentic tasks, which ideally reflect meaningful real-world problems (Traxler 2007: 7-8). Discovering new ways of utilising the devices and available materials ensures enduring learner motivation. As smartphone ownership and use is high amongst children and teenagers nowadays (Statista 2022a), employing mobile learning makes it possible to bring the learner's real life context to the classroom (McQuiggan et al. 2015: 12).

In addition, proponents assert that mobile learning and the use of mobile devices can facilitate the improvement of learners' higher order thinking skills, such as critical thinking and problem solving, communication, collaboration, creativity, and innovation. Sharing information, using resources and materials creatively, and evaluating the value, authenticity, and accuracy of available sources represent some of the activities inherent in the use of mobile devices and mobile learning in the educational context (McQuiggan et al. 2015: 11-12).

Finally, mobile technology can often reach underserved children and schools in much more effective ways than alternative technological initiatives. Tablets are usually available at lower cost than regular computers and can be updated quite easily. Portable Mobile devices thus stand out as reasonable alternatives to stationary computers (McQuiggan et al. 2015: 11, Chinnery 2006: 13).

2.3. Challenges of mobile learning

In contrast to the benefits, mobile learning also presents challenges. Firstly, access to mobile devices and the internet is an issue. Although smartphone ownership and internet access seem to be universal for huge parts of society, many children and teenagers from underprivileged economic backgrounds do not possess a personal device or have only limited or no access to the internet. This impairs their chance of engaging in mobile learning, so schools and the state should try to provide the necessary equipment (McQuiggan et al. 2015: 13-14).

Furthermore, learners should ideally be monitored while using mobile devices in formal and informal learning settings. Despite the beneficial possibilities that smartphones with access to the internet provide, it is necessary to control whether learners use devices for the

purpose of distraction, or to engage in any unethical behaviour, or if they unintentionally reveal too much of their personal data online. To anticipate some of these digital dangers, it is necessary to teach digital literacy to all learners in order to ensure that they use devices and consume content responsibly (McQuiggan et al. 2015: 14).

The effective incorporation of mobile learning is also influenced by attitudes and prejudice towards digital tools. The ways in which lessons nowadays are organised still frequently uphold traditional structures of education. Incorporating mobile learning effectively into lessons is a cultural shift that entails breaking down some traditional structures, which is something that stakeholders can be hesitant to agree to. Although much anecdotal evidence and many empirical studies (e.g. Persson & Nouri 2018, Mihaylova et al. 2022, Tommerdahl, Dragonflame & Olsen 2022, Sung, Chang & Liu 2016) emphasise the power of mobile learning, no established and accepted theory of mobile learning has been developed yet. Consequently, mobile learning receives a mixed perception amongst teachers, stakeholders and researchers with regard to this paradigm. Bans on smartphones in schools and school policies impeding the use of smartphones during lessons show the bias certain educational institutions hold and further impede effective employment of mobile learning (McQuiggan et al. 2015: 14-16).

Other disadvantages of mobile learning revolve around practical issues and the physical properties of mobile handheld devices. Smaller screens and no external keyboards can make them less comfortable to use, and limited power, limited quality of visuals and audio as well as internet connection issues pose other drawbacks (Chinnery 2006: 13). Sharing devices among groups, e.g. in the form of a device cart with 25 tablets that is shared among different classes, is sometimes arranged due to tight budgets but is not ideal, since students can benefit most from mobile learning when they can engage with their own device. However, sharing devices is better than no devices, even though overall ubiquity should be the aim (McQuiggan et al. 2015: 15-16).

Finally, the implementation of mobile learning in the educational context can easily impede its effectiveness and benefits. It is not effective to simply add a mobile device to the curriculum in exchange for an older tool. Mobile learning and mobile devices should be used to change the ways lessons are structured in order to engage learners in new ways. The goal is to create a mobile learning plan first, and then obtain the device, not the other way around (McQuiggan et al. 2015: 16).

3. Mobile assisted language learning (MALL)

3.1. Defining MALL

Mobile assisted language learning (MALL) is "one of the key application areas" within the field of mobile learning (Kukulska-Hulme 2013: 6). Furthermore, MALL is considered to be a successor to computer assisted language learning (CALL). For Kukulska-Hulme and Shield (2008: 273) "MALL differs from computer assisted language learning in its use of personal, portable devices that enable new ways of learning emphasising continuity or spontaneity of access and interaction across different contexts of use". Given the complexity of mobile learning, the concept of MALL can be defined in various ways and "seems to be all things to all people" (Díaz-Vera 2012: xiii). Although Díaz-Vera claims that "the unique nature of MALL is becoming very difficult to characterize" (2012: xiii), most definitions and conceptualisations highlight two aspects: the use of mobile devices for the purpose of language learning and the mobility of the learner (Viberg & Grönlund 2013: 73).

Viberg and Grönlund highlight the use of mobile devices when they suggest that "[m]obile assisted language learning refers to mobile technology use for learning language" (2013: 73). In contrast, Palalas (2011: 76-77) emphasises the aspect of mobility, stating that "MALL can be defined as language learning enabled by the mobility of the learner and [...] portability of handheld devices". Combining both aspects, Kukulska-Hulme defines MALL by the "use of smartphones and other mobile technologies in language learning, especially in situations where portability and situated learning offer specific advantages" (2020: 1). For the purpose of this study, I will adopt this definition for the concept of MALL.

In addition to the lack of a generally accepted definition of MALL, another issue arises out of the lack of clarity with regard to the use of the terms *mobile learning* and *mobile assisted language learning* that is observable in some of the research in the field. Sometimes used interchangeably, it is not always clear which concept is referred to, which eventually feeds into the issue of unclear conceptualisations. Thus, keeping the terms mobile learning and MALL delimited in the best possible manner should be the aim, even though conceptually they will always partly overlap.

3.2. Benefits of MALL

Since MALL represents a subset of mobile learning, many advantages of MALL are congruent with the benefits of mobile learning described in section 2.2. Nevertheless, the literature also points out various benefits that mobile devices in particular can have in the context of language learning.

Rapid technical developments over the last two decades have led to sophisticated mobile devices such as smartphones and tablets, which are able to perform most functions previously confined to stationary and expensive desktop and laptop computers (Kukulska-Hulme 2013: 1). In combination with the usually lower costs of mobile devices, this explains why smartphones and tablets are nowadays wide-spread amongst society and ubiquitous in all aspects of life (Kukulska-Hulme 2013: 1). Given the devices' inherent aspect of mobility, learners using smartphones or tablets with Internet access can therefore promptly access help and information, as well as retrieve language learning materials immediately. Moreover, mobile devices allow learners to use time and space more flexibly (Kukulska-Hulme 2020: 1).

This flexibility supports language practice while simultaneously conducting mundane actions and allows for a smoother continuation of learning between different spatial locations. It therefore enables location-specific and location-independent language learning. The former manifests itself in mobile applications that connect language learning to the learners' environment and surroundings (e.g. by means of access to location-specific language material), while the latter refers to language learning that does not depend on any predetermined location and can happen on the go, granted that the mobile device in use has either instant access to online learning resources or is equipped with language learning software (Kukulska-Hulme 2013: 2-3).

Another benefit of mobile devices in language learning is the possibility of participating in language activities individually or collaboratively. The latter incorporates discussion and discourse, which can occur in a synchronous or asynchronous setting by means of using speech, written text, or rich multi-media such as videos or pictures (Kim & Kwon 2012: 35, Traxler 2007: 7).

Regarding language acquisition and the practice of the four skills, MALL could provide several benefits. Focusing on the skills of reading and listening, it seems plausible that

increased engagement with authentic samples of language is possible due to technological advances and mobile device ownership. E-book software, online newspapers, or other textual news channels render reading in another language an attractive activity, for example during long, daily commutes. Similarly, listening to podcasts or other audio recordings, and watching Youtube videos on a regular basis in various situational contexts is also made possible through mobile device ownership and contributes to practising said skills. Additionally, dictionaries and tools for translation can provide real-time assistance and aid language comprehension (Kukulska-Hulme 2020: 2-3). With regard to speaking skills, mobile devices can allow for more privacy than stationary computers or less mobile laptops and might therefore benefit learners. Eventually, as far as written production is concerned, restricted input (e.g. character limit in various apps) and the physical properties of mobile devices, such as small screens and keyboards, pose a challenge for extensive writing. However, using mobile devices as tools that enable collaborative annotation of texts in class, for example by enabling students to add comments to a given text, shows how MALL can even positively contribute to writing practice (Kukulska-Hulme 2020: 3).

3.3. Challenges of MALL

There are also disadvantages to MALL that require discussion. Firstly, access to suitable devices and the internet is not universal and could be limited to certain groups. Furthermore, using mobile devices for teaching and learning purposes could potentially exclude learners with cognitive or physical impairments (e.g. visual impairments or dyslexia), as they might experience difficulties reading text from or interacting with others on a small screen. To assist these learners, listening materials could be offered instead, such as digital audiobooks, podcasts, or other recordings (Kukulska-Hulme 2013: 5).

The proliferation of mobile devices is advantageous, however, this omnipresence of mobile devices also poses a challenge for the conventional, traditionally structured modes of education, as it tends to obscure the boundaries between domains of daily life, entertainment, work, and learning (Kukulska-Hulme 2020: 1, 4-5).

4. Language learning apps

Prior to the trailblazing launch of Apple's first *iPhone* in 2007, the use of mobile phones for learning purposes was predominantly restricted by physical hardware and pre-installed software. This however changed with the introduction of smartphones and their new operating systems (Apple's *iOS* and Google's *Android OS*) in the late 2000s. Today, device-native software can be complemented and enhanced with additional software applications, simply known as apps or mobile apps, which are available for download in the app stores of different operating systems.

Educational apps for mobile devices have been increasingly developed in recent years. At the moment, the Google Play Store (Android) offers about 2.59 million apps (Statista 2023a), whereas the App Store (iOS) offers about 1.64 million apps (Statista 2022b), many of which promise to facilitate language learning. While countless educational apps can be immediately downloaded and used free of charge, some apps must be purchased first. However, the majority of these applications are available for purchase at a justifiable price. Additionally, some developers offer lite versions of their apps, which allows users to test the product before they acquire the full version. Certain mobile applications also offer the option of in-app purchases, which allows users to access additional content or remove advertisements.

4.1. A taxonomy of language learning apps

In light of the abundance of language learning apps (LLAs), several attempts to establish neat and clear-cut classifications have been made. Establishing taxonomies has a long tradition in education, as classifications provide us with a useful common language and frame of reference. Specifically, as Rosell-Aguilar describes, establishing a taxonomy of educational apps enables "learners, teachers, and researchers to conceptualize and visualize the different varieties of apps available, which in turn can help them to evaluate their potential" (2017: 248). Taking into consideration the different interests and perspectives of researchers, these classifications approach educational apps, and in particular LLAs, from a variety of angles.

In 2017 Rosell-Aguilar proposed a taxonomy of LLAs, which consists of three main categories: apps designed for language learning, apps not designed for language learning but

useful to language learners, and dictionaries and translation apps. This classification is shown in Figure 1 below.

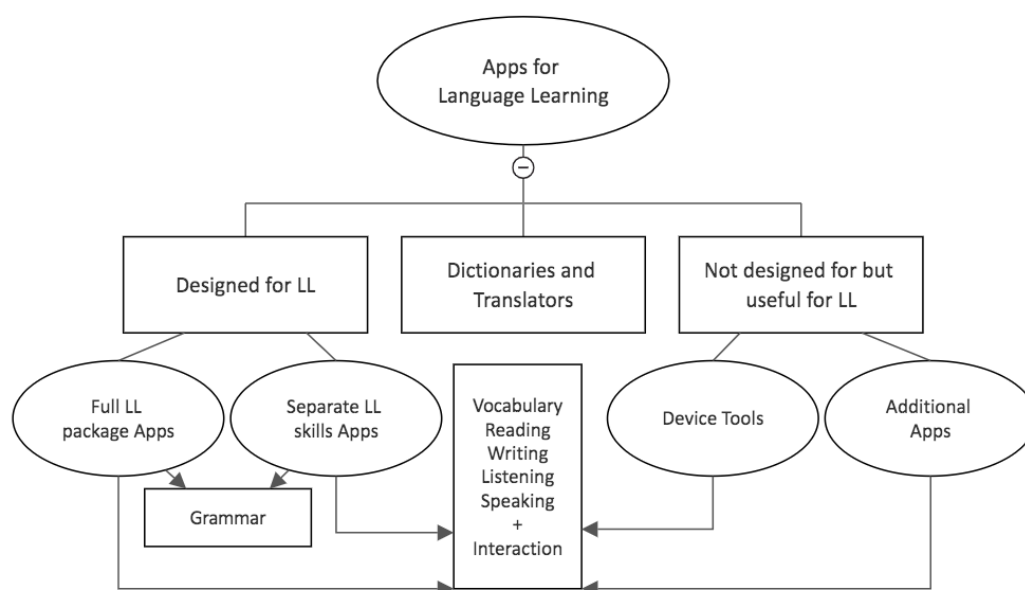


Figure 1 Rosell-Aguilar's taxonomy of apps for language learning (2017: 249)

The first category, apps specifically designed for language learning, comprises two subgroups. Firstly, there are apps which are intended to provide "full language learning solutions", and which involve "a variety of exercises, grammatical explanations, and interaction with other students and native speakers as well as support from communities of learners" (Rosell-Aguilar 2017: 248). Examples of such apps are *Duolingo*, *Mondly*, *Babbel*, or *Rosetta Stone*. The second subgroup of apps for language learning comprises apps which provide opportunities and exercises to improve and develop certain areas of language, such as vocabulary, grammar, reading, listening, speaking, and writing (Rosell-Aguilar 2017: 249).

The second category in this taxonomy of LLAs refers to apps which are not specifically designed for language learning but can still be used for language learning purposes. Again this category can be divided into two subgroups, namely tools which are pre-installed on a device and apps which are additionally installed at a later time. Pre-installed tools, such as web browsers, speech-to-text tools, messaging or communication tools, photo and video cameras, or satellite navigation tools can be utilised by learners to facilitate their language learning process in various areas. Moreover, language settings and multilingual text input can be employed in a similar fashion. Additional apps which are not device-native but might have value for language learning purposes are, for example, apps providing flashcard packages (*Quizlet* or *Memrise*), apps comprising reading materials in the target language,

blogging or microblogging apps that can be used for practising writing (*Twitter, Wordpress*), music and video streaming apps that can be used to practise listening skills (*Youtube, Spotify, Netflix*), voice recording apps to facilitate speaking skills, and social media platform apps that focus on social interaction and communication (*Instagram, Facebook, Skype, Whatsapp*). Finally, information sources such as news apps, apps providing maps and geography-related information, and games played in the target language can also be useful for language learners and teachers and belong to the second subgroup (Rosell-Aguilar 2017: 250).

The third category Rosell-Aguilar (2017: 251) proposes in his taxonomy encompasses dictionaries and translation apps. While in some cases dictionaries might be integrated in other apps, e.g. e-book readers, dictionaries as such often come in the form of separate apps that might also provide pronunciation examples (*Leo Wörterbuch, dict.cc Wörterbuch*). Translation apps, whose algorithms have improved in recent years, offer translations by entering written text or speech (*Google Translate, iTranslate*) and are, despite language teachers' criticism, the first place language learners frequently resort to when facing the task of writing.

In contrast to Rosell-Aguilar's taxonomy which categorises language learning apps according to their intended purpose in terms of language learning, Cherner, Dix and Lee's (2014) taxonomy of educational apps rather focuses on apps' functions. According to their taxonomy, educational apps can be divided into four main categories, namely skill-based apps, content-based apps, function-based apps, and educational misfits, as is shown in Figure 2 below.

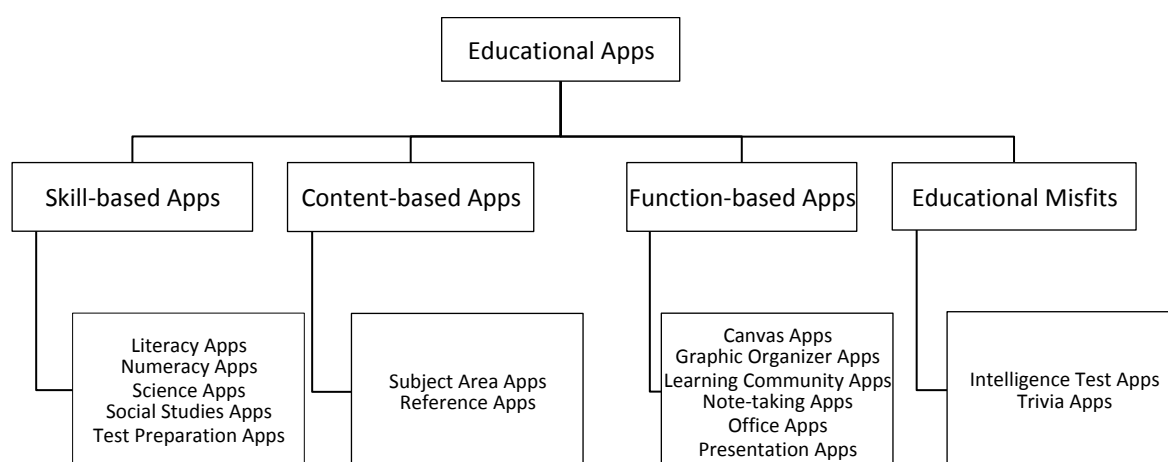


Figure 2 Cherner, Dix & Lee's taxonomy of apps for language learning (2014)

The first category, skill-based apps, encompasses 5 subgroups, namely literacy apps, numeracy apps, science apps, social studies apps, and test preparation apps. According to Cherner, Dix and Lee, "skill-based apps use recall, rote memorization, and skill-and-drill instructional strategies to build students' literacy abilities, numeracy skills, standardized test readiness, and subject area knowledge" (2014: 174). Since they mainly require learners to employ lower order thinking skills, they can be categorised as instructive apps. Nevertheless, by means of facilitating learners' basal skills, they create a valuable basis for prospective learning demands raised through the use of manipulative apps (content-based) and constructive apps (function-based). In the context of this study, the subgroup of literacy apps plays a significant role, as such apps contribute towards the development of learners' fluency, grammatical knowledge, handwriting skills, spelling, vocabulary knowledge, reading skills, and language proficiency overall (Cherner, Dix & Lee 2014: 174).

The second subgroup, namely content-based apps, offers the possibility to explore and utilise enormous portions of information and data. Learners can tap this reservoir of knowledge by means of self-directed enquiry (reference apps) or exploration of content that has been preprogrammed on apps (subject area apps). However, as mentioned before, learners must possess foundational literacy and numeracy skills in order to be able to access and properly process the exploited subject. In contrast to skill-based apps, content-based apps do not demand completion of a learning task but rather enable "students to read texts, view images, and watch videos related to specific topics, similar to walking from exhibit to exhibit in a museum" (Cherner, Dix & Lee 2014: 175-176.). Since such apps allow learners a certain amount of independence with regard to their choice of investigating information in a pre-established setting (this refers to the app's design), they can be called manipulable apps. Based on their design, content-based apps can function either top-down (limited to preprogrammed content) or bottom-up (ability to conduct own searches). Since content-based apps demand learners to assess the value of the explored information before synthesising it, they evoke and foster learners' higher order thinking skills (Cherner, Dix & Lee 2014: 176).

The third category, function-based apps, offers learners the possibility to employ their skills and knowledge acquired through the use of skill-based and content-based apps in order to create learning products, so-called learning artifacts. Such artifacts include texts, visuals, multimedia products, or a combination of these. Function-based apps can also be referred to

as constructive apps, since learners can deploy their literacy and numeracy skills developed through skill-based apps to transform their knowledge gained from content-based apps into learning products using flexible templates. Function-based apps can also be divided into subgroups, which reflect the various types of learning artifacts that can be generated. While using such apps learners create, evaluate, and synthesise, which reflects the highest order thinking skills. Moreover, they also address the development of *21st Century Skills* in learners, by allowing them to produce and distribute digital learning materials. (Cherner, Dix & Lee 2014: 176-177).

The last category introduced by Cherner, Dix and Lee comprises, what they call, educational misfits. This category encompasses apps that neither contribute to learners' development of literacy or numeracy skills nor provide them with the opportunity to explore valuable content. Moreover, they also do not enable students to create meaningful learning artifacts. Instead, these types of apps provide learners with the possibility to interact with content that has very limited value with regard to learning aims. Examples of such apps are trivia quizzes or intelligence tests (Cherner, Dix & Lee 2014: 177).

Applying Cherner, Dix and Lee's taxonomy in the English foreign language (EFL) classroom, teachers need to consider several things. To begin with, skill-based apps focused on developing skills require learners to simply retrieve and recall information that has been learned before in order to carry out tasks. As a result, they are considered less complex in Bloom's (1956) taxonomy since they mainly target lower-level cognitive abilities.

Nevertheless, learners receive immediate feedback when using such apps, which enables students to trace their individual learning progress. Content-based apps, which train higher order thinking skills, in turn, are characterised by a lack of feedback on how learners utilise information. Thus, it is the teacher's responsibility to instruct learners how to effectively use content-based apps and evaluate information gathered through independent research. Similar to content-based apps, function-based apps do not offer learners any feedback on the character and condition of their created artifacts. This means that ample time needs to be devoted to demonstrate learners how to use such apps effectively, for example, by introducing high quality model products. Providing too little support in this regard would hinder students from successfully creating and sharing learning artifacts (Cherner, Dix & Lee 2014: 175-177).

To summarise, the two taxonomies introduced are characterised by their different approaches towards classifying learning apps. While Rosell-Aguilar's (2017) classification draws its categories based on a very clear understanding of an app's purpose with regard to language learning, Cherner, Dix and Lee's (2014) classification divides apps according to their functions for learning purposes. Drawing on both taxonomies of LLAs, language learning apps can be characterised and categorised more clearly and effectively. Applying the taxonomies in the context of this study, skill-based LLAs specifically designed for language learning purposes will be compared and evaluated.

4.2. Evaluation rubrics for (language) learning apps

The rapid emergence of LLAs and their increasing adoption by teachers and researchers as pedagogical tools to supplement conventional classroom instruction have generated a surge of interest in the evaluation of these educational apps through empirical research studies. Many of these studies introduced and proposed evaluation rubrics, which should help teachers to compare and assess an app's quality and usability, in order to make decisions easier.

In this section, I shall turn towards app evaluation studies and discuss their proposed app evaluation rubrics. It is important to note that the studies did not necessarily evaluate language learning apps in order to be considered for inclusion in this review, since proposed evaluation rubrics aiming at educational apps overall still deserve to be discussed due to their quality and impact.

In the 1980s, computer-based educational software became increasingly popular, leading to the development of several evaluation frameworks by educators and researchers. Reeves and Harmon's (1994) *Systematic Evaluation of Computer-Based Education* framework, highlighted by Lee and Cherner (2015: 22), which consists of 14 pedagogical evaluative dimensions and 10 user-interface evaluative dimensions, was a prominent framework that has inspired other evaluation frameworks and has been frequently adapted.

Confronted with the extensive, yet growing offer of educational apps today, researchers and teachers can no longer rely upon Reeves and Harmon's (1994) original framework, since it is no longer sufficient for the evaluation of contemporary educational mobile software (Lee & Cherner 2015: 22-23). Firstly, some of the evaluative dimensions in Reeves and Harmon's (1994) framework, e.g. *overall functionality*, *origin of motivation*, and *mapping*, are too

vague and broad, making them difficult to measure. Secondly, the rubric lacks headings and clear indicators which should provide ratings for each evaluative dimension. Instead of providing such valuable indicators, which would support a proper evaluation of contemporary educational software, Reeves and Harmon's (1994) framework just offers brief descriptions of features that can be evaluated. Thirdly, the framework does not consider several functionalities carried by contemporary software, such as ability sharing capability, cross-platform integration, and the ability to save progress. Moreover, Reeves and Harmon's framework does not reference other frameworks such as Webb's (1997) *Depth of Knowledge* or *21st Century Skills*. While Reeves and Harmon's (1994) framework provides a strong foundation, the current post-PC era calls for a revision and adaptation of the framework to meet the standards of evaluating educational apps.

The advent of mobile technology, and in particular the introduction of mobile apps in the educational sector, has opened up the need for new evaluative frameworks, which are adjusted to contemporary affordances of mobile technology and help to assess the value of apps for learning purposes. In the last decade, several studies have proposed such evaluative frameworks. To outline the state-of-the-art in the field of educational app evaluation, the following section will introduce and discuss selected studies and their proposed app evaluation rubrics in greater detail.

Starting in 2011, Walker proposed one of the first (most likely the first) evaluation rubrics for educational apps on *iPods* in response to the already thriving app market of that time. Walker's rubric comprises six evaluative dimensions, namely (1) *Curriculum Connection*, (2) *Authenticity*, (3) *Feedback*, (4) *Differentiation*, (5) *User Friendliness*, and (6) *Student Motivation*. In order to be able to assess educational apps with regard to these proposed dimensions, Walker provides four indicator descriptors for each dimension, which constitute a four-point Likert scale. Using Walker's pioneering rubric as a template and foundation, Buckler (2012) developed another app evaluation framework. In contrast to Walker, her rubric is specifically geared to (educational) apps for adults with special needs. However, Buckler's framework also encompasses six evaluative dimensions, namely (1) *Application*, (2) *Feedback*, (3) *Adjustability*, (4) *Ease of Use*, (5) *Cost*, and (6) *Benefits*. In line with Walker's original 2011 rubric, Buckler's framework provides four indicator descriptors per evaluative dimension, which in turn create a four-point Likert scale.

Despite their novelty at the time of their creation and the resulting impact, both rubrics display several significant shortcomings with regard to their design. As Lee and Cherner (2015: 23-24) strikingly point out, both Walker (2011) and Buckler (2012) neglected to disclose explanations about the underlying theories and the previous research that influenced and shaped both rubrics in their studies. As a result, it is quite difficult to draw conclusions about the validity of the evaluative instruments. Furthermore, both rubrics are designed in ways that rather facilitate evaluations of educational apps on the surface level. This issue, for example, presents itself in Walker's (2011) rubric in the evaluative dimension (6) *student motivation*. While the indicator descriptors for this particular dimension provide four statements about the overall motivation of users evoked by the app, the rubric fails to differentiate between various factors that might significantly influence student motivation, such as the app's content, the pace with which the content is presented to the user, and the level of simplicity or difficulty experienced when using the app (Lee & Cherner 2015: 23). Another issue in Walker and Buckler's rubrics are the limited scopes of particular terms. Lee and Cherner (2015: 23) explicitly mention Walker's restricted definition and reductionist view of the term *differentiation*. Walker (2011: 61) refers to differentiation as "the ability to set the level of difficulty or target specific skills for individual students". However, as Lee and Cherner (2015: 23) rightfully point out, the notion of differentiation cannot only be limited to an app's ability to set the difficulty level of the presented content or taught skills, but should also consider differentiation in terms of "learners's background knowledge, learning style, and interest" (2015: 23). Moreover, Walker and Buckler's choice of aligning their created dimension to a four-point Likert scale format limits evaluations conducted with the rubrics, as a traditional five-point Likert scale would allow for a more differentiated assessment of the qualities and characteristics of selected apps (Lee & Cherner 2015: 24). Finally, limitations also exist with regard to the rubrics' range of comprehensiveness, as each only comprises six evaluative dimensions. These six dimensions are rather broadly defined and comprise factors and aspects that should ideally be independent evaluative domains (e.g. *app design*, *instructional value*, *potential to engage users*). This in turn limits the informative value of educational app evaluations, as conclusions about an app's assets or flaws cannot be drawn as finely nuanced as they should.

Turning towards a rather special field of language learning, namely early literacy education, Israelson (2015) proposed a two-part evaluative framework encompassing another

evaluation rubric that comprises four evaluative dimension, namely (1) *Multimodal Features*, (2) *Literacy Content*, (3) *Intuitiveness of App Navigation*, and (4) *User Interactivity*. Similar to the two rubrics discussed above, four descriptors are assigned to each evaluative dimension in this particular rubric, again providing the rater with a four-point Likert scale when rating apps that should facilitate multimodal early literacy learning. The issue of comprehensiveness, pointed out in the criticism of the two rubrics above, also partly applies to Israelson's rubric. While the rubric is merely oriented towards evaluating a very particular type of language learning app, which should be viewed rather positively as it enables raters to really focus on the particular characteristics and features such apps should hold, its dimensions and descriptors appear to be characterised by a certain shallowness that only allow for a surface-level evaluation of said apps. In addition, Israelson does not advise her readers on how to interpret or use final scores after an app has been evaluated, which is another issue that also applies to Walker's (2011) and Buckler's (2012) studies. Even though Walker (2011: 61-62) attempts to provide some instructions on what score separates valuable apps that facilitate learning from apps that simply do not (he suggests that good educational apps must score four points in the categories (1) *Curriculum Connections*, (5) *User Friendliness*, and (6) *Student Motivation*, the reader is left wondering how to meaningfully interpret rating scores.

The earlier mentioned issue of comprehensiveness with regard to the number of evaluative dimensions does not apply to Lee and Cherner's (2015) evaluation rubric for educational apps, which comprises 24 evaluative dimensions categorised into three domains, namely (A) *Instruction*, (B) *Design*, and (C) *Engagement*. Criticising Walker (2011) and Buckler (2012) for their lack of a theoretical anchorage in their rubrics, Lee and Cherner (2015) try to base their dimensions on theoretical frameworks, such as Vygotsky's (1978) *Zone of Proximal Development* or Webb's (1997) *Depth of Knowledge*. The rubric's 24 dimensions follow a precise format, which "includes a prompt that focuses the dimension on a central question, and five indicator descriptors that describe the ways in which an app's functionality or design may behave in response to a prompt" (Lee & Cherner 2015: 25). The five indicator descriptors of each dimension are aligned to a traditional five-point Likert scale, which allows for the establishment of quantitative measures in a more differentiated manner. While allocating a five on the Likert scale reflects the high quality of an app and allocating a one reflects the low quality of an app, the rating scale for each dimension also provides a *not*

applicable (N/A) option, which comes into effect when an app, by virtue of its design, cannot be assessed on a particular dimension (Lee & Cherner 2015: 25-26).

Diving deeper into the rubric's design, we shall now discuss its three domains in more detail. According to Lee and Cherner (2015: 26) the first domain, (A) *Instruction*, refers to "[t]he instructional worth of an app", with its underlying dimensions intended to "measure an app's educational value". This educational value of apps manifests in their "analysis of the cognitive demands placed on learners and the support offered to them by an app as they work to meet its learning objective" (Lee & Cherner 2015: 26). The eight dimensions which are used to assess an educational app's instructional worth comprise (A1) *Rigor*, (A2) *21st Century Skills*, (A3) *Connections to Future Learning*, (A4) *Value of Errors*, (A5) *Feedback to Teacher*, (A6) *Level of Material*, (A7) *Cooperative Learning*, and (A8) *Accommodation of Individual Differences* (Cherner & Lee 2015: 26-29).

Besides an app's instructional worth, the aspect of functionality is also important. This aspect is covered by the rubric's second domain (B) *Design*, which "assesses the functionality of an app's design, and functionality used in this context is defined as an analysis of how learners interact with an app" (Lee & Cherner 2015: 30). By measuring an app's performance in terms of its (B1) *Ability to Save Progress*, (B2) *Platform Integration*, (B3) *Screen Design*, (B4) *Ease of Use*, (B5) *Navigation*, (B6) *Goal Orientation*, (B7) *Information Presentation*, (B8) *Media Integration*, and (B9) *Cultural Sensitivity*, raters aim to evaluate an app's functional quality, which is, for example, expressed in a smooth and efficient interaction between user and app (Lee & Cherner 2015: 30).

The third and final domain (C) *Engagement* proposed by Lee & Cherner (2015) is concerned with an app's ability to motivate users to engage with and learn from the offered learning tasks. An app's appeal to learners' motivation is measured in terms of seven dimensions including (C1) *Learner Control*, (C2) *Interactivity*, (C3) *Pace*, (C4) *Personal Preferences*, (C5) *Interest*, (C6) *Aesthetics*, and (C7) *Utility*. These dimensions, however, only reflect a few representative reasons why an app might or might not motivate and engage users, as user motivation depends on a variety of context-specific factors that lie beyond the scope of this domain and the evaluation rubric in general (Lee & Cherner 2015: 33).

In contrast to the other evaluation rubrics introduced so far, Lee and Cherner's (2015) initially devised rubric was subjected to a critical examination by experts in the course of

their study in order to verify face and content validity of the evaluation instrument. The two expert groups were composed of university faculty members teaching at the tertiary level who held a doctoral degree in the field of Educational Technology, Educational Psychology, or Literacy as well as public school teachers and graduate students holding a teaching license with at least five years of teaching experience. These two expert groups were invited to comment on and critically rate the 24 dimensions by means of an assessment form. The feedback led to the modification of the initial instrument, after which the rubric was presented to the expert groups for a final verdict, which attributed the finalised version of the evaluation rubric with increased clarity in terms of its evaluative dimensions (Lee & Cherner 2014: 24). One may argue that such an extensive rubric is too comprehensive for practical use. The N/A option is useful as some apps might not be assessable on certain dimensions, however, an N/A option can also be problematic as users might find it difficult to decide whether a lack of information within an app should be rated as N/A or as a flaw (Papadakis, Kalogiannakis and Zaranis 2017: 3159).

Unlike Lee and Cherner's (2015) universal evaluation rubric for educational apps, Ok et al. (2016) devised an evaluation form that focuses on apps for a very particular target group, namely instructional apps for students with learning disabilities. In their study, the researchers introduce a three-step evaluation which is based on a three-part evaluation form. In the first step, basic information concerning the app is identified. In the second step, the app's qualities are evaluated, and in the last step, the app is graded. The information gathered in step one comprises details concerning (1) *the app's name and publisher*, (2) *the content area*, (3) *the objectives*, (4) *the content level*, (5) *the graphic/theme level* (i.e. the learner level), and (6) *the type of application* (Ok et al. 2016: 245-248). Step two in the evaluation process is based on an evaluation rubric which provides raters with an assessment instrument that measures app's for learners with disabilities in terms of 13 dimensions, which include (1) *Objective*, (2) *Strategy*, (3) *Examples*, (4) *Practice*, (5) *Error Correction and Feedback*, (6) *Error Analysis*, (7) *Progress Monitoring*, (8) *Motivation*, (9) *Navigation*, (10) *Visual and Auditory Stimuli*, (11) *Font*, (12) *Customized Settings*, (13) *Content Error and Bias*. These 13 dimensions fall within the scope of three overarching domains, namely (A) *Instructional Elements*, (B) *Personal Factors*, and (C) *Environmental Factors*. Every dimension comprises three indicator descriptors which are aligned to a three-point Likert scale (Ok et al. 2016: 247-251). In order to arrive at a grade, the scores for each

dimension are added, then divided by the total number of possible points, and multiplied by 100 in order to arrive at a percentage. Apps are then graded on a five-part scale (A-F), which reflects traditional grading systems in the educational sector (Ok et al. 2016: 251). According to Ok et al. (2016: 245) the face and content validity of the evaluation rubric was established through five professionals from the field of instructional technology and disability, who examined each dimension in terms of wording and relevance, which provided the researchers with feedback to improve the evaluation form.

Despite the evaluation form's overall clear structure and comprehensibility, the included rubric and its evaluative dimensions hold a few issues. As discussed earlier, one major problem is once again the apparent vagueness of evaluative dimension (8) *motivation*. While the three indicator descriptors provided in the rubric deal with learner engagement in general, the dimension fails to consider which features or aspects of an app motivate and engage learners. Therefore, scores from this dimension carry little meaning and do not help to define which app features promote learner's motivation in particular. Moreover, the three-point Likert scale of Ok et al.'s (2016) rubric does not allow for a differentiated assessment compared to a five-point Likert scale used in several other rubrics.

Looking at a different audience of learners, Papadakis, Kalogiannakis and Zaranis (2017) created an evaluation rubric assessing educational apps for preschool children. Their comprehensive evaluation instrument comprises 18 evaluative dimensions which can be split into four "key areas" (i.e. domains), namely (A) *Educational Content*, (B) *Design*, (C) *Functionality*, and (D) *Technical Characteristics* (Papadakis, Kalogiannakis and Zaranis 2017: 3154). Domain (A) *Educational Content* covers seven evaluative dimensions, which are (A1) *Knowledge Package Appropriateness*, (A2) *Learning Provision*, (A3) *Levelling*, (A4) *Motivation/Engagement*, (A5) *Error Correction/Feedback Provision*, (A6) *Progress Monitoring/Sharing*, and (A7) *Bias Free*. Domain (B) *Design* comprises four dimensions, namely (B1) *Graphics*, (B2) *Sound*, (B3) *Layout/Scenery*, and (B4) *App/Menu Design*. Domain (C) *Functionality* is divided into another four dimensions, which are (C1) *Child-friendliness*, (C2) *Autonomy*, (C3) *Instructions*, and (C4) *Configuration Ability*. Finally, domain (d) *Technical Characteristics* includes the last three dimension, namely (D1) *Performance and Reliability*, (D2) *Advertisement/Electronic Transactions*, and (C3) *Social Interactions* (Papadakis, Kalogiannakis & Zaranis 2017: 3160-3162).

In line with Lee and Cherner (2015) as well as Ok et al. (2016), Papadakis, Kalogiannakis, and Zaranis (2017) tried to verify the face and content validity of their rubric by involving two experts ("experienced instructional technology experts") in the creation process, whose feedback in terms of wording and relevance of the examined dimensions was incorporated. Furthermore, the researchers also conducted statistical analysis of the rubric's internal consistency, convergent validity, and inter-rater reliability. As a result of the small sample of apps evaluated, the statistical analysis does not provide results that can be generalised, even though it offers an optimistic outlook on the usability of the rubric (Papadakis, Kalogiannakis & Zaranis 2017: 3156-3158).

Two limitations of the rubric are directly pointed out in the study itself. First, the evaluative dimensions lack an N/A option (see Lee & Cherner 2015), which would be useful in situations where an app, by virtue of its design, cannot be assessed on a particular dimension, and second, as a result of its comprehensiveness, one must critically question the rubric's deployability outside an educational or research environment (Papadakis, Kalogiannakis & Zaranis 2017: 3159). Moreover, the rubric's evaluative dimension (A4) *Motivation/Engagement* carries the same inherent issue described in the context of Walker (2011) and Ok et al. (2016). This particular dimension is characterised by a vague formulation of indicator descriptors with regard to the app's potential to motivate its users, whereby the different aspects and factors that contribute to learners' motivation and engagement are overlooked. Additionally, the wording of the indicator descriptors implies that a rater would have to observe a child while using the app in order to be able to score dimension (A4) *Motivation/Engagement*. Lastly, dimensions (C1) *Child-friendliness* and (C3) *Instructions Existence* seem to overlap thematically in terms of their indicator descriptors. In this regard, in order to avoid a lack of clarity for raters, the rubric would have to be revised.

Not a fully-fledged evaluation rubric, but a short and helpful "app feature checklist" for reviewing and designing LLAs in particular was proposed by Sweeney and Moore (2012: 13). This checklist comprises seven bullet points, which in their wording remind one of rubrics and their evaluative dimensions. Using this checklist for the assessment of LLAs, educators and researchers would measure LLAs in terms of their (1) *potential for personalization*, (2) *visible indication of progress*, (3) *coverage of relevant language*, (4) *coverage of more than one skill*, (5) *maximization of exposure to the target language*, (6) *appropriateness for the device with regard to content, activities, and user interface*, and (7) *encouragement of*

learning behaviours which correspond to what we know about general mobile-enabled behaviour patterns (including social and gamification aspects) (Sweeney & Moore 2012: 13).

Chen (2016) developed an evaluation rubric specifically crafted for the purpose of assessing language learning apps. This rubric encompasses seven distinct categories that serve as criteria for evaluating these apps. The seven categories include (1) *Content Quality*, (2) *Pedagogical Coherence*, (3) *Feedback and Self-correction*, (4) *Motivation*, (5) *Usability*, (6) *Customization*, and (7) *Sharing*. Within each category, the rubric provides a guiding statement that outlines the desired characteristics and behaviours an app should ideally possess. Chen's rubric adopts a rating scale consisting of three bands, each representing a different level of suitability. The bands are labeled as *least suitable*, *average*, and *most suitable*. Each band encompasses a range of points that can be assigned for the corresponding category. Apps deemed to be least suitable may receive a score ranging from 1 to 3 points, while apps of average quality fall within the 4 to 7 point range. Apps considered to be most suitable can obtain a score ranging from 8 to 10 points. To assist evaluators in making accurate assessments, each band is accompanied by an indicator descriptor that aids in selecting the most appropriate band for scoring. However, some criticism can be raised regarding the imprecise nature of the three-part rating scale, the vagueness of some of the evaluative categories such as (4) *Motivation*, and the overall scope of the evaluation rubric.

Part II: EMPIRICAL PART

5. Methodology

5.1. Research interests and research design

As previously stated, this study is guided by two primary objectives. The first objective is to create a tailored evaluation rubric for LLAs. This rubric aims to assist teachers and app developers in conducting systematic and comprehensive evaluations of LLAs. The second objective is to apply this newly created rubric to assess two popular LLAs, namely *Duolingo* and *Mondly*. Through this assessment, the study aims to scrutinise their design and mechanics, while demonstrating the usability and practicality of the evaluation rubric for LLAs.

While pursuing the second primary objective, which involves the analysis of the apps, this study aims to answer several questions. Firstly, it will try to explore how popular LLAs work and which principles and methods of EFL teaching underpin their design. Secondly, it will address how LLAs can potentially facilitate language learning. Finally, during the discussion of the app analysis results, this study will try to outline potential ways of integrating LLAs into the EFL classroom and address how they can enhance the language learning experience in such a setting.

To achieve the two main objectives of this study, a combination of various qualitative research methods was employed. The first draft of the theory-based evaluation rubric for LLAs underwent a usability test, which involved soliciting feedback on its design through a peer review process. The obtained feedback was later utilised to refine the initial draft of the evaluation rubric, leading to the production of the final version. Subsequently, this final version of the evaluation rubric was applied during the evaluation of two LLAs, *Duolingo* and *Mondly*. The nature of this evaluation can be best described as systematic-qualitative-interpretive. Further details concerning the rubric's development process, the participants involved in the peer review, and the selection of the apps will be discussed in the respective sections.

5.2. Developing an evaluation rubric for language learning apps

As discussed in section 4.2., several app evaluation rubrics have been developed during the last decade. Many of these rubrics were created to evaluate educational apps in general (e.g. Lee & Cherner 2015) and only a few rubrics were designed for language learning apps (e.g. Chen 2016). This study assesses the value of LLAs for secondary level teaching. However, among the limited pool of rubrics aimed at evaluating LLAs, no rubrics have been created to specifically measure the usefulness of LLAs at the secondary education level. Therefore, a new rubric for evaluation of such LLAs has been devised.

Evaluation rubrics can be divided into two overarching categories: holistic rubrics and analytical rubrics. Holistic rubrics are deployed to arrive at an overall judgement about the quality of the evaluated material or performance, whereas analytical rubrics score the quality of each defined dimension of the material or performance, arriving at a total score by summing the individual scores (Allen & Tanner 2006: 198, Jonsson & Svingby 2007: 131-132, Mertler 2000: 1). The novel rubric for assessing LLAs in secondary teaching is an analytical scoring rubric with four main domains.

Rubric design can be guided by a few principles. Firstly, rubrics should aim at an appropriate and optimal number of evaluative dimensions. Comprising too many dimensions might render a rubric dysfunctional, whereas too few dimensions might not provide sufficient information in terms of the quality of the evaluated app. Furthermore, rubrics should include operational criteria and performance-level descriptors, which should constitute a sensible rating, for example, a scale ranging from one, the worst performance, to five, the best performance. Moreover, the performance-level descriptors must be clear and informative, allowing the evaluators to access and apply the rubric easily and effectively (Papadakis, Kalogiannakis & Zaranis 2017: 3152-3154).

5.3. First draft

Based on Lee and Cherner's (2015) comprehensive Evaluation Rubric for Educational Apps, a first draft of the Evaluation Rubric for Language Learning Apps for Secondary Level Learners (ERLLASLL) was created. This new evaluation instrument builds on the three domains and the 24 individual dimensions introduced in Lee and Cherner's (2015) rubric. Since Lee and Cherner's (2015) rubric aims at evaluating educational apps in general, a fourth domain

concerned with language learning was added in the Evaluation Rubric for Language Learning Apps.

The four domains of the first draft of the ERLASLL are (A) *Educational Value* (formerly called *Instruction*), (B) *Design*, (C) *Engagement*, and (D) *Language Learning*. The first draft of the rubric comprised 26 dimensions in total, with six dimensions belonging to domain (A) *Educational Value*, eight dimensions belonging to domain (B) *Design*, seven dimensions belonging to domain (C) *Engagement*, and five dimensions belonging to domain (D) *Language Learning*.

Six dimensions are used to evaluate (A) *Educational Value*: (A1) *Thinking Skills* (formerly called *Rigor*), (A2) *21st Century Skills*, (A3) *Value of Errors*, (A4) *Feedback to Teachers*, (A5) *Cooperative Learning*, and (A6) *Accommodation of Individual Differences*. In comparison with Lee and Cherner's (2015) rubric domain (A) *Instruction*, alterations have been made to dimensions (A1) *Rigor*, (A3) *Connections to Future Learning* and (A6) *Level of Material*. While the first dimension (A1) was simply re-named, the third dimension (A3) was abandoned for the sake of the rubric's scope. The sixth dimension (A6) was transformed into a new dimension called (D3) *Appropriateness of Language Learning Material* in the newly created domain (D) *Language Learning*.

Eight dimensions are used to evaluate (B) *Design*: (B1) *Ability to save progress*, (B2) *Integration*, (B3) *Screen Design*, (B4) *Ease of Use*, (B5) *Navigation*, (B6) *Information Presentation*, (B7) *Media Integration*, (B8) *Cultural Sensitivity*. In comparison with Lee and Cherner's rubric domain (B) *Design*, dimension (B6) *Goal Orientation* was also abandoned for the sake of the rubric's scope.

Seven dimensions are used to evaluate (C) *Engagement*: (C1) *Learner Control*, (C2) *Interactivity*, (C3) *Pace*, (C4) *Personalisation*, (C5) *Interest*, (C6) *Aesthetics*, and (C7) *Gamification*. In comparison with Lee and Cherner's rubric domain (C) *Engagement*, alterations were made to (C4) *Flexibility* and (C7) *Utility*. (C4) was renamed to *Personal Preferences* and (C7) *Utility* was replaced with *Gamification* in order to evaluate an app's ability to engage its users through the incorporation of game-like elements.

Finally, five dimensions are used to evaluate (D) *Language Learning*: (D1) *Quality of Input*, (D2) *Opportunity to Produce Output*, (D3) *Appropriateness of Language Learning Material*, (D4) *Skills*, and (D5) *Exposure to the Target Language*. However, after receiving peer

feedback, dimension (D3) *Appropriateness of Language Learning Material* was renamed to (D3) *Level of Language Learning Material* in the final draft of the rubric. The five dimensions of domain (D) *Language Learning* reference SLA theories such as: Krashen's *Input Hypothesis* (1982, 1985) and Swain's *Output Hypothesis* (1995), constructivist learning theory, in particular Vygotsky's *Zone of Proximal Development* (1978), as well as CLT teaching principles, including maximisation of exposure to the target language and integrated-skill instruction.

5.4. Peer review

To verify the rubric's face and content validity as well as gather valuable user feedback, the first draft of the Evaluation Rubric for Language Learning Apps for Secondary Level Learners was distributed among a sample of five teaching professionals, trained at the English Department of the University of Vienna, who teach in different schools at the secondary level. Together with the first draft of the rubric, the participants were provided with two language learning apps. While the first app, *ELSA Speak*, claims to focus primarily on facilitating speaking skills, especially pronunciation, the second app, *Mondly*, claims to focus on training listening and speaking skills as well as building vocabulary and grammatical knowledge, thus representing a more comprehensive language learning package.

The five participants were asked to test one of the two provided apps, thereby gaining a good overview of the app's design and the activities and topics covered. Afterwards, each participant was invited to partake in an observed usability test, where they applied the first draft of the evaluation rubric during their evaluation of the tested app. As a result of the ongoing COVID-19 pandemic, the usability tests were conducted in the form of online Skype calls. During the usability test, each participant was asked to think aloud while evaluating the selected app, thereby providing valuable user insight and feedback. The aim of the think-aloud usability tests was to elicit thoughts regarding the rubric's overall usability and design, as well as the appropriateness, comprehensiveness and relevance of individual dimensions and indicator descriptors. Following the usability tests, participants were engaged in brief follow-up interviews to address any potential questions and gather more feedback on the evaluation rubric. The recorded five think-aloud usability tests and follow-up interviews were later transcribed and the data was analysed. Transcripts of all five sessions can be found in the appendix.

5.5. Final draft

After conducting the peer review stage in the development process of the rubric, the first draft of the Evaluation Rubric for Language Learning Apps for Secondary Level Learners was adapted and optimised based on the participants' feedback and the authors observations.

Domain (A) Educational Value

Domain (A) *Educational Value* attempts to measure LLAs' instructional worth in terms of the thinking skills they demand and the instructional support they provide so that learners can achieve the app's intended learning objective. Domain (A) comprises dimensions (A1) *Thinking Skills*, (A2) *21st Century Skills*, (A3) *Value of Errors*, (A4) *Feedback to Teacher*, (A5) *Cooperative Learning*, and (A6) *Accommodation of Individual Differences*.

Dimension (A1) Thinking Skills (Lee & Cherner 2015, adapted)

Dimension (A1) *Thinking Skills* intends to evaluate the rigour or difficulty of learning tasks in an app in terms of the level of thinking skills a learner must apply to meet a task's learning objective, as is illustrated in Table 1 below.

Table 1 Dimension (A1) Thinking Skills (Lee & Cherner 2015, adapted)

Dimension (A1) Thinking Skills: Which level of thinking skills (Webb's <i>Depth of Knowledge</i>) must learners apply to meet the app's learning objectives?					
5	4	3	2	1	N/A
Learners are required to apply extended thinking skills (DoK-4) to meet the app's learning objectives. key words: <i>create, synthesise, design, analyse</i>	Learners are required to apply strategic thinking skills (DoK-3) to meet the app's learning objectives. key words: <i>assess, critique, formulate, hypothesise</i>	Learners are required to apply existing skills and conceptual knowledge (DoK-2) to meet the app's learning objectives. key words: <i>classify, compare, predict summarise</i>	Learners are required to recall and reproduce information (DoK-1) to meet the app's learning objectives. key words: <i>recall, recognise, match, memorise</i>	Learners are not required to apply any level of thinking skills because the app is geared more towards entertainment than education.	Not Applicable

As Table 1 shows, dimension (A1) *Thinking Skills* references Webb's (1997, 2002) *Depth of Knowledge (DoK)* framework. Webb's (1997, 2002) *DoK* model introduces four levels of thinking skills, which are positioned on a spectrum ranging from level 1, recall and reproduction, to level 4, extended thinking skills. These four levels of thinking skills are mirrored in the rating scale and its indicator descriptors. While apps on level 1 would ask learners to recall and reproduce information to accomplish a learning task, apps on level 2 would demand the application of learners' existing skills and conceptual knowledge to reach the task objective. Apps on level 3 would challenge learners' strategic thinking skills "when

assessing, formulating and making logical statements about ideas and phenomena" (Lee & Cherner 2015: 26). Apps on level 4, which generally comprise the most cognitively demanding learning tasks, would provoke learners' extended thinking skills as they accomplish the given task by means of "creating, critiquing, [designing], and synthesising information" (Lee & Cherner 2015: 26).

The first draft of dimension (A1) highlighted Webb's *DoK* framework, as was reflected in the formulations of the guiding question and the indicator descriptors. However, the peer review disclosed that Webb's *DoK* framework was unknown to all five participants. Despite this initial information gap, all participants were able to work with the scale and the dimension in general, as the indicator descriptors provided them with examples of typical activities for each level of thinking skills. Considering the peer review feedback, the design of the final draft, in particular the formulations of the guiding question and the indicator descriptors, was adapted. Moreover, the key words for all four levels of thinking skills were highlighted in the rating scale.

Dimension (A2) 21st Century Skills (Lee & Cherner 2015, adapted)

Dimension (A2) 21st Century Skills, illustrated in Table 2 below, intends to evaluate apps in terms of their capability to train particular learner skills needed to navigate and succeed in a modernised, rapidly changing, and technology-based world.

Table 2 Dimension (A2) 21st Century Skills (Lee & Cherner 2015, adapted)

Dimension (A2) 21st Century Skills: Does the app require learners to engage 21st century skills, such as <i>critical thinking</i> and <i>problem solving</i> (e.g. reasoning effectively, solving problems innovatively), <i>communication</i> and <i>collaboration</i> (e.g. communicating clearly and effectively, collaborating with others), <i>creativity</i> and <i>innovation</i> (e.g. thinking creatively, working creatively with others), <i>initiative</i> and <i>self-direction</i> (e.g. managing goals and time, working independently), and <i>information</i> and <i>media literacy</i> (e.g. accessing and evaluating information, applying technology effectively)?					
5	4	3	2	1	N/A
The app frequently requires learners to engage 21st century skills.	The app sometimes requires learners to engage 21st century skills.	The app rarely requires learners to engage 21st century skills.	The app does not require learners to engage 21st century skills.	The app promotes rote memorisation skills.	Not Applicable

As Table 2 above shows, 21st century skills comprise learner skills such as critical thinking and problem solving, communication and collaboration, creativity and innovation, initiative and self-direction, as well as information and media literacy. Ownership of such skills aims to help learners succeed in their future lives in a constantly changing environment. The particular set of skills used in dimension (A2) was adopted from the *Framework for 21st*

Century Learning published by the Partnership for 21st Century Learning (2019a, 2019b). The framework was selected on account of its clarity and conciseness. However, it should be mentioned that only a limited number of skills from the aforementioned framework were chosen, which were believed to be applicable in an evaluation of language learning apps. Furthermore, professional materials comprising definitions of 21st century skills are manifold.

The first draft of dimension (A2) comprised a hyperlink to the *Framework for 21st Century Learning* and a vague list of the relevant 21st century skills. Peer feedback questioned the logic of including hyperlink references to external sources in a rubric that should be as concise as possible and revealed that most participants had only a very general idea of potential 21st century skills. Considering this valuable feedback and the fact that most participants had only a very general idea of potential 21st century skills, the final draft of dimension A2 was altered in terms of the hyperlink provided in the first draft and the formulation of the guiding question. As Table 2 above shows, the hyperlink to the *Framework for 21st Century Learning* was excluded from the final draft. Furthermore, as was suggested by one participant, the guiding question now clearly states and explains which 21st century skills are to be considered in the evaluation process.

Dimension (A3) *Value of Errors* (Lee & Cherner 2015, adapted)

As Table 3 below states, dimension (A3) *Value of Errors* intends to assess apps' agency in terms of delivering constructive feedback about a learner's performance or understanding in order to foster and promote the language learning process.

Table 3 Dimension (A3) *Value of Errors* (Lee & Cherner 2015, adapted)

Dimension (A3) <i>Value of Errors</i>: If learners answer a question incorrectly, does the app use the incorrect answer to advance learning?					
5	4	3	2	1	N/A
The app offers learners an explanation for their incorrect answer, allows learners to review the material related to the question, and then gives them an opportunity to answer the question again	The app offers learners either an explanation for their incorrect answer or directs them to the material related to the question before giving learners an opportunity to answer the question again	The app offers learners an explanation for why their answer was incorrect, but the app does not give them an opportunity to answer the question again.	The app offers learners an opportunity to answer the question again, but the app does not provide information to explain why their answer was incorrect the first time.	The app does not give learners an opportunity to answer the question again or provide information about why their answer was incorrect.	Not Applicable

The role and importance of constructive feedback for language learning processes cannot be denied. However, arriving at a concise and straightforward definition of feedback is, unfortunately, not that simple. As Hattie and Clarke (2019) point out, feedback as a concept is somewhat of a conundrum, since it seemingly influences the learning process significantly while being extremely variable in its appearance. For this research project, Hattie and Timperley's (2007) conceptualisation of feedback shall be used. For them feedback is "information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding" (Hattie & Timperley 2007: 81). Moreover, Hattie and Timperley establish a four-level distinction in terms of feedback, dividing it into "feedback about the task (FT), about the processing of the task (FP), about self-regulation (FR), and about the self as a person (FS)" (2007: 90). Looking at language learning apps and their characteristics, most feedback will likely be provided on the task level (FT). This kind of feedback can be powerful, especially when the information is afterwards utilised to improve learners' task processing or refine their self-regulation (Hattie & Timperley 2007: 91).

Turning towards the dimension at hand, language learning apps that score the highest on this dimension would provide immediate, valuable, and constructive feedback in the sense that they would explain why the learner's answer was incorrect, would allow revision of the learning material, and would offer the opportunity to attempt the given task once again. Furthermore, subsequent tasks or learning materials would then ideally be adapted to the learner's level of proficiency.

Peer feedback for dimension A3 indicated that the guiding question as well as the indicator descriptors were comprehensible, clearly formulated, and easy to apply in the evaluation process.

Dimension (A4) *Feedback to Teacher* (Lee & Cherner 2015, adapted)

Dimension (A4) *Feedback to Teacher* evaluates whether an app provides teachers with sufficient information about their learners' mistakes and advance in the language learning process, as is stated in Table 4 below.

Table 4 Dimension (A4) Feedback to Teacher (Lee & Cherner 2015, adapted)

Dimension (A4) Feedback to Teacher: How does the app allow teachers to monitor their students' academic growth (language proficiency level), achievements (e.g. weekly challenges), and progress (advance in the language course)?					
5	4	3	2	1	N/A
The app tracks learners' academic growth, achievements, and progress made while using the app. Teachers can access the reports either through email or by visiting the app's website.	The app tracks learners' achievements or progress made when using the app, but it does not track learners' growth. Teachers can access the reports using email or by visiting the app's website.	The app tracks learners' achievements or progress made when using the app, but it does not track learners' growth. Also, learner progress can only be seen in the app and not by using email or a website.	The app only tracks learners' achievements or progress when they are using the app. Once the app is exited, progress is not saved, and teachers do not have any way to access it.	The app does not track students' academic growth, achievements, or progress made.	Not Applicable

While apps should ideally provide constructive feedback on the language learning process to the learners, crucial information about the learners' growth, achievements, and progress should also be accessible to the teachers to inform the design of effective future instruction, thereby meeting their students' individual needs. Sufficient feedback on learners' growth seems to be of particular significance in mobile-assisted language learning settings as a) learning in such contexts "often [extends] beyond the reach of the teacher" (Kukulska-Hulme 2009: 164), and b) the role of the teacher alters from being the superior "[information and] content provider at the centre of instruction to [being] a facilitator" (Baran 2014: 25) of more egalitarian, autonomous, and learner-centred educational settings (Cowan & Butler 2013: 3, 8). As Table 4 above indicates, effective language learning apps on dimension (A4) would therefore provide teachers with easily accessible, sufficient, valuable, and adequate data about learners' growth, achievements, and progress.

Peer feedback for dimension (A4), again, stated that the guiding question and the evaluation scale comprising the indicator descriptors were well designed and clearly formulated.

Dimension (A5) Cooperative Learning (Lee & Cherner 2015, adapted)

Dimension (A5) *Cooperative Learning*, illustrated in Table 5 below, evaluates if and how an app offers learners the possibility to cooperate during the language learning process and beyond.

Table 5 Dimension (A5) Cooperative Learning (Lee & Cherner 2015, adapted)

Dimension (A5) Cooperative Learning: Does the app allow learners to communicate (e.g. send/receive messages and post to message boards), share learning artifacts (e.g. images, recordings, and documents), and collaborate (e.g. create presentations and larger projects) with other learners?					
5	4	3	2	1	N/A
The app features an online community where learners can communicate freely with other users, share learning artifacts, and collaborate on projects.	The app features an online community where learners may be able to communicate with other users, share learning artifacts, or collaborate on projects.	The app features an online community where learners can communicate freely with other users, but they are not able to share learning artifacts or collaborate on projects.	The app includes an online community, but learners have little or no control over how they communicate with other users. Also, the ability to share learning artifacts or collaborate on projects is not available.	The app does not allow learners to communicate or interact with other users.	Not Applicable

As Oxford (1997) points out, potential positive impacts of cooperative learning on the language learning process have been proposed by various studies. The benefits of cooperative learning described in these studies concern a wide range of factors:

[C]ooperative learning is more effective in promoting intrinsic motivation and task achievement, generating higher-order thinking skills, improving attitudes toward the subject, [...] heightening self-esteem, increasing time on task, creating caring and altruistic relationships, and lowering anxiety and prejudice (Oxford 1997: 445).

Additionally, Zhang (2010) raises the issue of learners' individual accountability in the context of cooperative learning. As the interacting learners notice that progress and success in the language learning process depend on their individual contributions, a sense of shared responsibility is experienced, which eventually leads to greater learner autonomy, responsibility, and independence (Zhang 2010: 83).

Moreover, cooperative learning provides opportunities to produce written or spoken output more frequently, as meaning needs to be negotiated between interacting learners. This negotiation of meaning, as research suggests, can provoke certain "cognitive mechanisms [...] crucial for language learning" (García Mayo 2021: 1). Finally, the negotiation of meaning in a student interaction that emulates a real-life exchange will serve a communicative purpose with more opportunities to produce written or spoken language focusing rather on function than form (García Mayo 2021: 1; Zhang 2010: 83).

Considering the possible benefits of cooperative learning, apps that score high on this dimension would provide learners with the opportunity to connect with other learners

through communicating freely, sharing learning products with each other, or collaborating on learning projects inside an online community.

Similar to dimensions (A3) and (A4), the peer feedback for dimension (A5) *Cooperative Learning* did not reveal any ambiguity in terms of the phrasing. All participants managed to work with the dimension, and some participants even highlighted the clarity of the guiding question and the indicator descriptors.

Dimension (A6) Accomodation of Individual Differences (Lee & Cherner 2015, adapted)

Dimension (A6) *Accomodation of Individual Differences* intends to assess an app in terms of its potential to cater to students' unique needs and ways of learning, as is described in Table 6 below.

Table 6 Dimension (A6) Accomodation of Individual Differences (Lee & Cherner 2015, adapted)

Dimension (A6) Accomodation of Individual Differences: Does the presentation of the app's content recognise that learners learn in uniquely different ways (e.g. visual, auditory, and/or kinaesthetic)?					
5	4	3	2	1	N/A
All of the app's content recognises that learners learn in different ways and caters to differentiated instructional practices.	Most of the app's content recognises that learners learn in different ways and caters to differentiated instructional practices.	Some of the app's content recognises that learners learn in different ways and caters to differentiated instructional practices.	Very little of the app's content recognises that learners learn in different ways and caters to differentiated instructional practices.	The app's content is presented in a format that assumes all learners learn in the same way.	Not Applicable

Following the advent of *Communicative Language Teaching* (CLT), language learners and their unique and individual differences have become the reference point for language teaching. Upon entering a classroom, may it be analogue or digital, a student's language learning process will commence from an individual point of departure, as individual learners differ in terms of "their prior experiences, languages they know, interests, social background, motivation, cognitive abilities, strategies, learning style, etc." (Kolb & Rath 2021: 199). The effectiveness of any language learning process depends on how well these individual aspects are factored in. Thus, any mobile-based language learning provided by an app must also be assessed against the app's potential to consider learners' unique differences and cater to them accordingly. Looking at dimension (A6) an effective app that scores high on this dimension would therefore recognise learners' individual needs and adapt/customise its content accordingly.

The peer review of dimension (A6) highlighted the clear phrasing of the guiding question and the indicator descriptors. The participants agreed on the straightforward nature of this rubric dimension. Despite the participants' consensus on the design of dimension (A6), one participant raised an interesting question concerning the nature of apps and the resulting inherent difficulty to cater to all different learning styles. Although many LLAs attempt to present their content in various ways, for example visually and auditory, criticism in terms of the lack of kinaesthetic learning, e.g. through incorporating elements of physical activity, is valid. However, there is a growing number of LLAs which comprise augmented reality features that can be utilised to include some form of kinaesthetic learning. Certain LLAs might, for example, ask their users to scan their physical environment using the mobile device's camera. While the camera scans the surroundings, the app generates a virtual object, for example an animal, that users can then interact with on a language level by walking around the animal in order to explore new words, e.g. the animal's body parts.

Domain (B) *Design*

While domain (A) intends to evaluate LLAs' educational value, domain (B) *Design* intends to assess LLAs' functionality in terms of their design, as learners should ideally be afforded the chance to interact easily and effortlessly with an app's language learning content. Domain (B) comprises dimensions (B1) *Ability to Save Progress*, (B2) *Platform Integration*, (B3) *Screen Design*, (B4) *Ease of Use*, (B5) *Navigation*, (B6) *Information Presentation*, (B7) *Media Integration*, and (B8) *Cultural Sensitivity*.

Dimension (B1) *Ability to Save Progress* (Lee & Cherner 2015, adapted)

Dimension (B1) *Ability to Save Progress* looks at LLAs' ability to save learners' progress in order to allow them to proceed from the previously completed learning activity, as is described in Table 7 below.

Table 7 Dimension (B1) Ability to Save Progress (Lee & Cherner 2015, adapted)

Dimension (B1) Ability to Save Progress: Does the app allow learners to return to the content they were last engaging after exiting the app?					
5	4	3	2	1	N/A
After exiting an app, learners can reopen the app and automatically return to the content they were last engaging when they logged off or can manually select their desired content.	After exiting an app, learners can reopen the app and resume it automatically in close proximity to where they were when they logged off or can manually select their desired content.	After exiting an app, learners can reopen the app and manually select the content they were last engaging when they logged off.	After exiting an app, learners can reopen the app and manually select the content they were last engaging when they logged off, but the content may be different.	After exiting an app, learners must begin on the first level when returning to the app.	Not Applicable

As Table 7 illustrates, LLAs that score high on this dimension afford learners the opportunity to immediately resume their learning activity from their last point of access when they left the app. Enabling learners to resume smoothly from where they last stopped fosters learners' motivation, as interruptions in the learning process do not lead to tiring repetitions of previously handled content. Moreover, learners' time is spent more efficiently and learning opportunities are increased.

Peer review of dimension (B1) indicated that the guiding question was clearly formulated. Nevertheless, two participants raised questions regarding the phrasing of indicator descriptors 3 and 5. One participant commented on the mutual exclusivity of the terms *automatically* and *manually*, which in this case leads to a hierarchy, as *automatically* is rewarded with more points than *manually* in the first draft of the rubric. This issue was highlighted by another participant, who suggested that learners might prefer to be able to select the content manually after returning to the app. In light of these valid arguments, indicator descriptors 5 and 4 were rephrased to level the value of both options, that is to be able to return to the content *automatically* or select the desired content *manually*.

Dimension (B2) Platform Integration (Lee & Cherner 2015, adapted)

Dimension (B2) *Platform Integration* evaluates the degree to which a LLA incorporates different connections to other apps, online communities, independent websites, and learners' email to enhance the learning experience, as is illustrated in Table 8 below.

Table 8 Dimension (B2) Platform Integration (Lee & Cherner 2015, adapted)

Dimension (B2) Platform Integration: Is the app enhanced by how it connects to (1) other apps, (2) online communities, (3) independent websites, and (4) learners' email?					
5	4	3	2	1	N/A
The app is enhanced with how it integrates with all four of the listed connections.	The app is enhanced with how it integrates with three of the listed connections.	The app is enhanced with how it integrates with two of the listed connections.	The app is enhanced with how it integrates with only one of the listed connections.	The app does not integrate with any of the listed connections.	Not Applicable

As Table 8 shows, apps which score high on this dimension would integrate with all of the four connections listed in the guiding question, e.g. by allowing users to transfer their created learning artifacts to other LLAs or share the artifacts in online communities, on independent websites or through users' email. Allowing users to connect with each other beyond the users' individual in-app learning experiences promotes potential learner cooperation in learning processes. Furthermore, by allowing learning artifacts to be transferred to other apps, online communities or independent websites access to users' individual learning experience is provided and the acquired knowledge is shared.

Peer review of dimension (B2) suggested that the guiding question and the indicator descriptors were, in general, rather clearly and precisely formulated. However, comments provided by two of the participants highlighted two concerns. Firstly, the guiding question does not define whether online communities must be necessarily external or can exist inside the app as well, and secondly, evaluating the degree to which LLAs successfully integrate the listed connections in order to enhance the learning experience seems to demand a comprehensive and thorough understanding of a LLA's mechanics. With regard to the formulation of the guiding question, the term *online communities* was deliberately left without further definition, as it refers to external as well as internal communities. Concerning the second issue, it can only be agreed that a thorough and comprehensive insight into the LLA's mechanics will be necessary in order to be able to deliver a proper evaluation of dimension (B2).

Dimension (B3) Screen Design (Lee & Cherner 2015, adapted)

Dimension (B3) *Screen Design* intends to assess how well a LLA's user interface incorporates design aspects such as text, graphics, videos, sound, and speech in order to enhance and improve the app's content and thus the learner's user experience, as is indicated in Table 9 below.

Table 9 Dimension (B3) Screen Design (Lee & Cherner 2015, adapted)

Dimension (B3) Screen Design: Is the app's text, graphics, videos, sound, and speech well-organised? (focus: user interface)					
5	4	3	2	1	N/A
The app's text, graphics, videos, sound, and speech are always organised in a way that enhances the app's content.	The app's text, graphics, videos, sound, and speech are usually organised in a way that complements the app's content.	The app's text, graphics, videos, sound, and speech are organised in a way that does not enhance or detract from the app's content.	The app's text, graphics, videos, sound, and speech are not well organised and may detract from the app's content.	The app's text, graphics, videos, sound, and speech are cluttered and confusing, which detracts from the app's content.	Not Applicable

As Table 9 shows, for LLAs to score high on this dimension all individual constituents of an app's screen design need to be well-organised in order to complement the LLA's content and improve the user experience. Research (e.g. Galitz 2007) on user interface design in recent decades has highlighted the importance of good interface design, as it does not only potentially affect successful communication between user and device, but also impacts the users' motivation and performance in general, since screen design "is [...] the vehicle through which [...] [learning] tasks are presented" (Galitz 2007: 5). Screen design comprises various individual design aspects of an app's user interface, such as pictures, videos, graphics, icons, text, sounds, colours, sizes, etc. which should integrate successfully to ensure an enjoyable user experience. Assessment of each individual design aspect of the overall screen design, however, would go beyond the scope of the rubric. Therefore, screen design and its impact on a LLA's content and the user experience are observed and assessed holistically through this dimension.

Peer review of dimension (B3) yielded some stimulating comments. Two participants highlighted the clear and precise formulation of the guiding question and the indicator descriptors, which constitute a well-structured scale, while one participant proposed the idea to split up the holistic dimension screen design and create several more specific sub-dimensions which focus, for example, on the text, graphics or sound in particular. However, as mentioned above, this would reach beyond the scope of the present rubric. Finally, another participant questioned in how far the dimension of screen design can impact a LLA's content. While it is true that an app's design will not impact the actual programmed language learning content directly, it will alter the way the content is presented, and thus enhance or impair the learner's user experience.

Dimension (B4) *Ease of Use* (Lee & Cherner 2015, adapted)

As presented in Table 10 below, dimension (B4) *Ease of Use* examines an app's design in terms of its intuitiveness and accessibility for its potential users.

Table 10 Dimension (B4) *Ease of Use* (Lee & Cherner 2015, adapted)

Dimension (B4) <i>Ease of Use</i>: Is the app intuitive and are learners able to engage it with minimal guidance?					
5	4	3	2	1	N/A
Learners are able to engage the app immediately with no guidance.	Learners are able to engage the app immediately with minimal guidance.	Learners are able to engage the app, but only after some guidance.	Learners are able to engage the app, but only after substantial guidance.	Learners are able to engage the app, but training materials are continually needed to do so.	Not Applicable

As Table 10 above exemplifies, the design of LLAs which score highly on this dimension ensures that users can work effectively and independently without any additional guidance or with only minimal guidance required. Ease of use is a significant aspect with regard to any mobile learning app, as its effect on user motivation is profound. Being in need of substantial guidance in order to be able to use an app effectively will discourage potential users from engaging with the particular app any further, as this is tedious and minimises the users' real learning time considerably. Furthermore, the excess of available apps ready to be downloaded signals potential users that a more attractive alternative can be found at any time. Thus, being intuitively accessible is a main design aim for any successful LLA.

Peer review of dimension (B4) raised two questions. Firstly, one participant noted that the indicator descriptors do not clearly state the agent that provides user guidance. In terms of instruction, a teacher, an external person, or the app itself could provide guidance to the learners. A rephrasing of the indicator descriptors, however, was not considered, as the omission of a particular agent in the formulation in fact allows for a broader evaluation of this dimension since guidance could be provided by either a teacher, an external person, the app itself or a combination of these. Secondly, another participant directed attention towards the gradation in the indicator descriptors. For said participant, the terms *minimal*, *some*, and *substantial* lack precision and are too vague, since there is presumably no widespread consensus with regard to their meaning. To counter this argument, the importance of a certain vagueness in the formulation of some of the dimensions' indicator descriptors should be highlighted. In the case of dimension (B4) the terms *minimal*, *some*, and *substantial* are to be understood as three sections constituting a bipolar scale with

rather fluid boundaries between each section. It is assumed that educated raters who evaluate a LLA by means of this rubric can be trusted to arrive at a sensible conclusion in terms of the most fitting indicator descriptor by using their common sense.

Dimension (B5) *Navigation* (Lee & Cherner 2015, adapted)

Dimension (B5) *Navigation* aims to assess how effortlessly learners can navigate through a LLA's content and options, as is explained in Table 11 below.

Table 11 Dimension (B5) *Navigation* (Lee & Cherner 2015, adapted)

Dimension (B5) <i>Navigation</i>: How easily can learners move through the app's content and options?					
5	4	3	2	1	N/A
Learners can move through the app's content and options fluidly.	Learners need to put forth a little effort to move through the app's content and options.	Learners need to put forth some effort to move through the app's content and options.	Learners are somewhat impeded from moving fluidly through the app's content and options because of its organisation	Learners encounter substantial challenges when trying to move through the app's content and options because of its disjointed organisation.	Not Applicable

As Table 11 specifies, apps that score high on this dimension would enable learners to manoeuvre the app's content and options without difficulties. Apps can provide user support with regard to easy orientation in terms of having a straightforward navigation menu, which includes elements such as bottom or top navigation bars or tab bars, sidebar navigation menus, floating navigation buttons, e.g. navigation arrows, content overview maps, etc. A LLA's navigation menu design should ideally enable learners to move through the app's content easily without losing orientation and therefore becoming discouraged. Intuitive and easy navigation provided by a clear design supports learners in maintaining their focus on the actual learning task.

Peer review of dimension (B5) signalled that the guiding question and the indicator descriptors were clearly formulated. Nevertheless, one participant made an interesting remark concerning the formulation of indicator descriptor for score 3. From his point of view, the need to make multiple clicks or swipes to move through the content does not directly imply that a LLA's navigation menu is poorly designed and should therefore be assessed negatively or rather negatively, as multiple clicks or swipes might in fact be necessary to help increase a clear on-screen arrangement for apps that, for example, are very dense in terms of content provided. In the case of such apps, making multiple clicks or

swipes might positively impacting the overall screen organisation. Considering this valid argument, the indicator descriptors for score 3 and score 4 were rephrased.

Dimension (B6) Information Presentation (Lee & Cherner 2015, adapted)

Dimension (B6) *Information Presentation* evaluates if an app's content is presented to learners in a logical manner to promote understanding and knowledge acquisition, as is described in Table 12 below.

Table 12 Dimension (B6) Information Presentation (Lee & Cherner 2015, adapted)

Dimension (B6) Information Presentation: Is the app's content presented in a logical manner? (e.g. the app's content grows increasingly rigorous as learners experience success, the app activates learners' background knowledge before presenting them new information, and/or the app provides an overview of its content before learners engage in specific tutorials or activities.)					
5	4	3	2	1	N/A
The app's content is presented in a logical manner.	The app's content is presented in a manner that is mostly logical.	The app's content is presented in a manner that is somewhat logical.	The app's content is presented in a manner that is somewhat illogical.	The app's content is presented in an illogical manner.	Not Applicable

As Table 12 above shows, apps that score highly on this dimension present their content in a logical manner, which includes an increase in rigour as learners successfully progress between tasks, the activation of learners' prior knowledge ahead of the presentation of new information as well as the provision of an overview of the available content. Apps which do not comply with these principles of information presentation can provide very substantial and rich content to potential learners but will still fail to create meaningful and valuable learning opportunities, as the information presented will not be comprehensible, ultimately causing a decrease of the app's overall educational value. Besides considering an increase in rigour from one task to another, which, essentially, establishes a space for learners to enter the zone of proximal development proposed by Vygotsky (1978) in which "the level of potential development [...] [can be achieved] through problem solving under adult guidance [or app guidance] or in collaboration with more capable peers" (Vygotsky et al. 1978: 86), information presentation should acknowledge "the beneficial effect of [appropriate] knowledge activation in classroom settings [identified by research in] educational psychology/instructional science" (Shing & Brod 2016: 157). The role of the teacher, who, as Lucariello et al. (2016: 57) succinctly outline, assists the learners' conceptual growth or conceptual change caused by the impact of their pre-existing knowledge in the knowledge acquisition process, is transferred to the app.

Peer review of dimension (B6) indicated that the dimension's guiding question and the individual indicator descriptors were clearly and comprehensibly formulated. Observation also revealed that all participants managed to apply the dimension successfully during their evaluation. One participant highlighted the convenience of the more detailed explanation of the guiding question provided in brackets.

Dimension (B7) *Media Integration* (Lee & Cherner 2015, adapted)

Dimension (B7) *Media Integration* scrutinises how effectively an app's design integrates its individual media constituents, i.e. texts, graphics, videos, sound, and speech, "to form a cohesive program" (Lee & Cherner 2015: 32) with educational value for the learners, as is presented in Table 13 below.

Table 13 Dimension (B7) *Media Integration* (Lee & Cherner 2015, adapted)

Dimension (B7) <i>Media Integration</i>: Are the app's texts, graphics, videos, sounds, and speech integrated effectively so each of the app's media components complements each other and forms a cohesive program? (focus: learning content)					
5	4	3	2	1	N/A
The app's text, graphics, videos, sounds, and speech are integrated seamlessly to form a cohesive program.	The app's text, graphics, videos, sounds, and speech are integrated adequately to form a cohesive program.	The app's text, graphics, videos, sounds, and speech are integrated poorly, but they still form a mostly cohesive program.	The app's text, graphics, videos, sound, and speech are mismatched to the app's content, but they still form a somewhat cohesive program.	The app's text, graphics, videos, sound, and speech are mismatched to the app's content, and it does not form a cohesive program.	Not Applicable

As is evident in Table 13, an app scoring high on this dimension would combine the different media components seamlessly to create an educationally valuable learning experience. The call for a successful integration of individual media parts can be based on some of the basic design principles (e.g. the principles of spatial and temporal contiguity as well as the multimedia principle) of multimedia learning proposed by Mayer (2021b: 399-401). These principles are in turn deduced from the cognitive theory of multimedia learning (CTML), which describes the learner's mind as a dual-channel system with limited capacity for active processing (Mayer 2021a: 59). In this context it is imperative to consider the construction of multimedia messages that align with the cognitive capacity of the learner's mind to manage the three types of cognitive load, namely intrinsic, extraneous, and germane (Mayer 2021a: 66-67).

Peer review for dimension (B7) yielded interesting feedback. Overall, observation during the usability tests indicated that all participants understood the guiding question and could apply the indicator descriptors rather easily. However, while one participant confidently

suggested that no professionally developed app available to download from the App Store or Google Play would lack a successful integration of media components, thus questioning the rationale for including dimension (B7) in the rubric, three of the other participants commented on a perceived similarity or relatedness between dimension (B3) *Screen Design* and dimension (B7) *Media Integration*. Addressing the former comment, a quick research of available apps suggests that numerous LLAs display poor integration of media components leading to noncohesive, educationally valueless learning programmes. Addressing the latter comment, even though dimensions (B3) and (B7) might appear to evaluate similar aspects on the surface level, both dimensions evaluate apps on different levels. While dimension (B3) operates on the level of the user interface, dimension (B7) looks at the learning content.

Dimension (B8) *Cultural Sensitivity* (Lee & Cherner 2015, adapted)

As Table 14 below illustrates, dimension (B8) *Cultural Sensitivity* examines if a LLA incorporates culturally responsive teaching methods and principles to represent diverse populations and account for learners' differing cultural backgrounds.

Table 14 Dimension (B8) *Cultural Sensitivity* (Lee & Cherner 2015, adapted)

Dimension (B8) <i>Cultural Sensitivity</i>: Does the app use culturally responsive teaching methods to represent diverse populations?					
5	4	3	2	1	N/A
The app allows learners to connect to and share ideas with people from diverse communities across the globe.	The app allows learners to explore diverse cultures from across the globe using authentic pictures, images, and/or texts.	The app makes general references to one or more cultures.	The app makes reference to diverse cultures, but the information contains some cultural biases.	The app presents stereotypical images and information about cultures.	Not Applicable

As Table 14 explains, apps that score high on this dimension either afford learners access to diverse communities in which they can connect and communicate with people from different cultural backgrounds or allow them to delve into diverse cultures by means of utilising authentic input such as pictures, images, and texts, therefore employing culturally responsive teaching principles. Gay (2002: 106) defines culturally responsive teaching (CRT) "as using the cultural characteristics, experiences, and perspectives of ethnically diverse students as conduits for teaching them more effectively". Vavrus (2008: 49) suggests that the aim of CRT is "to help empower children and youth by using meaningful cultural connections to convey academic and social knowledge and attitudes". All five of the essential principles of CRT, i.e. establishing awareness about cultural diversity, adapting the curriculum to incorporate ethnic and cultural diversity, establishing learning communities,

connecting with ethnically diverse students, and adjusting instructions to accommodate ethnic diversity, proposed by Gay (2002: 106), should be factored in when designing a LLA in order to create a culturally sensitive 'digital' learning space, which is in line with Porto's (2010: 46) call for culturally sensitive classrooms. To arrive at a better understanding of how the content or learning tasks of LLAs might be evaluated in terms of their cultural responsiveness, the *Culturally Responsive Curriculum Scorecard* (Bryan-Gooden, Hester & Peoples 2019) or the *Fairness, Bias, and Cultural-Responsiveness Checklist* (Center for Collaborative Education 2017) can be consulted.

Peer review of dimension (B8) provided varied feedback. Overall, all participants were able to apply the dimension without great difficulty, which implies a rather straightforward and clear phrasing of the guiding question and the indicator descriptors. However, three of the five participants opted for the *not applicable* (NA) option in their evaluation, offering a number of reasons for their choice. Firstly, app 2 provided for evaluation covers pronunciation and intonation practice and its intended purpose does not allow for a profound incorporation of cultural diversity. Nevertheless, participants commented on the app's lack of choice in terms of being able to practice different varieties of English, as only American English pronunciation practice, in particular General American pronunciation practice, is covered. Thus, participants rightly pointed towards the lack of representation of different populations. Secondly, although not limited to the practice of pronunciation or intonation, one participant felt that app 1 lacked references to different cultures and authentic material to promote cultural diversity. However, this app did not promote stereotypical or biased depictions either, which persuaded him to select the *not applicable* option. The option to choose between American English and British English spelling and pronunciation as well as a perceived Eurocentrism, e.g. in terms of food choices at a restaurant, were highlighted by the participant. Therefore, score 3 might have been granted for this dimension since the app references one culture without presenting stereotypical or biased depictions of other cultures.

Domain (C) Engagement

While domain (A) intends to measure LLAs' educational value and domain (B) aims at evaluating LLAs' design, domain (C) *Engagement* attempts to scrutinise the means through which an app generates and maintains learners' motivation to employ it. Apps that fail to

perform well in this domain lack value for learners. Despite the multitude of reasons which can influence an app's appeal to learners, a number of assumed rather general motivational factors were included in this domain. These factors are covered by dimensions (C1) *Learner Control*, (C2) *Interactivity*, (C3) *Pace*, (C4) *Personal Preferences*, (C5) *Interest*, (C6) *Aesthetics*, and (C7) *Gamification*.

Dimension (C1) *Learner Control* (Lee & Cherner 2015, adapted)

Dimension (C1) *Learner Control* intends to assess if a LLA affords learners the opportunity to choose the starting point for their learning process by means of allowing them to select the content appropriate for them in terms of the grade level, age level, or difficulty level, as is described in Table 15 below.

Table 15 Dimension (C1) *Learner Control* (Lee & Cherner 2015, adapted)

Dimension (C1) <i>Learner Control</i>: Does the app allow learners to select the level (e.g. grade level, age level, or difficulty level) where they will begin engaging with content?					
5	4	3	2	1	N/A
The app allows learners complete control related to the level at which they begin engaging with content.	The app allows learners significant control related to the level at which they begin engaging with content.	The app allows learners some control related to the level at which they begin engaging with content.	The app allows learners little control related to the level at which they begin engaging with content.	The app allows learners no control related to the level at which they begin engaging with content.	Not Applicable

As Table 15 shows, apps that grant learners a high degree of learner control, in terms of choosing the level at which their learning process commences, perform well on this dimension. According to Simsek "[l]earner control refers to instructional strategies through which learners can exercise some level of control over the events of instruction" (2012: 1748). Approaching learner control from the point of view of self-determination theory (Deci et al. 1991, Ryan & Deci 2000), the assumption is that perceived control over elements of the learning process will yield satisfaction of learners' innate need for autonomy leading to enhanced intrinsic motivation for learning.

However, much of the empirical research on the effects of learner control on learning thus far has produced ambiguous or inconclusive results. While some studies report an overall positive effect on learning or a positive impact on some aspects of it, others observe no impact or even describe negative effects (Karim & Behrend 2013: 88, Katz & Assor 2007: 429, Schnackenberg & Sullivan 200: 19-20, Simsek 2012: 1749, Sorgenfrei & Smolnik 2016:

155, Vandewaetere & Clarebout 2011: 2323). Rather than arguing for an unlimited employment of learner control, many researchers and educators now emphasise

that the effectiveness of learner control depends on what (e.g., which elements of instruction, such as pace, display or task features) is controlled by whom (e.g., novice or more experienced learners) and, moreover, is only realized if learners recognize the control that is given to them (Corbalan, Kester & Merriënboer 2009: 455).

LLAs that transfer a perceived degree of instructional control at predetermined, strategic, and appropriate points in the learning process to learners, for example by affording them choice over the level of content or the progression of instruction, thereby meeting learners' inherent need for autonomy, will likely boost learners' intrinsic motivation for language learning.

Peer review of dimension (C1) yielded generally positive feedback. Three of the participants applied dimension (C1) without any trouble and positively remarked on the clear phrasing of the guiding question as well as the sound choice of modifiers for the indicator descriptors. Another participant also noted that modifiers for the indicator descriptors were well-chosen but simultaneously challenged the rationale behind the dimension. In her view, in a language learning setting at school, learners should not have to make decisions in terms of the content level. Based on an evaluative test, an app or a teacher should assign the matching content level to each learner. Therefore, this participant suggested a different evaluation objective for dimension (C1) and a rephrasing of the guiding questions. This valid argument neatly illustrates the scientific debate on the degree of control given to learners. Nonetheless, learners will likely use LLAs beyond classroom settings, which supports a call for some degree of control given to them in terms of the choice of the content level. Finally, task-involved learners who strive to prove their ability and increase their mastery can be trusted to generally opt for learning tasks whose difficulty level is "close to their own perceived level of competence" (Nicholls 1984: 333).

Dimension (C2) *Interactivity* (Lee & Cherner 2015, adapted)

As Table 16 below shows, dimension (C2) *Interactivity* examines whether a LLA merely presents content to learners or if it provides them with opportunities to engage in the content through learning activities.

Table 16 Dimension (C2) Interactivity (Lee & Cherner 2015, adapted)

Dimension (C2) Interactivity: Does the app only present content to learners, or does the app allow learners to engage with the content through activities?					
5	4	3	2	1	N/A
The app always requires learners to engage with its content through activities.	The app often requires learners to engage with its content through activities.	The app occasionally requires learners to engage with its content through activities	The app only allows learners to interact with it by searching for specific content.	The app only presents content to learners. Learners are never asked to engage in activities.	Not Applicable

As explained by Table 16, LLAs that score high on this dimension will provide learning opportunities by allowing learners to actively engage in the content through activities, which constitutes a key idea central to constructivist learning theories. Constructivist learning theories are rooted in both psychology and philosophy and have been influenced by a range of disciplines including linguistics, anthropology, and sociology (Kaufman 2004: 303, Olsen 1999: 347). Despite being described as "a broad church" (Fox 2001: 24) which encompasses different variations of constructivism, e.g. cognitive constructivism (Piaget 1971), social constructivism (Vygotsky 1978), and radical constructivism (Glaserfeld 1995), constructivist learning theories share a common core claim which is "that (human) knowledge is acquired through a process of active construction" (Fox 2001: 24). Approaching learning from a constructivist understanding, the assumption is that learners are not passive recipients of knowledge but are actively engaged in the learning process and construct new meaning through engaging in activities and building on prior knowledge and experiences (Moreno 2010: 298).

Peer review of dimension (C2) indicates that the dimension's guiding questions as well as the indicator descriptors are clearly formulated. With the exception of one participant, who struggled to determine whether repeating words as part of the pronunciation practice offered by the app under evaluation could be recognised as 'engagement' and therefore choose the *not applicable* option, all other participants managed to apply the dimension without difficulty. Two of the participants, however, commented on the phrasing of score 4 and 5. To them both scores were too similar in terms of the choice of adverb of frequency, with score 4 using the adverb *often* and score 5 using the adverb *frequently*. Both participants recommended a replacement of the adverb *frequently* with the adverb *always* to increase the grading, which was incorporated in the final draft.

Dimension (C3) *Pace* (Lee & Cherner 2015, adapted)

Dimension (C3) *Pace* intends to assess the extent to which learners are afforded control over the pace of instruction, i.e. the rate at which learners move through the app's content, or whether the app introduces the content at an appropriate pace for secondary level learners, as Table 17 below specifies.

Table 17 Dimension (C3) *Pace* (Lee & Cherner 2015, adapted)

Dimension (C3) <i>Pace</i>: Does the app allow learners to control the rate at which they are presented content? If the app does not allow learners to control the rate at which they move through the content, does the app present the content at an appropriate rate for secondary level learners?					
5	4	3	2	1	N/A
The app allows learners to completely control the rate at which they move through the app's content.	The app allows learners to mostly control the rate at which they move through the app's content.	The app does not allow learners to control the rate at which they move through the app's content, but the content is presented at an appropriate rate for secondary level learners.	The app does not allow learners to control the rate at which they move through the app's content, but the content is presented at a mostly appropriate rate for secondary level learners.	The app does not allow learners to control the rate at which they move through the app's content, and the content is presented at an inappropriate rate for secondary level learners.	Not Applicable

As is described by Table 17, apps that offer learners full control over the pace at which they progress through the content receive the highest score. If learners cannot control the pace, the app should provide content at an appropriate rate to anticipate cognitive overload, which is a phenomenon described by cognitive load theory (CLT). Cognitive load theory, introduced by Sweller (1988), seeks to demonstrate "how the information processing load induced by learning tasks can affect students' ability to process new information and to construct knowledge in long-term memory" (Sweller, Merriënboer & Paas 2019: 261). Cognitive load theory suggests that humans' cognitive processing of information is rather limited owing to the restricted capacities of our working memory (Sweller, Merriënboer & Paas 1998: 252).

As an alternative to allowing learners control over the pace of instruction, which research in computer-based instruction has found to significantly improve learning, effective pacing, a method which refers to dividing large units of information into smaller chunks, can be employed as strategy to decrease cognitive workload for learners (Moreno 2010: 221). Giving learners control over the pace of content might help to improve learner retention and memory, reduce pressure and learner anxiety, and facilitate learner confidence and motivation.

Peer review of dimension (C3) yielded positive feedback. All participants were able to apply the dimension in their evaluation without any difficulty, which implies that they understood the dimension's guiding question and the indicator descriptors. This assumption is further supported by one participant, who, during her evaluation, highlighted the clearly formulated guiding question as well as the well-designed indicator descriptors.

Dimension (C4) *Personal Preferences* (Lee & Cherner 2015, adapted)

Dimension (C4) *Personal Preferences* intends to assess the extent to which apps afford learners control over personalisation of certain features, e.g. the background music, images, colour scheme, font style or size, navigation menu, app interface language or language style, notifications, avatar, etc., as is shown by Table 18 below.

Table 18 Dimension (C4) *Personal Preferences* (Lee & Cherner 2015, adapted)

Dimension (C4) <i>Personal Preferences</i>: Can learners personalise the app by setting individual preferences (e.g. background music, images, colour scheme, font style or size, navigation menu, app interface language or language style, avatar, etc.)?					
5	4	3	2	1	N/A
Learners can personalise four or more features of the app.	Learners can personalise three features of the app.	Learners can personalise two features of the app.	Learners can personalise one feature of the app.	Learners cannot personalise the app.	Not Applicable

As Table 18 explains, high scoring apps on this dimension would allow learners to express their individual preferences by means of personalisation of four or more features. According to a meta-analysis of research findings on the effects of choice on intrinsic motivation conducted by Patall, Cooper and Civey Robinson (2008), choice, in general, has a positive effect on motivation, whereas "instructionally irrelevant choice", in particular, most strongly affects intrinsic motivation (2008: 295). One potential explanation for this surprising finding is that humans might perceive the choices they make, which are not related to the content or sequence of the actual learning task, e.g. the background music during an exercise or an avatar's name, as a way to show their personal identity or express their individuality (Patall, Cooper & Civey Robinson 2008: 295). In light of a self-regulatory strength model, which predicts that choices, especially ones that hold significant personal or instructional consequences, would have a negative effect on motivation owing to the effort necessary in that process, instructionally irrelevant choices might impact motivation more strongly as they are inherently low-risk and low-effort and do not require great effort or entail significant consequences (Patall, Cooper & Civey Robinson 2008: 294-295).

Peer review of dimension (C4) indicated that the dimension is easy to apply and that the guiding question and the indicator descriptors are clearly formulated. However, two participants commented on the number of features provided as examples in the guiding questions and the formulation of indicator descriptor score 5. In the first draft of the rubric, the guiding question comprised a list of three potential features that could be personalised, namely background music, images, and avatar. In comparison indicator descriptor score 5 states that learners can personalise four or more features. To help evaluators apply dimension (C4), both participants suggested an expansion of the list of personalisable features in the guiding question, which has been incorporated in the final draft shown in Table 18 above.

Dimension (C5) *Interest* (Lee & Cherner 2015, adapted)

Dimension (C5) *Interest* evaluates if an app's learning material, i.e. the content topics and learning activities, will match and attract learners' interest, as Table 19 below illustrates.

Table 19 Dimension (C5) *Interest* (Lee & Cherner 2015, adapted)

Dimension (C5) <i>Interest</i>: Will the app's language learning material (i.e. topics, activities) likely appeal to secondary level learners?					
5	4	3	2	1	N/A
The app's language learning material might be particularly interesting and engaging to secondary level learners.	Most of the app's language learning material might be interesting and engaging to secondary level learners.	Some of the app's language learning material might be interesting and engaging to secondary level learners.	The app's language learning material might not be interesting and engaging to secondary level learners.	The app's language learning material will likely deter secondary level learners from engaging with it.	Not Applicable

As Table 19 shows, apps that score high on this dimension would attract and maintain learners' intrinsic motivation through providing interesting and engaging language learning materials. "According to interest theory, the key to motivating students is finding activities that are aligned with their personal interests" (Moreno 2010: 334). Researchers in the area of interest separate *dispositional interest*, also known as personal or individual interest, which refers to "a relatively enduring preference for certain topics, subject areas, or activities" (Schiefele 1991: 302), from *situational interest*, which refers to "a transient state that arises when environmental features are perceived as novel, vivid, or intense" (Walkington & Bernacki 2014: 146).

The latter, situational interest, which apps could potentially trigger through appealing topics and learning activities, often precedes and facilitates the formation of dispositional interest.

Therefore, if a topic or activity does not align with a learner's personal interest, triggering situational interest can be a key factor to catch this learner's attention. However, research on seductive details, i.e. interesting or shocking pictures, videos, etc., has demonstrated a downside to situational interest. By adding seductive details which are not necessary for the understanding of the relevant information in the learning material, focus can be pulled from the actual learning task, leading to impeding effects on learning (Krapp 1999: 28, Moreno 2010: 335, Plass et al. 2014: 129). If apps' learning materials successfully cater to learners' dispositional interest, or generate situational interest without distracting from the actual learning task, learners' intrinsic motivation will rise, which should eventually result in an increased performance.

Peer review for dimension (C5) yielded positive feedback. All participants managed to apply the dimension in their evaluations easily, which indicates that the guiding question and the indicator descriptors are well formulated and thought-out. An interesting remark was provided by one participant, who questioned why the dimension's guiding question only aimed at lower secondary level learners. This comment pushed an already initiated thought process about the rubrics' intended learner level even further, eventually leading to a change of the intended learners in this study in general. Instead of solely focusing on lower secondary level learners, the rubric, and consequently this study in general, now evaluates apps for secondary level learners, encompassing both lower and upper secondary levels.

Dimension (C6) *Aesthetics* (Lee & Cherner 2015, adapted)

Dimension (C6) *Aesthetics* intends to scrutinise an app's potential to motivate learners through its interface design and aesthetic appeal, as Table 20 below clarifies.

Table 20 Dimension (C6) *Aesthetics* (Lee & Cherner 2015, adapted)

Dimension (C6) <i>Aesthetics</i>: Will the app's graphics and interface likely motivate learners to engage with it?					
5	4	3	2	1	N/A
The app's graphics and interface will motivate learners to engage with it.	The app's graphics and interface will most likely motivate learners to engage with it.	The app's graphics and interface may possibly motivate learners to engage with it.	The app's graphics and interface will likely have no impact on whether or not learners are motivated to engage with it.	The app's graphics and interface will likely deter learners from engaging with it.	Not Applicable

As Table 20 explains, if a LLA manages to motivate learners by virtue of an aesthetically appealing user interface, it scores highly on dimension (C6). For any successful user-centered design, usability is a key factor required. Even if an app's design is effective, efficient, and

satisfying, it is not automatically emotionally attractive to learners (Still & Crane 2017: 57). The emotional appeal of an app's design, however, should not be disregarded, as learners' feelings, which can be positively affected, for example, by colours (Shneiderman 2005: 515), will influence "what they want to use and how well they use it" (Still & Crane 2017: 59). Therefore, an effective way to cultivate learners' positive attitudes towards a LLA is through aesthetic design (Lidwell, Holden & Butler 2010: 20).

This aligns with Reyna's assumption, which states that "[d]esign and aesthetics have a profound impact on how users perceive information and learn, judge credibility and usability, and ultimately assign value to an online experience" (2013: 28). Research on the usability of websites (Phillips & Chaparro 2009, quoted in Still & Crane 2017: 56) has shown that users perceived an aesthetic design as easier to use compared to a non-aesthetic design, regardless of the actual ease of use. Similar effects of aesthetic design were observed with regard to websites' credibility (Reyna 2013: 30). In conclusion, a LLA whose user interface is aesthetically appealing might attract attention more easily and increase learners' level of motivation and commitment.

Peer review of dimension (C6) showed that the majority of participants were able to apply the dimension in their evaluations without difficulty. Two of the participants even praised the straightforward formulation of the guiding question and the indicator descriptors. Nonetheless, valuable constructive criticism was given by one participant. Although she was able to apply dimension (C6) in her evaluation, the participant criticised the similarity of the indicator descriptors for score 3 (orig.: *may possibly motivate*) and score 4 (orig.: *may motivate*) in terms of their formulation. To increase the degree between scores 3 and 4, she suggested to upvalue score 5 (orig.: *will most likely motivate*) and rephrase score 4 using score 5's original phrasing. To help evaluators choose the correct indicator descriptor score more easily, these suggestions were incorporated in the final draft of dimension (C6), as can be seen in Table 20 above.

Dimension (C7) *Gamification*

Dimension (C7) *Gamification*, a newly introduced dimension which went unaddressed in Lee and Cherner's (2015) original rubric, evaluates the extent to which LLAs incorporate game-like elements to increase learner motivation and engagement, as is explained in Table 21 below.

Table 21 Dimension (C7) Gamification

Dimension (C7) Gamification: Does the app use game-like elements (e.g. prizes, points, ranking, avatars, etc.) to boost learner engagement?					
5	4	3	2	1	N/A
The app uses four or more game-like elements which will potentially boost learner engagement.	The app uses three game-like elements which will potentially boost learner engagement.	The app uses two game-like elements which will potentially boost learner engagement.	The app uses one game-like element which will potentially boost learner engagement.	The app does not use game-like elements.	Not Applicable

As Table 21 shows, LLAs which provide gamification through incorporating several game-like features, such as badges, points, leaderboards, avatars, and others, score highly on this dimension. The concept of gamification has progressively attracted interest in education and educational research over the past decade, on the grounds of its hypothesised ability to boost motivation and engagement in learners (Sailer & Homner 2020: 78). Despite its growing popularity and increased use, Seaborn and Fels' (2015: 16) claim that gamification has not yet received an agreed upon standard conceptualisation still seems to be true today. Nevertheless, in search for a conceptualisation of the phenomenon, Deterding et al. (2011) provide a frequently cited and seminal definition. According to them, gamification in its core "is the use of game design elements in non-game contexts" (Deterding et al. 2011: 10). This definition clearly separates the concept of gamification from the concept of game-based learning, which can be defined as "learning that is facilitated by the use of a [serious] game" (Whitton 2012: 1337). Gamification, therefore, is not a product, i.e. a (serious) game, but rather a design process to enhance instruction and learning by means of incorporating game elements and mechanics.

A recent meta-analysis conducted by Sailer and Homner (2020) provides a synthesis of research findings on the effects of gamification on cognitive, motivational, and behavioural learning outcomes. According to the researchers, an analysis of 38 studies revealed that gamification of learning, indeed, has significant, positive effects on learning outcomes. Despite these promising claims regarding cognitive, motivational, and behavioural learning outcomes, the meta-analysis could not entirely answer the question of which factors (e.g. contextual and situational) play a significant role for the success of gamification. Nonetheless, given the results of the meta-analysis, a call for inclusion of game-like elements in LLAs to facilitate learner motivation can be made.

Peer review of dimension (C7) yielded positive feedback. All participants could apply the dimension in their evaluation and seemed to understand the guiding question as well as the indicator descriptors. Providing a list of examples of game-like elements in brackets which accompanies the guiding question appeared to be helpful during evaluation. One participant suggested a rephrasing of the indicator descriptors using adverbs to signal the different degrees of gamification. However, this suggestion was not incorporated in the final draft since using countable numbers to evaluate the number of game-like elements and therefore the degree of gamification has proven to be successful for dimension (C7), as indicated by the other evaluations.

Domain (D) *Language Learning*

While domain (A) *Educational Value*, domain (B) *Design*, and domain (C) *Engagement* were adopted from Lee and Cherner's (2015) comprehensive Evaluation Rubric for Educational Apps and partially adapted for the evaluation rubric created for this study, domain (D) *Language Learning*, not addressed in previous rubrics evaluating educational apps in general, was added as a novel domain. Overall, domain (D) intends to assess LLAs in terms of their value for language learning. The individual dimensions included in this domain evaluate LLAs with regard to their way of considering and incorporating some of the fundamental principles of language acquisition, learning, and teaching. These principles are covered by dimensions (D1) *Quality of Input*, (D2) *Opportunity to Produce Output*, (D3) *Appropriateness of Language Learning Material*, (D4) *Skills*, and (D5) *Exposure to the Target Language* and shall be discussed in more detail in the following sections.

Dimension (D1) *Quality of Input*

Dimension (D1) *Quality of Input* intends to examine if the language and topics provided by a LLA are comprehensible and relevant to learners, as is described by Table 22 below.

Table 22 Dimension (D1) *Quality of Input*

Dimension (D1) <i>Quality of Input</i>: Does the app provide comprehensible and relevant input (language and topics) for secondary level learners?					
5	4	3	2	1	N/A
The app continuously provides comprehensible input. It covers language that is highly relevant to secondary level learners.	The app provides mostly comprehensible input. It covers language that is mostly relevant to secondary level learners.	The app provides rather comprehensible input. It covers language that is rather relevant to secondary level learners.	The app sometimes provides input that is incomprehensible. It covers language that is aimed at the general public.	The app often provides input that is incomprehensible. It covers language that is aimed at a specific audience (e.g. people working in business).	Not Applicable

As Table 22 illustrates, apps that continuously offer comprehensible input that is highly relevant to secondary level learners score highly on this dimension. According to Krashen's Input Hypothesis, which is a part of his theory of second language acquisition, language is acquired "in only one way – by [...] receiving 'comprehensible input'" (1985: 2). If learners are to progress along the line of the "natural order", which is another one of Krashen's hypotheses, this comprehensible input must contain structures that are slightly beyond the learner's current level of linguistic competence. In other words, if a learner is at stage 'i', acquisition ensues if the learner is provided with comprehensible input of stage 'i + 1' (Krashen 1982: 21). Teachers and LLAs alike can facilitate the comprehensibility of input in linguistic (e.g. through simplified syntax, use of high frequency vocabulary) and non-linguistic (e.g. through use of pictures, familiar topics) ways. Krashen furthermore argues that optimal input, besides being comprehensible, should be interesting and relevant to learners, should not be grammatically sequenced, and should be provided in sufficient quantity (1982: 63-73).

Applying Krashen's Input Hypothesis and consequent suggestions regarding optimal input to the design of LLAs entails that apps should continuously provide enough interesting and relevant language input which is challenging yet comprehensible in order to ensure language acquisition. Furthermore, this optimal input should be "[grammatically] [u]nsequenced but natural", as such input will inherently comprise 'i + 1' structures (1982: 68). While LLAs, in principle, could meet all of the above needs regarding input, a shallow and superficial investigation into the ways in which many LLAs operate, reveals that language input tends to be highly grammatically sequenced. Nevertheless, as many LLAs continuously draw conclusions about linguistic competence from learners' output, they tend to provide quantitatively sufficient input that is just at the right linguistic level for the individual learner. This justifies employing dimension D1, with its rationale being Krashen's Input Hypothesis, in the evaluation of LLAs to assess their value for language learning.

Peer review of dimension (D1) provided constructive feedback. Participant 1 applied the dimension successfully during his evaluation and recognised that the input provided by the app was comprehensible and valuable, but questioned the relevance for secondary level learners. Participant 2 seemed to understand the guiding question easily and praised the formulation of the indicator descriptors in terms of their gradation. Participant 3 provided

critical comments on the formulation of the guiding question and the construct behind it. They acknowledged the clarity behind the term *comprehensible input* but questioned the meaning behind the terms *valuable* and *relevant*, suggesting that individual understandings could vary significantly between evaluators. They recommended a clarification of these terms in the guiding question. Participant 4 believed that, overall, the dimension would work well. Participant 5 suggested to split up the guiding question and to create two individual dimensions so that dimension (D1) would not assess three qualities, i.e. comprehensibility, value, and relevance, of the language input. Furthermore they also raised concerns about how to deal with input (language or topics) that is comprehensible but not relevant or valuable or the other way around.

In response to the feedback from participant 3, modifications were made to the guiding question of dimension (D1) in the rubric's final version. Since the term *valuable* could arguably lead to rather subjective interpretations during evaluation, it was decided, after careful consideration, to eliminate it for the sake of clarity. With regard to the term *relevant*, I am confident that teachers will have the capacity to determine which language and topics are *relevant* to their students while applying dimension (D1), taking into account their students' language proficiency, interests, and individual requirements. While participant 5's critique regarding language or input that might be comprehensible but not relevant underpins the call for a split of dimension (D1) into two separate dimensions, creating another dimension would stretch the scope of this rubric even further. If an instance of comprehensible yet irrelevant input arises, evaluators are trusted to make a sound judgement and solve this issue by selecting a score that lies between the higher and lower score.

Dimension (D2) *Opportunity to Produce Output*

Dimension (D2) *Opportunity to Produce Output* evaluates the extent to which a LLA offers learners the chance to produce written or spoken output, as is explained by Table 23 below.

Table 23 Dimension (D2) Opportunity to Produce Output

Dimension (D2) Opportunity to Produce Output: Does the app provide learners with the opportunity to produce written or spoken output?					
5	4	3	2	1	N/A
Learners are offered the opportunity to produce written or spoken output freely. Their output is not limited to predefined answers. The output goes beyond the sentence level.	Learners are offered the opportunity to produce written or spoken output rather freely. Their output is sometimes limited to predefined answers. The output is limited to the sentence level.	Learners are offered the opportunity to produce predefined written or spoken output. The output is limited to the sentence level.	Learners are offered the opportunity to produce predefined written or spoken output. The output is limited to the word level.	Learners are not offered the opportunity to produce any output when engaging the app. The app merely provides content for reception.	Not Applicable

As Table 23 specifies, apps which afford learners the opportunity to produce written or spoken output freely, without being restricted to predefined answers, going beyond the sentence level, perform highly on this dimension. In answer to Krashen's Input Hypothesis, Swain's Output Hypothesis (1985, 2005) postulates "that the act of producing [spoken or written] language [...] constitutes, under certain circumstances, part of the process of second language learning" (Swain 2005: 471). The Output Hypothesis thus suggests that learners improve their language proficiency when they engage in activities that require them to actively produce spoken or written language, rather than simply receiving comprehensible input.

While learners' fluency might increase by means of practising the target language through speaking and writing, Swain furthermore claims that linguistic output has three particular functions which relate to the improvement of learners' accuracy: "the 'noticing'/'triggering' function[,] the hypothesis-testing function[,] [and] the metalinguistic function" (1995: 128). Firstly, learners who produce spoken or written output might notice gaps in their linguistic knowledge, which could provoke them to discover what they are missing (1995: 129). Secondly, by trying out new structures and language forms while producing output, learners will be able to test their own hypothesis about the ways the target language operates (1995: 130-131). Thirdly, given particular task conditions, spoken or written output will not only reveal learners' hypotheses, but will also enable learners to reflect on them in order to "control and internalize" the target language (1995: 132). In summary, according to the Output Hypothesis, active production of output can contribute to the improvement of learners' fluency and accuracy in the target language.

Applying Swain's Output Hypothesis to the design of LLAs entails that apps should provide learners with ample opportunities to produce output to facilitate language learning in the ways described above. Learners should be able to produce spoken or written output freely, i.e. without being restricted to predefined answers, and this output should ideally go beyond the sentence level. If an app only allows for predefined answers to be given or only provides free output on the word level, the learners' target language output will not be able to take on the three functions described by Swain.

Peer review of dimension (D2) offered positive feedback. All participants appeared to understand the guiding question and the indicator descriptors rather effortlessly and therefore applied dimension (D2) successfully during their evaluation. Participant 3 praised the design and formulation of the indicator descriptors, as they outline a progression from offering no chance to produce output to offering the chance to produce free output reaching beyond the sentence level.

Dimension (D3) *Level of Language Learning Material*

Dimension (D3) *Level of Language Learning Material* intends to assess whether an app's language learning material (i.e. language, activities, topics) for specific language proficiency levels is appropriate for the intended target learners, as is described in Table 24 below.

Table 24 Dimension (D3) *Level of Language Learning Material*

Dimension (D3) <i>Level of Language Learning Material</i>: Is the language learning material (i.e. language, activities, topics) for specific language proficiency levels (e.g. A1) appropriate for the intended target learners (e.g. secondary level learners on level A1)?					
5	4	3	2	1	N/A
The level of the language learning material is always matched appropriately to the learners' level.	The level of the language learning material is usually matched appropriately to the learners' level.	The level of the language learning material is rarely matched appropriately to the learners' level.	The level of the language learning material is not matched appropriately to the learners' level.	The app's language learning material is inappropriate for language learning purposes.	Not Applicable

As Table 24 explains, apps that score highly on dimension (D3) will provide language learning materials that are always matched appropriately to the intended learners in terms of their language proficiency level. The essential assumption with regard to dimension (D3) is that apps explicitly disclose whom their language learning material is designed for. An app that, for example, offers language learning on level A1, needs to provide language learning material (i.e. language, activities, topics) that is appropriate for learners of said level. Mismatches between the learners' language proficiency level and the level of the provided

language learning material will not yield beneficial learning outcomes, since learners will fail to enter their zone of proximal development (Vygotsky 1978), in which learning could take place through support (e.g. scaffolding) provided by the app. Such mismatches, i.e. learners perceiving a learning activity as being too challenging or as being not challenging enough, will possibly lead to either boredom or frustration (cf. Moreno 2010: 90), which subsequently will lead to a decrease in learner motivation and interest, as well as a potential increase in learner anxiety. To assess the correct level of learning material needed, apps might use placement tests, progress checks, and, most importantly, the learners' written or spoken output.

Peer review of dimension (D3) indicated that the dimension's initial design presented in the first draft of the rubric needed revision. The original guiding question *Is the app's language learning material (i.e. activities) for specific grade-levels appropriate for its target audience?* caused confusion for two of the participants. Furthermore, the follow-up interviews revealed that all five participants appeared to interpret the question in slightly different ways. Participant 1 admitted that they did not understand what *specific grade-levels* referred to. Participant 2 expressed concerns while applying the dimension because the evaluated app did not explicitly state its intended learners, which rendered the application of dimension (D3) challenging for him. In contrast, participant 3 applied the dimension without difficulties, stating that the guiding question as well as the indicator descriptors were comprehensible and clear. Furthermore, she also positively highlighted the progression between the individual indicator descriptors. Finally, the guiding question caused a complete breakdown of the evaluation process for participant 4, as she did not understand the question and the construct behind it. Thus, she could not apply the dimension successfully. For the final draft of the rubric, the guiding question was therefore rephrased as can be seen in Table 24 above. To facilitate comprehension even more, additional explanations, given in brackets, were added.

Dimension (D4) Skills

Dimension (D4) *Skills* intends to evaluate the extent to which a LLA integrates the teaching of receptive skills, i.e. reading and listening, and productive skills, i.e. writing and speaking, in its language learning programme, as is specified by Table 25 below.

Table 25 Dimension (D4) Skills

Dimension (D4) Skills: How many of the four core skills (reading, listening, writing, speaking) does the app train?					
5	4	3	2	1	N/A
The app integrates all four skills. Vocabulary or grammar practice might be included in the learning activities.	The app integrates three of the four skills. Vocabulary or grammar practice might be included in the learning activities.	The app integrates two of the four skills. Vocabulary or grammar practice might be included in the language activities.	The app trains at least one of the four skills. Vocabulary or grammar practice might be included in the language activities.	The app only provides vocabulary or grammar practice. None of the four skills are trained explicitly.	Not Applicable

As described by Table 25, apps that teach all four core skills integratively through the design of their language learning activities score highly on this dimension. Oxford points out that "[l]iteracy – as well as language itself – is sometimes viewed as a set of skills" (2001: 6). Different approaches to language instruction have, for much of language teaching history, treated these skills as individual areas of competence that needed to be developed independently (Hirvela 2013: 1). Oxford refers to this as "segregated-skill instruction" (2001: 7). In answer to this view of skills as independent areas of language competence and expanding upon her preliminary observation, Oxford suggests that language proficiency develops from the interconnection of the main core skills, i.e. listening, reading, speaking, and writing, and their related knowledge areas, i.e. grammatical knowledge, and vocabulary knowledge, during language instruction (2001: 6).

The call for "integrated-skill instruction" (Oxford 2001: 9) basically mirrors Widdowson's earlier proposed "principle of integration" (1978: 160), a notion which preceded the communicative turn of the 1980s and the establishment of Communicative Language Teaching (CLT). At the heart of CLT, despite it being such a comprehensive approach, resides the objective of developing communicative competence through language teaching (Richards & Rodgers 2014: 85). As Hirvela aptly states "[e]mbedded in the notion of communicative competence is the reciprocity between language skills that constitutes real-life communication" (2013: 1). Put differently, successful everyday communication on the basis of communicative competence depends on learners' ability to use core language skills in an integrated way to convey their message. Applying these convictions to the design of LLAs, it only seems reasonable to evaluate the extent to which apps teach the four core skills integratively.

Peer review of dimension (D4) generated helpful feedback. Participant 1 expressed uncertainty regarding whether the pronunciation practice provided by the evaluated app

belonged to the core skill of speaking but based his score on the indicator descriptors' phrasing. However, he did not seem to realise that pronunciation practice as it was provided by the app integrated the practice of at least two skills, namely listening to conversations or reading written input and consequently practising one's pronunciation. Participants 2 and 3 both successfully applied the dimension during their evaluations, which implies a clear and comprehensible formulation of the guiding question and the indicator descriptors, but suggested that listing the four core skills in the guiding question might help evaluators. This proposal was furthermore corroborated by participant 4's expressed initial confusion about the four skills referred to in the guiding question. However, after reading the indicator descriptors, participant 4 grasped which skills the guiding question referred to and applied the dimension successfully. Participant 5 appeared to understand the guiding question as well as the indicator descriptors without problems and, too, applied the dimension successfully. Based on the received feedback, the guiding question was extended to include a list of the four core skills and minor changes were made to the formulations of the indicator descriptors.

Dimension (D5) *Exposure to the Target Language*

Dimension (D5) *Exposure to the Target Language* assesses the extent to which apps maximise learner's immersion in the target language, as Table 26 below explains.

Table 26 Dimension (D5) *Exposure to the Target Language*

Dimension (D5) <i>Exposure to the Target Language</i>: Does the app maximise exposure to the target language?					
5	4	3	2	1	N/A
The app only uses the target language in learning activities. No translations in the native language are provided, or they are only provided upon request. Certain app features (e.g. settings menu) might make use of the native language.	The app mostly uses the target language in learning activities. Some translations in the native language are provided. Certain app features might make use of the native language.	The app uses the target language and translations in the native language in learning activities to the same extent. Certain app features might make use of only the native language.	The app provides some exposure to the target language in learning activities. However, translations are very frequent. Other areas of the app are dominated by the native language.	The app only provides minimal exposure to the target language. The chosen native language dominates all areas of the app (e.g. learning activities, option menu, user interface).	Not Applicable

As Table 26 shows, apps that score highly on this dimension provide learners with maximised immersion in the target language (TL) and only exploit the learner's L1 for particular areas outside the actual learning activities, such as the app's settings menu. The

aim to maximise learners' exposure to the target language corresponds with a long-standing tradition in *second language acquisition* (SLA) which emphasises the role of target language input for language acquisition and is observable in most of the language teaching approaches of the 20th century, aside from the grammar-translation method and the reading method (cf. Hall & Cook 2012: 274-277). This monolingual paradigm, also referred to as *monolingual habitus* (Gogolin 2008, Kolb & Raith 2021) or *monolingual teaching* (Hall & Cook 2012), which suggests that everything happening in the classroom should happen in the target language, is underpinned by SLA theories such as Krashen's (1982, 1985) Input Hypothesis as well as research on the effectiveness of maximised use of the target language in SLA (cf. Thompson & Harrison 2014: 322).

The role, value, purpose, and extent of learners' L1 in the language classroom, however, has been an area of intense debate with researchers and educators seemingly not finding a consensus (Arnett 2013: 1). In their state-of-the-art article, Hall and Cook (2012), apparent advocates of a *bilingual approach*, report several instances where teachers code switch from the target language to the L1. According to them, teachers use learners' L1

for grammar instruction, classroom management and administration, to demonstrate empathy or show solidarity with the learners, to provide translations for unknown words and to compensate for learners' apparent lack of understanding and in response to learners already speaking in their own language (Hall & Cook 2012: 286).

Similar categories for teachers' L1 use are reported by Thompson and Harrison (2014: 323). Beyond SLA theories' influence on L1 use in the language classroom, Arnett (2013: 2) identifies various reasons for teachers' L1 use, such as teachers' beliefs about SLA, teachers' education and training, school policies regulating the language use, the target language and L1 relationship, lesson aims, the nature and characteristics of the activity or task at hand, the access to suitable and appropriate educational resources, and others.

However, the question of the optimal ratio between target language and L1 in the language classroom basically remains unanswered. While Hall and Cook (2012: 292) advocate a "judicious" use of the L1, i.e. at strategic points to support language acquisition, they simultaneously acknowledge the vagueness of this answer and point out that more research to determine a favourable ratio is needed. Furthermore, they report on studies which revealed significant differences in terms of quantity of L1 use in language classrooms in different educational contexts, ranging from 0% to 90% (Hall & Cook 2012: 284). These

reported high percentages of L1 use might be realisations of Turnbolls (2001: 536) expressed concern "that licensing teachers to speak the L1 in their SL [second language] or FL [foreign language] classes will lead to an overuse of the L1 by many teachers" as well as the result of a lack of research findings. Considering the ongoing debate on the role and extent of L1 use in SLA, this study will follow Krashen's (2004: 8) evolved idea which states "that the first language helps when it is used to make input more comprehensible" but should not discourage the use of comprehensible input.

In terms of apps' design of learning activities, we need to consider how far the app will mirror the role of the teacher when we want to make judgements about the extent to which the learners' L1 can or should be incorporated. In a blended language classroom, i.e. a classroom that would represent a hybrid of what is considered to be more traditional forms of language learning and digital mobile learning, the app would operate as language learning material with the teacher being a resource for clarification and for providing translations. In such a setting, I would argue for a maximisation of target language input with the learners' L1 stepping into the background. However, when learners use LLAs at home or on the move, meaning no teacher is present to provide clarification, scaffolding, or translations, the learners' L1 might be used in a *judicious* way in the language learning programme. Thus, dimension (D5) grants the two highest scores for apps which maximise target language input without providing any translations, some translations, or translations upon request with the L1 being employed only outside of the learning activities.

Peer review of dimension (D5) offered positive feedback. All participants seemed to understand the construct of dimension (D5) effortlessly and successfully applied the dimension during their evaluations. Participant 4 called the dimension "straightforward" and participant 3 praised the progression between the individual indicator descriptors. Nevertheless, one comment made by participant 3 raised a question. During the evaluation they pointed out that the app allowed learners to choose whether or not to be provided with translations in their L1. This possibility was not mirrored in the first draft of the dimension but was added to the final draft of the indicator descriptor for score 5.

5.6. Interpreting the evaluation result

The final score obtained through the evaluation can be interpreted using a general ranking system similar to school grades, based on points and percentages. However, it is important

to note that how teachers evaluate and compare apps may vary depending on what they prioritise in an app. Certain dimensions may hold more significance in specific cases, as teachers' preferences and requirements can influence the evaluation process. Therefore, the interpretation of the final score should consider the specific context and priorities of the evaluating teachers.

6. Analysis and Evaluation of language learning apps

As this study aims to analyse and evaluate language learning apps with regard to their usability in a secondary level language learning context, two mobile apps were selected and compared. The selection was based on the app's popularity in terms of number of downloads from the App Store (iOS) or Google Play (Android). Furthermore, the app's affordances played into the decision, as this study aims to scrutinise apps which claim to be holistic language learning packages. Thus, the two selected apps for this study are Duolingo and Mondly. In the following sections these two language learning apps will be a) analysed and b) evaluated using the devised evaluation rubric.

6.1. Duolingo

Duolingo, a household name in the area of language learning apps, is among the most popular educational apps globally. According to its annual report of 2022, Duolingo has over 500 million registered learners (Blanco 2022), with an average of about 54 million monthly active users and an average of about 14 million daily users (Curry 2023). It provides over 100 language courses that cover more than 40 different languages (Blanco 2022), such as English, French, Spanish, German, Chinese, and Swahili as well as constructed languages such as Klingon (Star Trek), High Valyrian (Game of Thrones), and Esperanto (Cardwell 2021). However, which of these 40+ languages is available as a language course depends on the learners' first language. Founded by Luis van Ahn and Severin Hacker in 2011 (Loewen et al. 2019: 295), Duolingo was launched to the public in 2012 and has been available as an online language learning platform as well as a mobile application for devices using Android and iOS operating systems (Teske 2017: 393). Duolingo is available as freemium software, which means that the main parts and functions of the app are available free of charge. However, certain features, such as personalised listening and speaking practice activities, unlimited lives, and ad-free lessons, can only be unlocked through subscribing to *Super Duolingo*, the premium version of the software, which requires payment of a monthly subscription fee. In the following sections Duolingo shall be analysed and evaluated in the domains of (A) *Educational Value*, (B) *Design*, (C) *Engagement*, and (d) *Language learning*.

6.1.1. Educational value

(A1) Thinking Skills. Turning towards cognitive demand, Duolingo requires learners to mainly apply thinking skills of level 1 (*DoK-1*) and level 2 (*DoK-2*) as they engage in the app's language learning programme. A more detailed analysis of Duolingo's exercise design and the underlying teaching and learning approaches, methods, and principles can be found in section 6.1.4. In summary, language learning activities in Duolingo ask learners to memorise, recall, and reproduce information, as well as recognise, predict or infer language patterns and grammatical structures. On dimension (A1), Duolingo therefore receives 3 points.

(A2) 21st Century Skills. If learners use Duolingo as means to advance their language learning, they will train several 21st century skills by default. Firstly, by managing their own learning time, setting appropriate learning goals, and working independently, learners will nurture their initiative and self-direction. Secondly, by accessing, using, and evaluating information provided through a mobile app, learners will train their information, media, and technology literacy, especially if they manage to successfully apply the technology to support their language learning. In many instances, Duolingo will provide scaffolding to its learners, e.g. by asking them to set a daily learning goal and reminding them through push notifications. Moreover, through its exercise design, Duolingo will foster learners' effective reasoning, thereby training their critical thinking and problem-solving skills. Therefore, Duolingo receives 5 points on dimension (A2).

(A3) Value of Errors. Duolingo provides its learners with immediate and explicit feedback, as is illustrated in Figure 3 below. After completing an exercise correctly, learners will receive praise (motivating phrases such as *Very good!*, *Super!*, *Well done!*, *Fantastic!*, *Brilliant!*, etc.) before they move on to the next exercise. If learners answer an exercise incorrectly, Duolingo immediately notifies them of their mistake and, depending on the type of exercise, either shows them the correct answer or asks them to try again. If the app detects that learners repeatedly make the same mistake (e.g. regularly forgetting to add the third person singular -s to verbs in the present simple tense), it interposes a so-called *smart tip*, which is a short explanation followed by an exercise. In terms of speaking exercises, Duolingo uses voice recognition that is based on artificial intelligence (AI) to analyse learners' spoken utterances and provide immediate feedback on their pronunciation (Moline & Blanco 2020). At the end of each lesson, Duolingo asks its learners to review the previously incorrectly

completed exercises, and they are only allowed to move on to the next lesson once all exercises are answered correctly. On dimension (A3), Duolingo therefore receives 5 points.

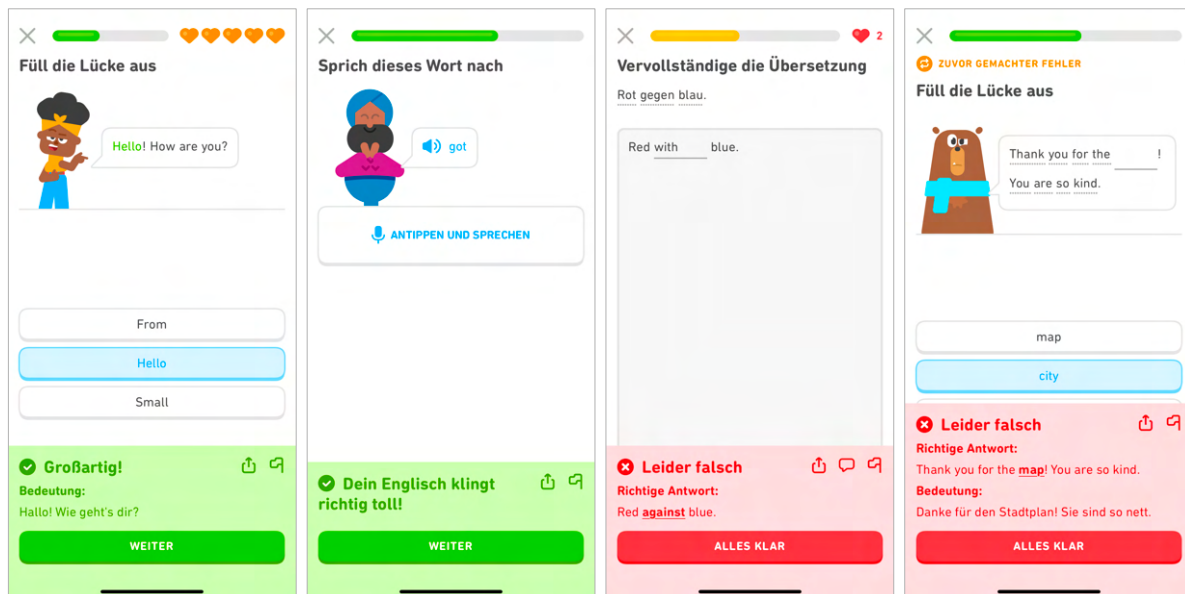


Figure 3 Examples of immediate feedback in Duolingo

(A4) Feedback to Teachers. Duolingo offers teachers feedback through "a free layer of management that sits on top of the Duolingo language learning app" called *Duolingo for Schools* (Duolingo 2023a). Through this online language course management tool, illustrated in Figure 4 below, teachers are able to create digital classrooms to which they can add students who possess a Duolingo account. In these digital classrooms, teachers can then assign tasks by either selecting the amount of XP (experience points) students should achieve or by selecting particular levels from a unit. When teachers assign a specific level, they are provided with an overview of the words and phrases that students will encounter for the first time (or review) in that level. Furthermore, they can filter levels according to the covered language proficiency level (CEFR A1-B2), the assigned activity type, and the defined learning objectives. Once assigned, Duolingo informs students about open tasks through push notifications and via email. After students complete an assigned task, the teacher receives feedback about their individual performances. However, it should be noted that the feedback provided is extremely limited and only reveals the time students spent on the task, their average success rate, and their gained XP. Detailed information, e.g. feedback about students' frequent mistakes or their overall growth in terms of language proficiency, is not provided. Therefore, Duolingo receives 4 points for dimension (A4).

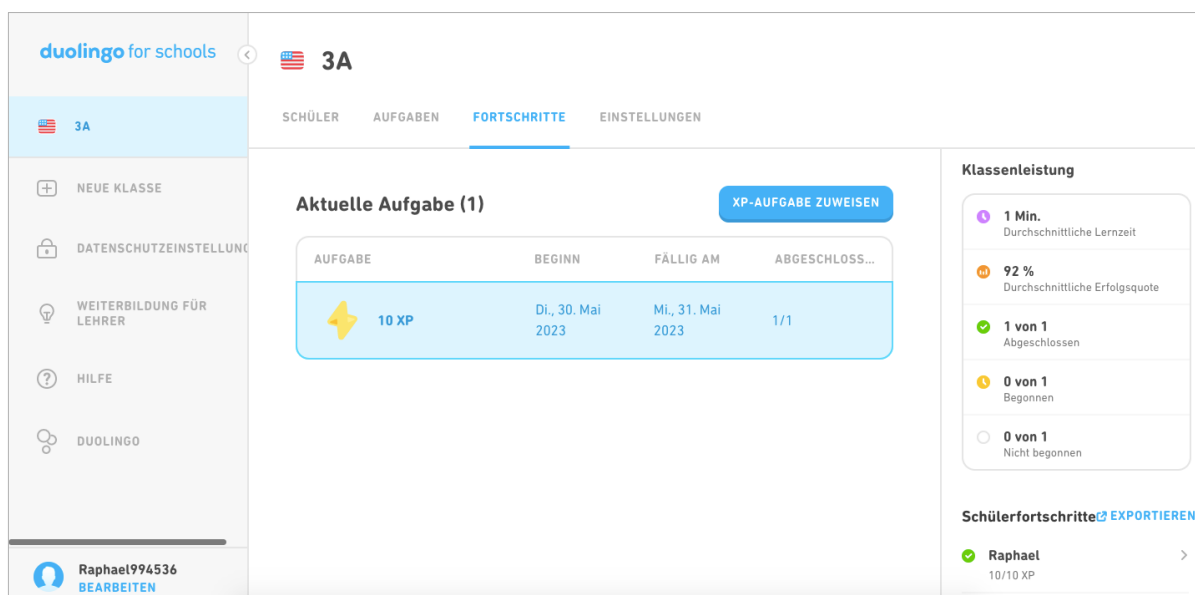


Figure 4 Duolingo for Schools management layer for teachers

(A5) Cooperative Learning. Duolingo does not afford its learners the opportunity to engage in real collaboration during the learning process and beyond. Furthermore, free communication between learners is not arranged for in the app and learning artifacts cannot be shared with each other. However, the app includes an online community where learners can follow each other, receive updates on learning milestones and achievements through a personalised social media-inspired newsfeed, and react to each other's milestones using emojis. Moreover, learners are invited to engage in so-called *friends quests*, which are weekly joint challenges that require two learners to earn a certain amount of XP through completing lessons and levels individually. These friends quests reward learners with XP boosts or gems upon successful completion. In the app, interaction and cooperation between learners is, nonetheless, extremely limited. On dimension (A5), Duolingo therefore receives only 2 points.

(A6) Accommodation of Individual Differences. Duolingo's language learning programme tries to cater to learners' individual differences and different learning styles. Many exercises in Duolingo are multimodal, as they combine written, spoken and visual input. While written translation exercises usually comprise a recording of the word, phrase or sentence, which can be accessed by tapping the speaker icon next to the item, language in use exercises and vocabulary exercises frequently merge written text and visuals. In addition, most listening exercises provide learners with recordings at two different speeds. Besides the regular recording that is played by default, learners can tap the tortoise icon and listen to a slower version of the same recording. Moreover, recordings can be played multiple times. Since the

app comprises some differentiated instructional practices, it receives 3 points for dimension (A6).

6.1.2. Design

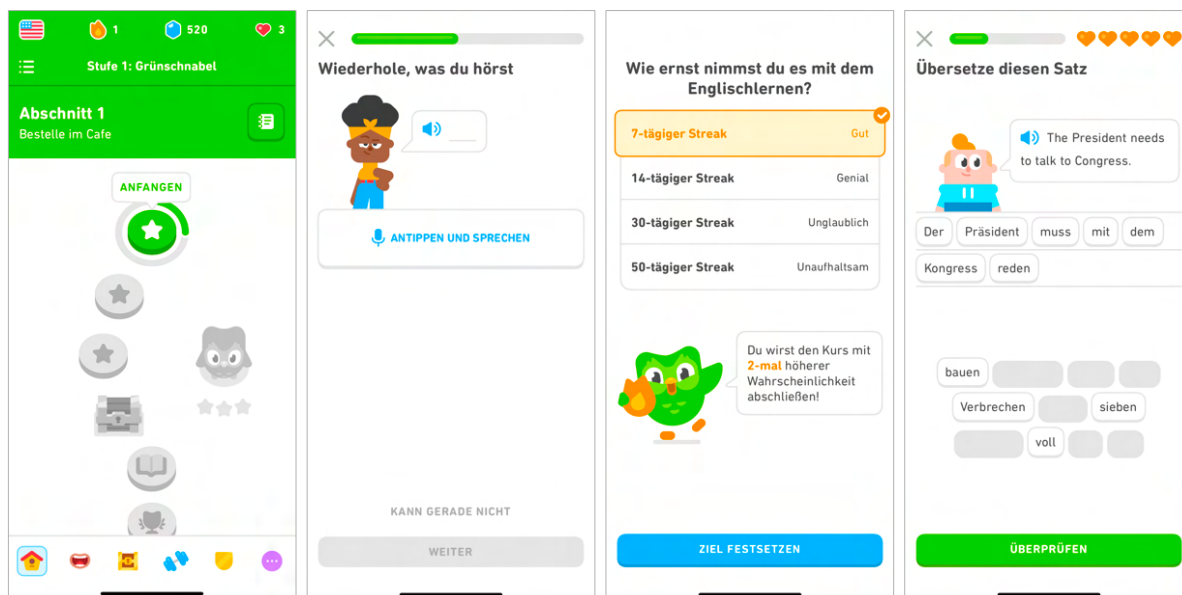
(B1) Ability to Save Progress. Duolingo only saves learners' progress if they set up an account. If learners exit the app and go to their phone's homescreen while working on a lesson, they will be able to continue from the same point they left after reopening the app. However, if learners completely close the app while they are working on a lesson and reopen the app, they will need to start the lesson from the very beginning. Similarly, if learners leave a lesson by clicking the cross icon (X), Duolingo will warn them about their progress being lost and they will need to start the lesson from the very beginning. Regarding dimension (B1) *Ability to Save Progress*, Duolingo allows its registered learners to automatically return to the content they were last working on when they exit the app or manually select their desired content. On dimension (B1), Duolingo therefore receives 5 points.

(B2) Platform Integration. While Duolingo does not officially and directly connect to other apps and independent websites, the app provides a rudimentary online community and employs e-mails to connect to its users. As mentioned above, Duolingo allows learners to follow other learners, post their own, as well as track others', progress and achievements in a social media-inspired in-app newsfeed, and collaborate in so-called friends quests. However, the app does not offer learners the space and tools for direct interaction and communication, e.g. through sending direct messages or posting on a message board. Duolingo owns accounts on popular social media platforms such as *Facebook*, *Twitter*, *Instagram*, and *TikTok*, which are used to update learners on news and developments. In addition, if not deactivated in the settings menu, learners also receive newsletters on app updates or friendly reminders about set learning goals, challenges, etc. via e-mail. Finally, Duolingo promotes the unofficial *Duolingo Discord Server*, which hosts real-time text, video, and voice chats. There, learners can connect and interact directly, ask questions, and find learning partners and learning resources. As Duolingo provides a basic online community within the learning app and connects to its learners through e-mails, it receives 3 points on dimension (B2).

(B3) Screen Design. In terms of screen design, Duolingo offers a visually appealing and well-organised user interface. The app successfully combines individual design aspects regarding

written text, graphics, illustrations, animations, and sounds to enhance its language learning activities. Duolingo's screen design is marked by consistent design choices in terms of typography, shapes, and colours. Besides the font *DIN Next Rounded*, which is used for the main parts of all written text, the app also employs a custom, copyrighted font called *Feather Bold* (cf. Duolingo 2023d). With regard to graphics, illustrations, and animations, Duolingo primarily employs three basic shapes, namely rounded rectangles, circles and rounded triangles, which are assembled in various combinations to create two-dimensional, rather simple characters, animals, and objects (cf. Duolingo 2023c). Moreover, all design elements used in Duolingo's user interface adhere to a limited colour palette comprising the core brand colours, some bold and vibrant secondary colours, and some neutral colours (cf. Duolingo 2023b), which contributes to its cohesive design. Besides the high quality of individual visual design aspects, Duolingo manages to integrate all design constituents successfully, which enhances its language learning content. Therefore, Duolingo receives 5 points on dimension (B3).

(B4) Ease of Use. Learners will require no guidance when using Duolingo, as the app is designed to be intuitively accessible. Starting a language course, Duolingo will guide learners through a step-by-step onboarding process in which learners are asked to select the target language, their language proficiency level, and their daily learning objective before they complete a placement test. Task instructions and prompts throughout the app are clearly formulated and use easy language. Progress bars for individual levels and lessons inform learners about their progress. Moreover, Duolingo employs intuitive or natural mapping. This means that it uses structures and patterns for its language courses which learners might recognise from (online) games and quizzes. Duolingo furthermore relies on learners' intuition with regard to elements that require learner interaction. Drop shadows, i.e. shadows behind clickable buttons or movable elements, as well as (empty) grey spaces, which indicate that an element has been moved successfully or that a level has not been activated yet, are examples of ways in which Duolingo tries to signal learners that interaction is required. Some of these elements which contribute to an intuitive use are illustrated in Figure 5 below. As secondary level learners are believed to be able to use the app without needing guidance, Duolingo receives 5 points on dimension (B4).



(a) Learning path

(b) Drop shadow button

(c) Simple instructions

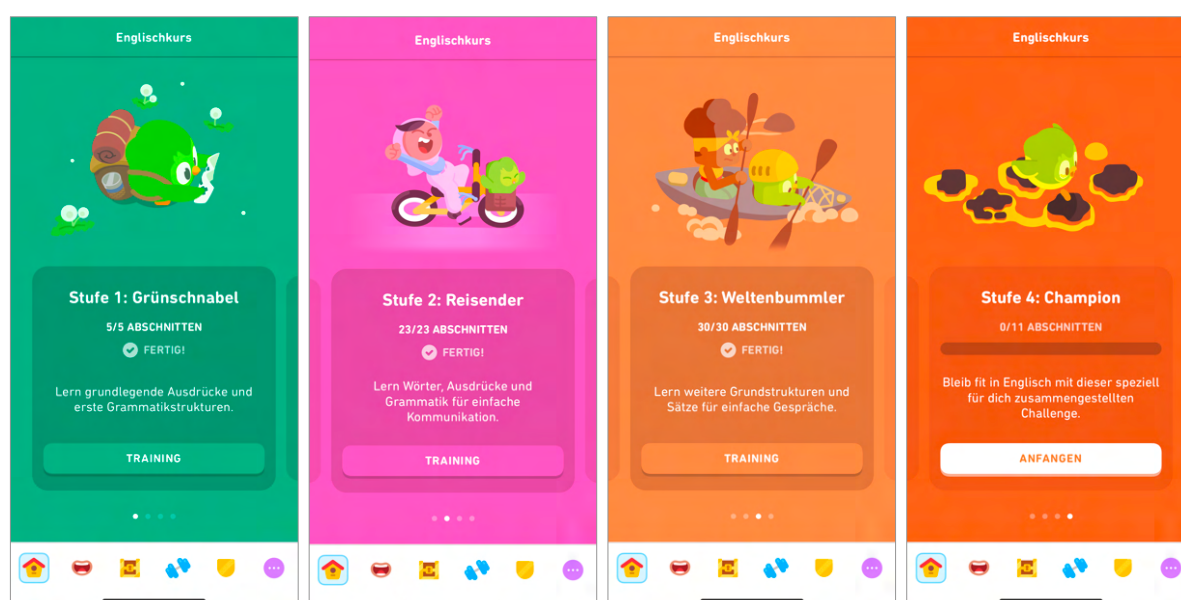
(d) Interactive elements

Figure 5 Elements contributing to ease of use and navigation in Duolingo

(B5) Navigation. Duolingo facilitates easy navigation on the basis of an intuitive navigation bar and linear learning paths, which are illustrated in Figure 5 above. The app's navigation bar, located at the bottom of the screen, allows learners to fluidly move between various spaces within the app, which are indicated by simple but expressive icons. Moreover, Duolingo's language courses take the shape of so-called learning paths. These have a rigid linear form and guide learners along a predetermined route that must be completed step-by-step. The learning path for Duolingo's English course is, for example, divided into four tiers, as is shown by Figure 6 below. Learners must successfully complete each tier before they are allowed to move on to the next. Alternatively, they are allowed to take a placement test, which, if passed, enables them to immediately move to the higher tier. On account of its simple but intuitive navigation bar and the predetermined, linear learning path, Duolingo receives 5 points on dimension (B5).

(B6) Information Presentation. Overall, Duolingo presents its content in a logical manner. Firstly, the learning path provides learners with an overview of the available content: tiers, illustrated in Figure 6 above, are organised into units, illustrated in Figure 7 below. Units comprise different topic areas (e.g. food, travelling, etc.) and aim at specific learning objectives (e.g. ordering at a cafe, making a reservation, etc.). They cover thematic vocabulary (e.g. food items, family members, etc.) as well as grammar (e.g. tenses, pronouns, plurals, etc.). Each unit encompasses several levels. Each level comprises five individual lessons. Each lesson, in turn, contains about ten exercises. After successfully

completing all exercises of a lesson, learners can access the subsequent one. All lessons of a level need to be completed before learners can advance to the next level. In between levels of a unit, as illustrated in Figure 7, learners receive reward chests, which provide them with additional gems (Duolingo's virtual currency) that can be spent on in-app purchases, e.g. for heart refills. Most units also comprise a story activity which is marked by the open book icon. Furthermore, learners can also access additional speed practice activities by tapping the animated Duolingo characters next to the learning path (e.g. the green Duolingo owl). Each unit concludes with a trophy challenge that unlocks the next unit. If learners want to progress to a higher unit without successfully completing the current one, they need to pass a placement test.



(a) Tier 1 *newbie*

(b) Tier 2 *traveller*

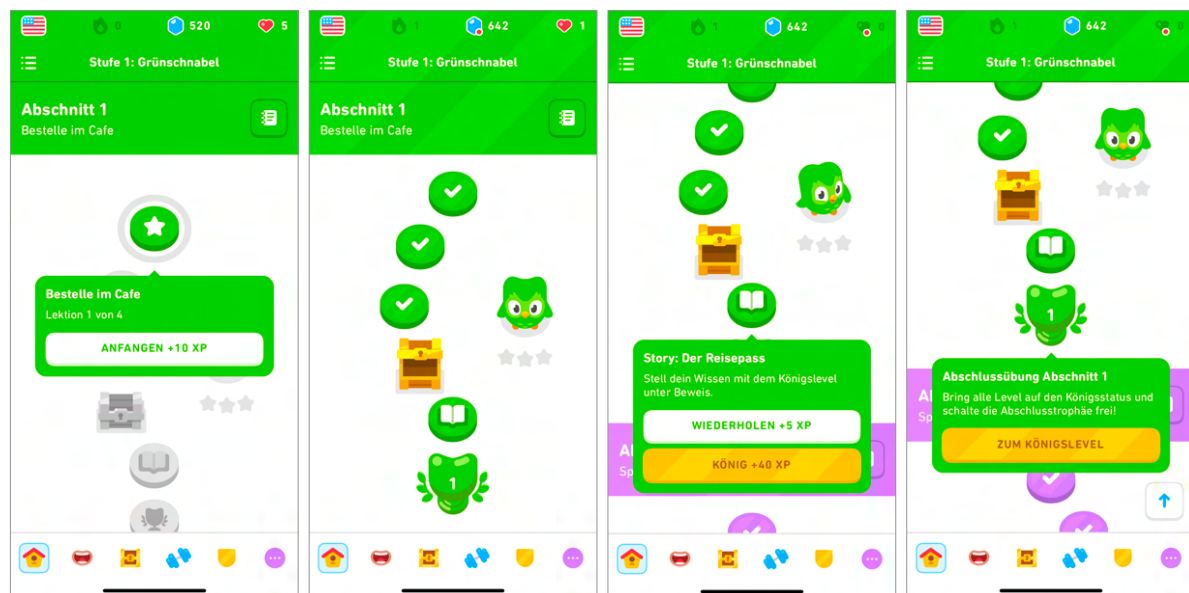
(c) Tier 3 *globetrotter*

(d) Tier 4 *champion*

Figure 6 Tiers in the learning path of Duolingo's English course

Secondly, as Freeman et al. (2023:3) claim, Duolingo's content grows increasingly more rigorous as learners move along the learning path. An analysis of the app's English course suggests that this claim is valid. While units become progressively more challenging in terms of the language level and the topic complexity, Duolingo also employs a machine learning model called Birdbrain to predict personalised difficulty levels for each learner based on its analysis of the provided teaching material and the learners' performance (Bicknell & Brust 2020). Given a learner's strong performance in a lesson, Duolingo's algorithm picks marginally more demanding exercises, thus aiming at the learner's zone of proximal development (cf. Freeman 2023: 5). Finally, Duolingo provides so-called *king challenges* for completed levels, which are more demanding than regular lessons. As Duolingo presents its

content in a logical manner, i.e. by providing an overview in the learning path, and gradually elevates the difficulty level while factoring in learners' individual performances, Duolingo receives 5 points on dimension (B6).

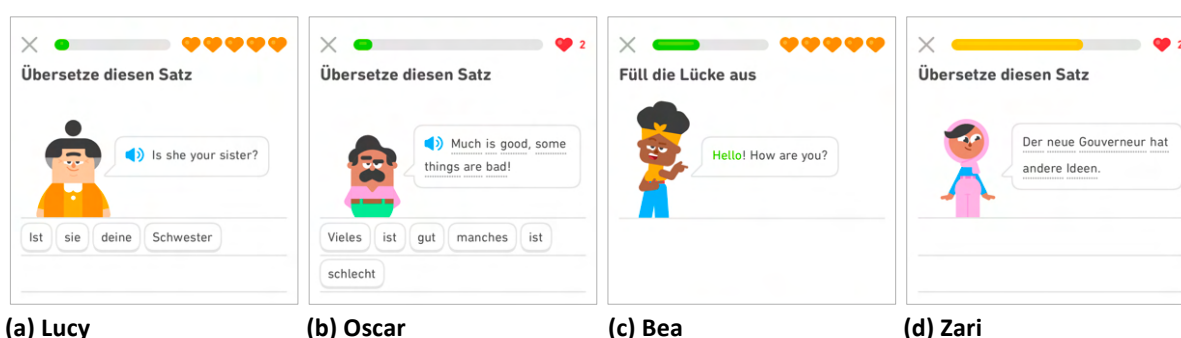


(a) Levels and lessons (b) Completed learning path (c) Story activity (d) Trophy challenge
Figure 7 Example of the Duolingo learning path for unit 1

(B7) Media Integration. Through seamless integration of all of its media components, Duolingo creates an educationally valuable and cohesive learning experience. Written texts, graphics, illustrations, animations, speech, and sounds are combined and aligned successfully. For instance, in exercises involving animated characters vocalising written text, lip movement accurately corresponds with the recording. Moreover, the voices of characters are tailored to match their age, gender and assumed ethnic background, resulting in the inclusion of diverse accents. Furthermore, characters respond accordingly to the learner's answer, providing distinct animated reactions for both correct and incorrect learner responses. In general, the analysis of Duolingo's language course revealed a consistent match between written text and recorded speech as well as written text and graphics. Disjointed media components were not found throughout the course. This consistency in terms of media integration greatly enhances language activities and contributes to the overall quality of the app. On dimension (B7), Duolingo therefore receives 5 points.

(B8) Cultural Sensitivity. Although "Duolingo does not make any claims about promoting intercultural competence or cultural awareness" (Teske 2017: 398), the app manages to represent diverse populations to a certain extent. This is mainly achieved by inclusion of diverse characters, which are illustrated in Figure 8 below. Apart from rather obvious

differences with regard to age, gender, and complexion, Duolingo's characters also allude to diversity in terms of sexual orientation and faith. According to Chiu (2021), Duolingo has included queer depictions and narratives in its story activities long before the app introduced its three queer characters Lin, Bea, and Oscar. However, the app does not draw particular attention to the character's sexual orientation or identity, it rather aims at normalising queerness through presenting it "in an unsensationalized and normal way" (Chiu 2021). With regard to linguistic diversity, Duolingo lacks range since it does not teach different varieties of English. All language learning activities follow American English spelling and pronunciation rules and conventions. Topics and language taught are rather generic, however, a tendency towards western culture is noticeable. Different cultures cannot be explicitly explored through Duolingo's content and the app does not employ authentic pictures, videos, or texts from across the globe. Furthermore, learners are not offered the chance to connect to learners from different communities, which entails that they cannot share ideas and engage in a transfer of cultural knowledge. On dimension (B8), Duolingo therefore receives 3 points.



(a) Lucy (b) Oscar (c) Bea (d) Zari
Figure 8 Examples of diversity in Duolingo characters

6.1.3. Engagement

(C1) Learner Control. When starting a language course, learners need to specify their language proficiency level in the onboarding process. They can select from four given answers (*I am learning English from scratch, I know some words and expressions, I can hold simple conversations, I am advanced*). Later on in the onboarding process, despite their previous answers about language proficiency, learners need to select a starting point. They can either state that they are beginners or that they have some language proficiency. Selecting the former will immediately place learners at the first lesson of the language course's easiest level. Selecting the latter leads them to a placement test which assesses their language proficiency. Upon completion, learners are placed at their most fitting level,

where they can start working on lessons. While the rigid learning path does not allow for high levels of learner control, the app allows learners to complete placement tests if they want to skip units or even entire tiers. Since Duolingo grants learners only some choice in terms of the starting point for learning, the app receives 3 points on dimension (C1).

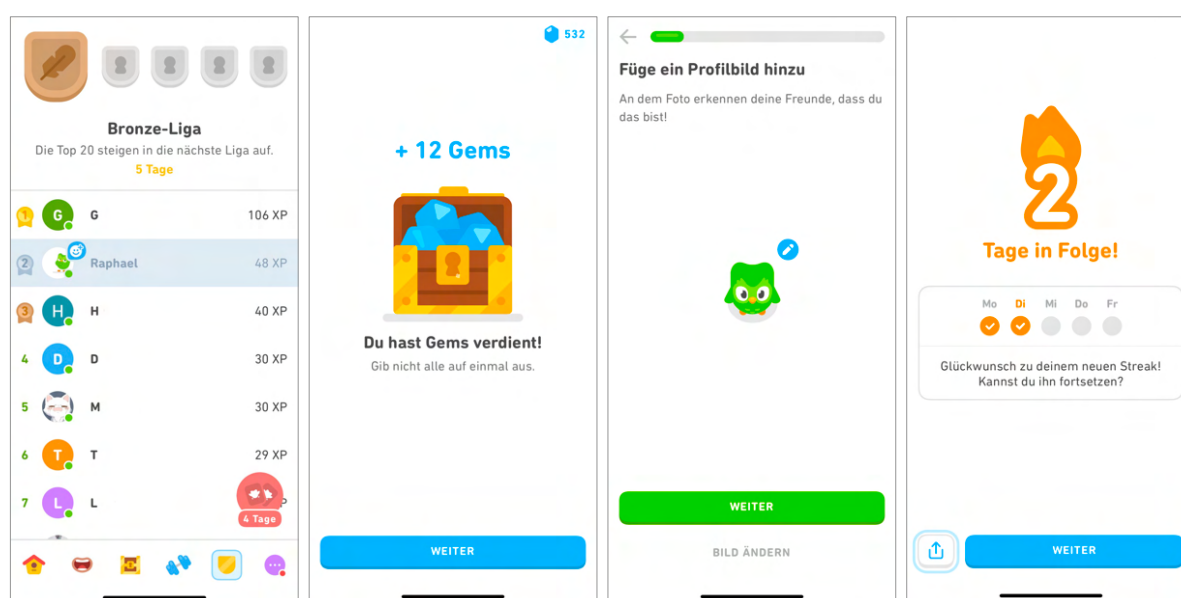
(C2) Interactivity. Throughout its language course, Duolingo follows a "learn by doing" approach (Freeman et al. 2023: 2). This means that the educational content covered by the app is mainly provided to learners through different types of interactive language learning exercises, which will be discussed in more detail in section 6.1.4. Since the app continuously requires learners to engage with the content interactively, Duolingo receives 5 points on dimension (C2).

(C3) Pace. Duolingo allows learners complete control over the rate at which they move through the app's content. In the learning path, learners must actively click on a level before they can work on its lessons. They are also required to tap the *start* button to commence each lesson, and upon completion of an exercise within a lesson they need to tap the *continue* button to advance to the next exercise. Apart from speed practice activities, which have a time limit of two minutes, learners are never given time limits within exercises. In speed practice activities learners also need to tap the *continue* button to move from one exercise to another. As these activities comprise a total of ten exercises, secondary level learners are expected to be able to complete them in the given time. Overall, learners have full control over the rate at which they are presented content. On dimension (C3), Duolingo therefore receives 5 points.

(C4) Personal Preferences. While Duolingo provides an overall cohesive learning programme wrapped in a well-designed user interface, the app lacks opportunities for user personalisation. Strictly looking at proper app features, learners are only afforded the opportunity to change their profile picture (avatar). App features such as the sounds, the colour scheme, the font size, the navigation menu, and others cannot be personalised in any way. Personalisation, however, is possible in terms of several settings: amongst others, learners are allowed to select and change their learning goals (e.g. in terms of the daily amount of XP achieved), to prioritise specific skills (speaking and listening exercises can be switched off temporarily or permanently), and to enable or disable push notifications (e.g. daily learning reminders) and motivating messages. As learners can only personalise one proper app feature, their avatar, Duolingo receives 2 points on dimension (C4).

(C5) Interest. Since Duolingo is not aimed towards a specific user group, its language learning material is rather generic and covers a broad range of topics and activities. Most of these are also covered by approved traditional EFL textbooks used in Austrian lower and upper secondary schools and might be interesting and engaging to secondary level learners. Therefore, Duolingo receives 4 points on dimension (C5).

(C6) Aesthetics. Duolingo's graphical user interface (GUI) and its individual components will likely motivate learners to engage in the app's language learning programme. Apart from its overall aesthetically pleasing and cohesive design, Duolingo's vibrant graphics, illustrations and animations are likely to appeal to younger learners and contribute to their motivation. On dimension (C6), Duolingo therefore receives 5 points.



(a) League ranking (b) Virtual currency (c) Personalised avatar (d) Learning streak
Figure 9 Elements of gamification in Duolingo

(C7) Gamification. Duolingo serves as a prime example of digital game-based language learning (DGBLL) (cf. Cornillie, Thorne & Desmet 2012), as the app employs several gamified elements in its language courses. A survey conducted by Govender and Arnedo-Moreno (2020: 676) identified 22 elements of gamification in Duolingo's language programme, making the app one of the most highly gamified LLAs on the market. These gamified elements, illustrated in Figure 9 above, include learner levels, quests, badges, lives (hearts), experience points (XP), league rankings, a virtual currency (gems), learning streaks, an avatar, and others that should foster learners' extrinsic motivation and boost learner engagement (cf. Freeman 2023). As Duolingo follows a highly gamified approach towards

language learning with more than four game-like elements, it receives 5 points on dimension (C7).

6.1.4. Language learning

(D1) Quality of Input. While Duolingo continuously provides comprehensible input, the relevance of topics and language throughout its courses requires scrutiny. Comprehensibility of language is guaranteed by Duolingo's own machine learning model *Birdbrain* (cf. Bicknell & Brust 2020), which analyses learners' responses and the material's difficulty level to arrive at conclusions about the appropriate difficulty level of prospective exercises for each individual learner. Furthermore, Duolingo sometimes facilitates comprehensibility of language input by including accompanying graphics, illustrations or animations, as well as by generally covering generic and familiar topics (e.g. family, travelling, food, emotions, jobs, etc.) that will activate learners' background knowledge.

In terms of the relevance of comprehensible language input, two issues arise. Firstly, language used in Duolingo's courses occasionally tends to be quite odd. Although units always cover topic-specific vocabulary and grammar and frequently teach useful common phrases and expressions (e.g. for ordering at a café, making a reservation, or giving one's opinion) learners might encounter what Duolingo calls "quirky" or "silly" sentences (e.g. *I am crying and the onion is laughing*), prompting the creation of numerous blogs and social media accounts which collect and share them online (cf. Hu 2021). The app's rationale behind including these odd constructions is based on the belief that such sentences will a) entertain and motivate learners and b) serve as memorable "grammatical anchor[s], helping [learners] [...] remember key examples of essential grammar concepts" (Leow, Bialostozky & Blanco 2021).

Secondly, the length of target language input provided in Duolingo's regular exercises could be increased. As a result of the app's short, sequenced, and disconnected language exercises, input never goes beyond the sentence level. This means that learners are left with single words, phrases or sentences that are presented in isolation without any further context. Longer sequences of connected language, i.e. two or more sentences forming short paragraphs, are only provided in Duolingo's story activities. In terms of quantity, the length of language input therefore may not be sufficient to really foster target language acquisition. Considering these issues, Duolingo therefore receives 4 points on dimension (D1).

(D2) Opportunity to Produce Output. Duolingo continuously affords learners opportunities to produce written and spoken output, but this output is always restricted. The app's language exercises, described and analysed in detail under point (D4) *Skills*, limit the freedom and length of learners' outputs. Learners are either required to select predefined answers to complete exercises or produce written or spoken output that is prompted by the task. This is freely produced output, however, it is ultimately predefined since the app expects a specific answer. Moreover, learners' output can never exceed the sentence level. On dimension (D4), Duolingo thus receives 3 points.

(D3) Level of Language Learning Material. While Duolingo claims to establish learning objectives for different proficiency levels based on the *Common European Framework of Reference for Languages* (CEFR) (cf. Blanco 2020), the app does not explicitly disclose which units in the language course correspond to specific CEFR levels. Furthermore, as a result of the learning path design and Duolingo's Birdbrain machine learning model, learners are not required to manually select their proficiency level. In fact, within the same lesson, different learners may encounter different exercises tailored to their individual language levels. Assuming the accuracy and effectiveness of Duolingo's machine learning model and the resulting algorithm, learners should consistently receive learning materials suitable for their personal language proficiency level. Consequently, Duolingo receives 5 points on dimension (D3).

(D4) Skills. Duolingo's English course does not equally prioritise the four language skills. The app places a stronger emphasis on developing learners' receptive skills, i.e. listening and reading. Nevertheless, it also addresses the development of learners' productive skills, i.e. speaking and writing. Additionally, not all aspects of a skill are necessarily covered comprehensively. Regarding speaking proficiency, Duolingo primarily focuses on improving learners' pronunciation rather than aiming for fluency. In addition to the four skills, Duolingo integrates the teaching of grammar and vocabulary knowledge. As a result, Duolingo receives 4 points on dimension (D4). To gain a better understanding of how each individual skill is trained by the app, the various types of language learning exercises will be analysed in detail.

Types of exercises. As of June 2023, lessons in Duolingo's English course encompass translation activities (a-f), listening activities (g-l), speaking activities (m-n), language in use activities (o), dialogue activities (p), and story activities (q). Translation activities and

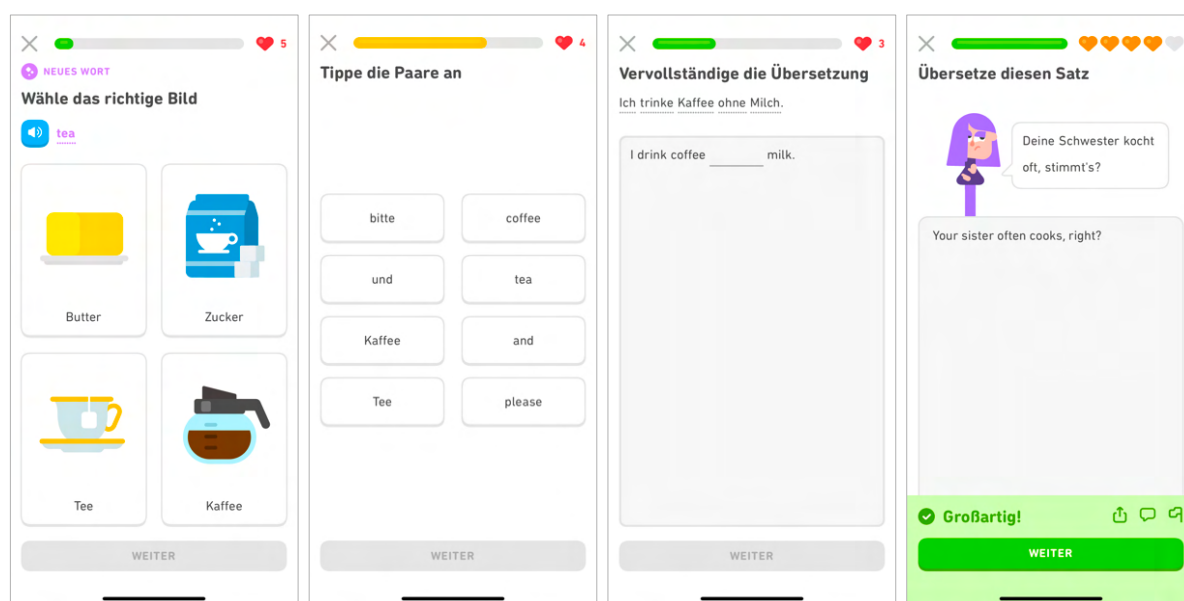
listening activities are the most common types of exercises employed in lessons, while the other four types of exercises occur less frequently. A complete list of the 17 different types of activities is provided in Table 27 below.

Table 27 Types of activities in Duolingo

Translation activities	
(a) Choose the correct picture	Learners are given a word in the L1 or the TL and must choose the correct translation from four flashcards.
(b) How do you say "xyz"?	Learners are given a word in the TL and three translations in the L1 to choose from.
(c) Tap on the word pairs	Learners are given eight words and must match the four L1 and TL pairs correctly.
(d) Complete the translation	Learners are given a sentence in the L1 and the corresponding sentence in the TL. The TL sentence has a blank that learners must complete by typing in the missing word.
(e) Translate the word	Learners are given a word in the TL or L1 and a matching picture and must type out the correct translation.
(f) Translate the sentence	Learners are given a sentence in the L1 or the TL and must translate it by typing it out or by arranging words to form the sentence.
Listening activities	
(g) What do you hear?	Learners listen to the recording of a spoken TL sentence and must type it out or arrange words to form it.
(h) What do you hear?	Learners listen to the recording of a TL word and must choose between two options with similar spelling.
(i) Listen and answer	Learners listen to two recordings of a TL word and must decide whether the two recordings are the same word or different words.
(j) Listen and answer	Learners listen to the recording of a TL sentence and must decide what the speaker is talking about.
(k) Write the missing word	Learners listen to the recording of a TL sentence and must type in the missing word to complete it.
(l) Listen and find the missing word	Learners are given a TL sentence with a blank and must choose the missing word from two recordings.
Speaking activities	
(m) Repeat what you hear	Learners listen to the recording of a TL word, phrase, or sentence and repeat it.

(n) Read the correct answer Learners are provided with a sentence or question in the TL. They must choose between two possible answers and must read the correct answer aloud.
Language in use activities
(o) Complete the sentence Learners are given a sentence in the TL with a blank and must choose the missing word from a number of options.
Dialogue activities
(p) Complete the dialogue Learners are given a sentence in the TL and must choose an answer out of two options to complete the dialogue.
Story activities
(q) Story activity Learners are provided with a story in the TL in the form of a dialogue with two speakers. In between listening to and reading the dialogue, learners must complete short comprehension tasks.

Translation activities. Translation exercises (a-f), exemplified in Figure 10 below, demonstrate how Duolingo blends principles of the *Grammar Translation Method* (GTM) and the *Audiolingual Method* (ALM) in its learning activities. While the GTM focuses on analysing grammar rules and translating sentences accurately (Richards & Rodgers 2014: 5), the ALM encourages learners to discover language rules independently and engage in rigorous practice sessions (drills) for reinforcement (Larsen-Freeman 2000: 35).



(a) Exercise type a

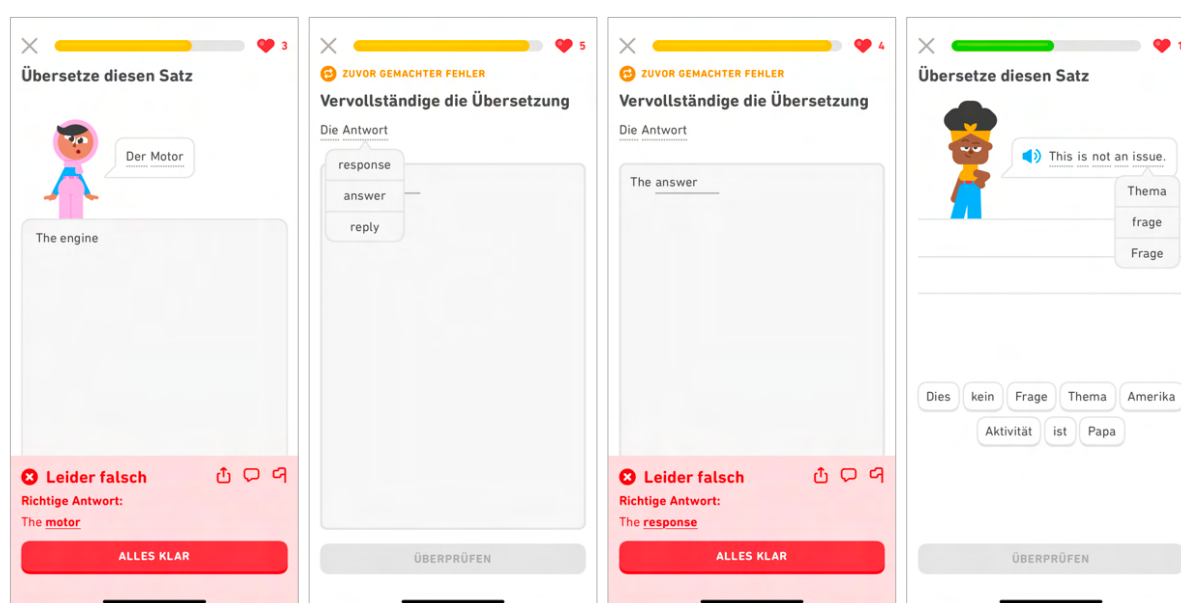
(b) Exercise type c

(c) Exercise type d

(d) Exercise type f

Figure 10 Examples of Duolingo's translation activities

Duolingo's translation exercises incorporate several GTM principles. Firstly, they utilise learners' L1 as a medium of instruction, to clarify unfamiliar words in the TL, and to draw comparisons between the two languages. Moreover, these exercises involve translations in both directions, requiring constant grammatical transfer between the two languages. Secondly, Duolingo's translation exercises primarily focus on isolated sentences, which "is a distinctive feature of the [grammar-translation] method" (Richards & Rodgers 2014: 6). Additionally, they aim to develop learners' vocabulary and grammar knowledge, which aligns with the GTM approach. Furthermore, in the GTM approach "words are taught through bilingual word lists, dictionary study, and memorization", and vocabulary is presented with translation counterparts (Richards & Rodgers 2014: 6). All of these strategies are present in Duolingo's translation activities.

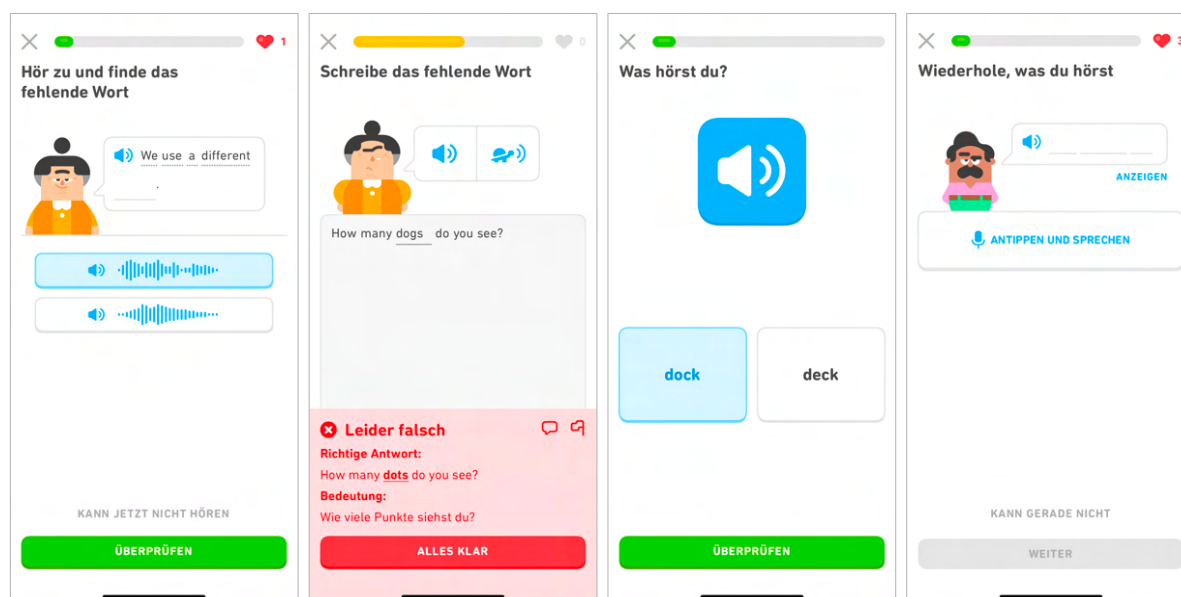


(a) Incorrect answer (b) Listed translations (c) Incorrect answer (d) Incorrect translations
Figure 11 Examples of flawed translation activities

Glossing, defined as "providing short definitions or explanations of the meaning of words in a given text" (Bowles 2004: 541), is an integral part of the design of translation exercises. Learners can easily access translations for unfamiliar items by tapping them. Duolingo also emphasises accuracy in learners' translations. For instance, selecting the English indefinite article *a* or *an* as a translation for the German definite article *die* is considered incorrect. Similarly, using a per se technically correct translation of a word (e.g. *the engine* for *der Motor*) that is not listed by Duolingo is marked as a mistake. Furthermore, some listed options, despite being provided by Duolingo, are not considered correct translations. Additionally, some of the listed options are themselves incorrect translations. Examples of

flawed translation exercises are shown in Figure 11 above. Regarding spelling mistakes in translation exercises, Duolingo is relatively lenient. If learners make a minor spelling mistake, the app notifies them but still considers the answer correct.

In addition to GTM principles, Duolingo's translation activities incorporate some tenets of the ALM. Instead of explicitly introducing grammar rules, learners are expected to identify the underlying grammar structures and apply them in translation exercises resembling typical ALM activities. These activities include replacement exercises (replacing one word with another of the same category or topic, e.g. *I drink coffee. – I drink tea.*), inflection exercises (changing the form of a word when the utterance is repeated, e.g. *I pet the cat. – I pet the cats.*), and restoration exercises (rearranging words to form a grammatically sound sentence, e.g. *is/sister/she/your – She is your sister.*), which all rely on drill and repetition (Savvani 2019: 41). Advocated by the ALM, repetition is believed to help foster language learning through the formation of habits. Correct habits should be positively reinforced, whereas incorrect habits should be corrected promptly (Larsen-Freeman 2000: 43). Duolingo incorporates these ideas in its translation activities through spaced repetition, praise, and immediate, explicit, corrective feedback.



(a) Exercise type l

(b) Exercise type k

(c) Exercise type h

(d) Exercise type m

Figure 12 Examples of Duolingo's listening and speaking activities

Listening activities. Listening exercises (g-k), exemplified in Figure 12 above, also integrate principles of the GTM and the ALM. As observed in the analysis of translation exercises, listening activities focus on individual words, lexical chunks or sentences. They can be categorised as dictation-based and transcription-based and merely train learners to listen for

specific information, such as the correct spelling of a word or the correct order of words in a lexical chunk. The development of a broader set of listening skills (e.g. listening for the main idea, listening for opinions, inferential listening, etc.) is not fostered. As listening exercises take the form of dictations by a single speaker, recordings can be described as being monologic. Recordings can be replayed multiple times, and to cater to learners with different needs, Duolingo provides dictations in two speeds. Learners can play the same recording at a slower pace by tapping the button indicated by a tortoise icon. If the learning setting does not permit learners to listen to audio recordings in a quiet environment, Duolingo allows them to temporarily turn off further listening activities for 15 minutes.

Speaking activities. Duolingo's speaking exercises (I), exemplified in Figure 12 above, are based on drill activities and pattern exercises representative of the ALM. In particular, they resemble repetition exercises where learners hear an utterance and repeat it aloud (e.g. *Walk towards him! – Walk towards him!*). Usually, the learner is not given a transcript of the spoken utterance and thus the utterance must be concise enough to be memorised easily. In repetition exercises "[s]ound is as important as form and order" (Richards & Rodgers 2014: 60). Duolingo uses voice recognition that is based on artificial intelligence (AI) to analyse learners' spoken utterances and provide immediate feedback (Moline & Blanco 2020). Since Duolingo's speaking exercises primarily train learners' pronunciation skills and ability to string words together to form correct sentences, learners' accuracy is a key objective. The limitation to single sentences, always provided without context, illustrates Duolingo's lack of facilitating learner's development of real communicative competence (fluency). If learners cannot talk because of a busy or loud environment or because they need to remain silent in public places, speaking exercises can be temporarily turned off for 15 minutes, which causes further speaking exercises to be provided as translation exercises.

Language in use activities. Other types of exercises found in Duolingo's English course focus on assessing learners' vocabulary and grammar knowledge. These language in use exercises are exemplified in Figure 13 below. All of Duolingo's language in use activities are gap-fill exercises, which means that the provided sentence contains a gap that needs to be filled by selecting the most appropriate word from a set of three or four options. Frequently, these exercises are accompanied by an illustration that provides scaffolding. In terms of grammar knowledge, learners are required to pick the correct answer based on word class (e.g. noun or adjective), or inflection (e.g. singular or plural). In terms of vocabulary knowledge,

learners need to choose the most suitable word. The primary objective of these exercises is to enhance learners' vocabulary and grammar knowledge through targeted practice.

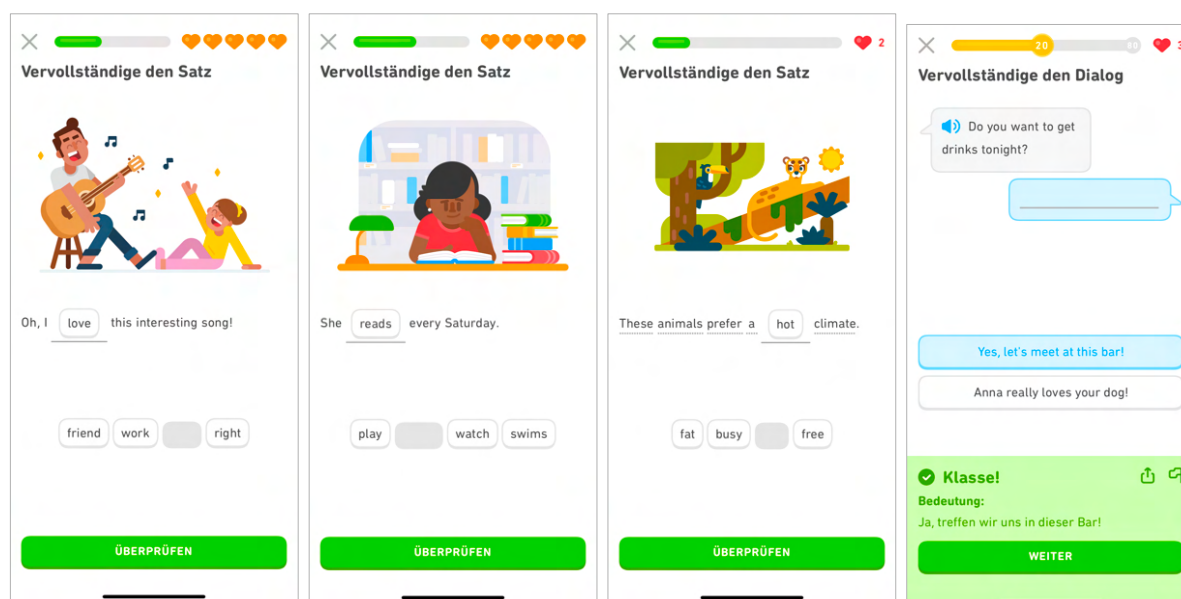


Figure 13 Examples of Duolingo's language in use and dialogue activities

In terms of its approach to building grammatical knowledge, Duolingo blends implicit and explicit grammar teaching methods. These two opposing approaches are also referred to as *non-interventionist instruction* (implicit) and *interventionist instruction* (explicit) (Hutz 2021: 149). While Duolingo predominantly adopts a non-interventionist approach to grammar instruction in its language learning activities, emphasising the subconscious acquisition of grammatical rules through exposure to language input, elements of the interventionist approach can be observed in the app's design. Firstly, Duolingo focuses on training learners' grammatical accuracy. Secondly, the app offers immediate and explicit error correction. Thirdly, many of the language learning activities resemble drill exercises that concentrate on practising new forms. Examples of Duolingo's explicit grammar teaching methods are illustrated in Figure 14 below. These include grammar tips, i.e. short grammar exercises, which are interposed when learners repeatedly make specific grammatical errors, basic sentences, which cover unit-specific grammar and can be accessed within the learning path, and smart tips, which provide explanations that appear after learners complete an exercise incorrectly.

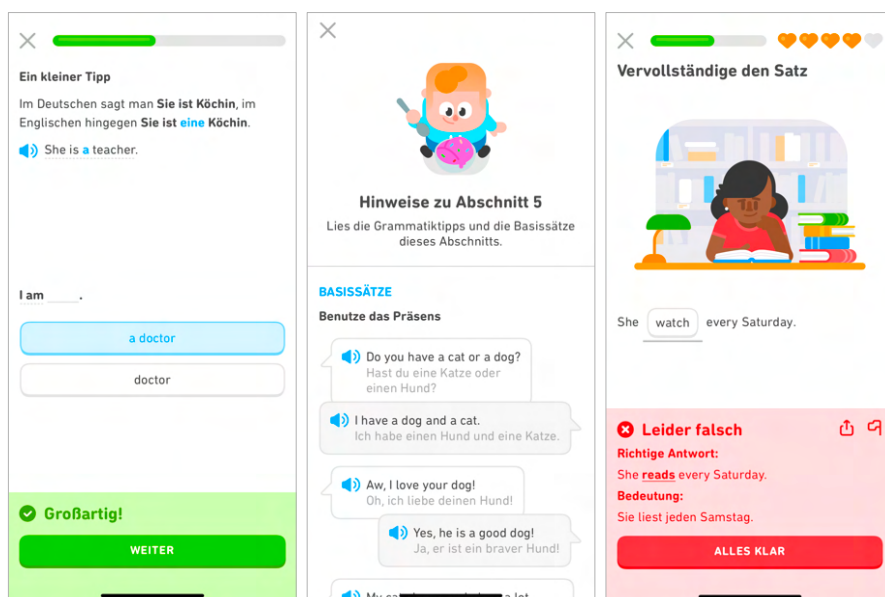


Figure 14 Examples of explicit grammar teaching methods in Duolingo

Dialogue activities. Duolingo also provides dialogue exercises, illustrated in Figure 13 above, as another exercise type. These activities involve presentation of a sentence in both written and spoken form, and learners are required to select the most appropriate answer from two options. The main focus of dialogue exercises is to train learners' reading and/or listening skills while also fostering the development of their vocabulary knowledge and understanding of pragmatics.

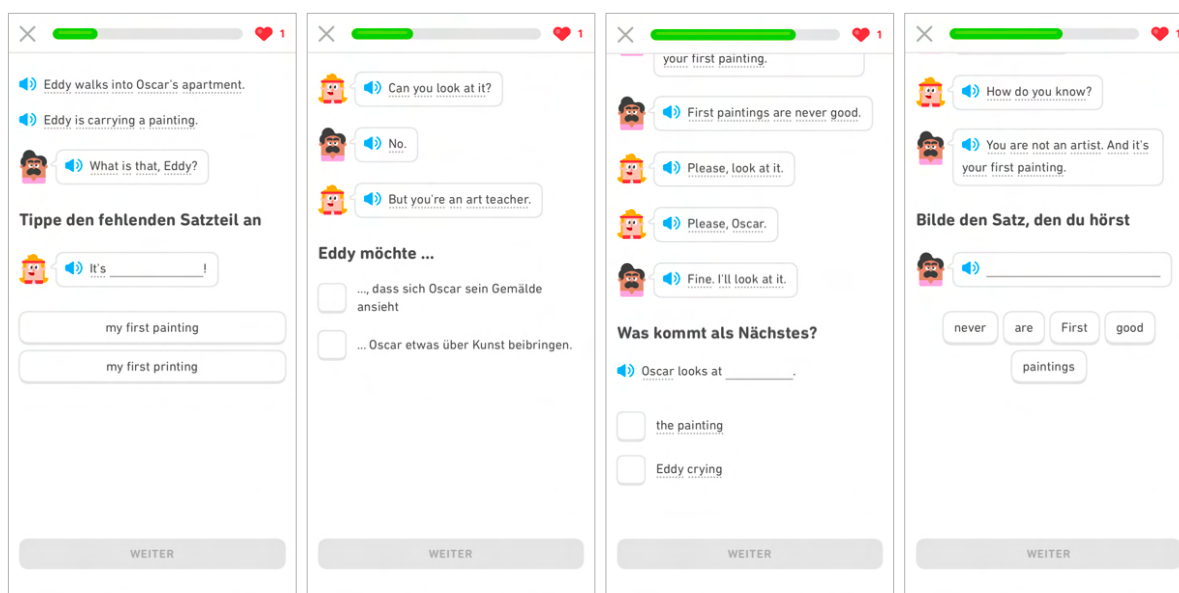


Figure 15 Examples of comprehension tasks in Duolingo's story activities

Story activities. Duolingo's story activities, exemplified in Figure 15 above, can be accessed as individual levels within the learning path. Each unit includes at least one, some feature two story activities. Once completed, story activities are added to learners' personalised

practice hub (barbell icon in the navigation bar) for future review. Story activities consist of short dialogues involving a minimum of two speakers engaging in conversations on various topics related to the respective unit. They are designed to train learners' listening and reading comprehension skills and include, but are not limited to comprehension tasks such as selecting the missing phrase, multiple choice questions about the conversation, predicting what comes next, reconstructing the heard sentence, or identifying the meaning of specific words. Each story activity concludes with a word pairs exercise, focusing on the key vocabulary presented in the dialogues. In contrast to Duolingo's other language learning activities, which are strictly limited to the sentence or even word level and arguably do not really facilitate learners' development of a communicative competence, story activities provide a more natural and authentic language input. They appear to align with the listening or reading comprehension tasks commonly used in a CLT approach. As of June 2023, story activities are included in Duolingo's English course right from the beginning. They progressively increase in difficulty as learners advance along the learning path.

(D5) Exposure to the Target Language. Duolingo applies a bilingual approach to teaching, which significantly limits learners' exposure to the target language. Throughout the app, learners' L1 is frequently employed. While certain exercise types tend to provide more exposure to the target language (e.g. story activities), others rely heavily on learners' L1 (e.g. translation activities). Considering that translation exercises account for a significant portion of the provided activities, target language exposure is limited in this regard. Furthermore, even in exercises that provide more TL input, instructions are solely given in the learners' L1. In addition, all other areas of the app, such as the newsfeed, rankings, option menu, etc. are dominated by the learners' L1. Consequently, Duolingo receives a rating of 2 points on dimension (D5).

Key principles. Duolingo's language learning programme is built on three key principles: implicit statistical learning, spaced repetition, and interleaving. Implicit statistical learning refers to "incidental learning of structured patterns encountered in the environment" (Conway 2020: 280), which is believed to play a significant role in language acquisition. Duolingo incorporates implicit statistical learning into its exercise design (Freeman et al. 2023), where learners immediately engage in language activities without initial lectures or explanations of underlying language rules. For instance, in Figure 16, multiple exercises across different levels of a unit focus on present simple tense 'yes/no' questions (*Do/Does +*

subject(s) + verb in base form), expecting learners to internalise the pattern as they move through the levels.

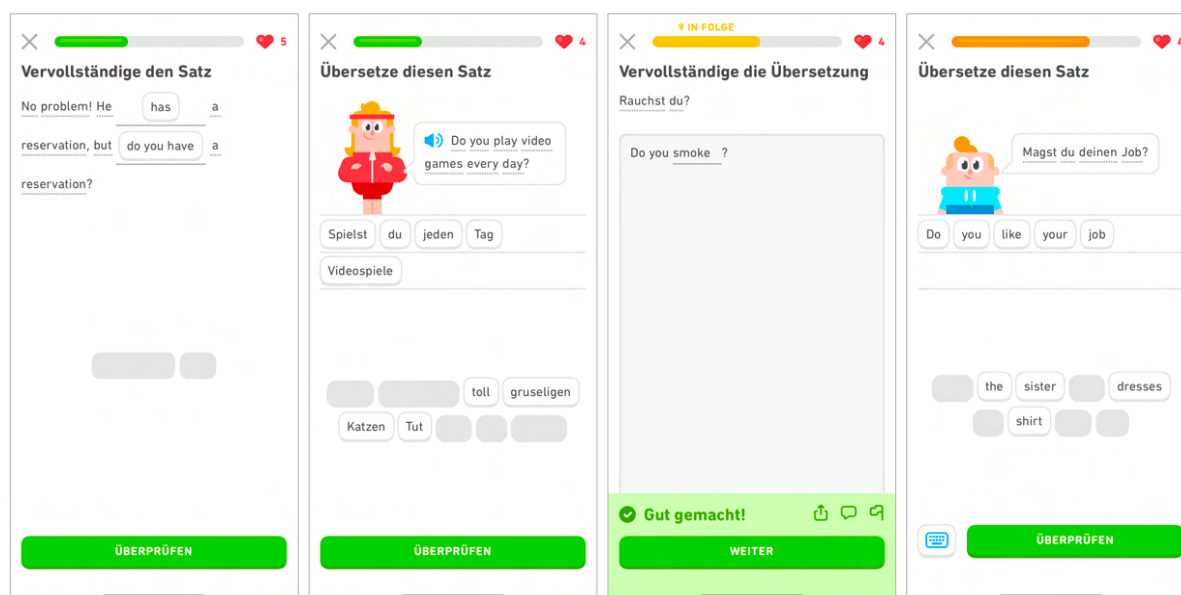


Figure 16 Examples of implicit statistical learning in Duolingo

The second principle, spaced repetition, refers to the systematic review of learning material at specific intervals. This strategy of review is proven to lead to better memorisation and recall outcomes compared to "massed practice" (Kang 2016: 13). Duolingo uses a spaced repetition algorithm called *half-life regression* (HLR), which merges "a psycholinguistic model of human memory with modern machine learning techniques" (Settles & Meeder 2016: 1855). In brief, Duolingo prompts learners to review recently learned content (i.e. grammar and vocabulary) more frequently, but gradually decreases the frequency of review as they demonstrate consistent knowledge and understanding of the material (Freeman et al. 2023: 5). The third principle, interleaving, refers to the alternation between two or more language skills in Duolingo's language courses. Instead of focusing on a single skill for an extended period, learners are exposed to a variety of language skills within a single lesson in a random or alternating manner.

Evaluation results. Results of the evaluation of Duolingo are shown in Table 28 below. It displays the app's individual scores awarded for each dimension, the app's tallied scores for each domain as well as the app's overall score.

Table 28 Evaluation results of Duolingo

Duolingo Evaluation							
A: Educational Value		B: Design		C: Engagement		D: Language Learning	
A1 Thinking Skills	3	B1 Ability to Save Progress	5	C1 Learner Control	3	D1 Input	4
A2 21st Century Skills	5	B2 Integration	3	C2 Interactivity	5	D2 Output	3
A3 Value of Errors	5	B3 Screen Design	5	C3 Pace	5	D3 Level of Material	5
A4 Feedback to Teachers	4	B4 Ease of Use	5	C4 Personalisation	2	D4 Skills	4
A5 Cooperative Learning	2	B5 Navigation	5	C5 Interest	4	D5 Exposure to Target Language	2
A6 Individual Differences	3	B6 Information Presentation	5	C6 Aesthetics	5		
		B7 Media Integration	5	C7 Gamification	5		
		B8 Cultural Sensitivity	3				
Total score per domain	22/30		36/40		29/35		18/25
Total score	105 pts. / 130 pts.						

As Table 28 exhibits, Duolingo receives 22 points for domain (A) *Educational Value*, 36 points for domain (B) *Design*, 29 points for domain (C) *Engagement*, and 18 points for domain (D) *Language Learning*, resulting in an overall score of 104 out of 130 points, which equals a percentage grade of about 80%. While Duolingo shows strengths in terms of its cohesive and appealing design and overall capacity to engage learners, i.e. through its heavily gamified language learning approach, its educational value and capacity to facilitate language learning perform on average or below average on certain dimensions.

Duolingo's biggest weaknesses are its lack of opportunities for real cooperation and collaboration between learners, its lack of opportunities for any real user personalisation as well as the lack of exposure to the target language. Regarding language learning, Duolingo makes bold claims: "[a]t Duolingo, we're developing our courses to get you to a level called B2, at which you can get a job in the language you're studying" (Duolingo 2023e). Put differently, Duolingo promises its users to be able to obtain a respectable level fluency in the target language.

Reviewing this claim, several studies have investigated Duolingo's effects on language learning. The first study to question the app's effectiveness, conducted by Vesselinov and Gregor in 2012, funded by Duolingo, found a statistical significant improvement of L1 English speakers' language proficiency in Spanish after using Duolingo for 30 hours over a two month period. The researchers estimated that it would take a beginner an average of 34 hours of Duolingo to cover the same material as a first semester college Spanish course. For years, Duolingo used these findings to claim its effectiveness online (Loewen et al. 2019:

295). Critiquing their research, Krashen (2014) notes that only 90 of the 156 participants finished the study, the participants were not typical college students, and the actual time spent using Duolingo varied significantly across the sample. Commenting on their estimation, Krashen (2014: 14) highlights the difficulty in comparing the study's results to conventional tertiary language instruction, as tertiary learners usually have different characteristics and constraints.

More recent studies, summarised in Shortt et al.'s (2023) systematic review of Duolingo research conducted between 2012 and early 2020, report mixed findings. While some studies found beneficial effects on different areas of learners' language performance, such as academic achievement (Abaunza, Martinez-Abad & Conde-Rodriguez 2019), vocabulary knowledge (Ajisoko 2020, Guaqueta & Castro-Garces 2018), listening skills (Bustillo et al. 2017), communicative skills (Armas Pesantez et al. 2019), performance on a holistic language test (Loewen et al. 2019), and writing skills (Salcedo Salcedo, Fernandez Morales & Duarte 2018), two studies (Rachels & Rockinson-Szapkiw 2018, James & Mayer 2019) report no significant differences in learners' language achievement and self-efficacy after using Duolingo. Owing to the studies' small sample sizes and a lack of control of confounding variables (e.g. gender, ethnicity, motivation, prior language proficiency, and others), Shortt et al. (2023) note that these findings should be treated cautiously and should not be used to make generalised claims about Duolingo's effectiveness for improving language proficiency. In addition to these studies, Alfuhaid (2021) and Jiang et al. (2021) report positive effects on the development of learners' speaking skills (Alfuhaid 2021) and their listening and reading proficiency levels (Jiang et al. 2021). While Alfuhaid's sample size is too small, Jiang et al. themselves caution against generalising their findings regarding Duolingo's overall effectiveness.

Revisiting Duolingo's claim about facilitating fluency, analysis of the app reveals that its language learning programme is not designed to help develop learners' overall communicative competence. While using the app might increase learners' knowledge of grammar and vocabulary or train their receptive skills, i.e. listening and reading proficiency, learners' productive skills, i.e. speaking and writing proficiency, are strongly neglected by Duolingo. Thus, it seems unlikely that learners will be able to develop their communicative competence to a level B2 by solely relying on Duolingo as a stand-alone language learning resource.

6.1.5. Conclusion

In conclusion, Duolingo is a prime example of an appealing and engaging language learning app that promises its users a holistic language learning experience resulting in a relatively high language proficiency. Unfortunately, the app does not deliver on these promises. Duolingo's language learning activities do not foster the development of learners' real communicative competence because of the limited, isolated, and partly invaluable language input, the limited possibility to produce output, and the app's focus on grammar and vocabulary knowledge. Despite Duolingo's attempt to interleave skills during lessons, only learners' receptive skills receive proper training. Aspects of learners' speaking proficiency (pronunciation) are fostered, while learners' writing skills are completely neglected. Although Duolingo is very intuitive in its use and provides a well-designed and attractive user interface as well as a highly gamified approach towards learning a language, which will likely speak to secondary level learners, using the app as a stand-alone resource will only benefit certain aspects of learners' language proficiency.

6.2. Mondly

Mondly, developed by ATi Studios and launched to the market in 2014 (cf. Jensen & Cadierno 2022: 4), has gained increasing popularity among language learners in recent years. As of February 2023, Mondly ranks as the 9th most downloaded language learning app globally (Statista 2023b). According to its official blog, Mondly surpassed 75 million learners worldwide in 2021 (Mondly 2021a), and it is anticipated that this number will continue to grow. In April 2022, Mondly was acquired by Pearson (Pearson 2022). The award-winning app (Mondly 2021b) offers language courses that cover 41 languages, such as British English, American English, Spanish, French, German, Russian, Japanese, Chinese and others (Mondly 2023a). Since its launch, Mondly has been accessible as both an online platform and a mobile application for devices using Android and iOS operating systems (Mondly 2023a). Similar to Duolingo, Mondly follows a freemium model, but its free content is considerably more limited and resembles a trial version. To access Mondly's full learning programme, learners have to subscribe to *Mondly Premium*, which involves a monthly or yearly subscription fee. In addition to its standard language learning app, Mondly also offers two innovative complementary apps. Firstly, there is Mondly VR (virtual reality), which provides learners with a virtual immersion experience by putting them in a simulated environment

that is used for language learning purposes. Secondly, there is Mondly AR (augmented reality), which utilises augmented reality technology to scan learners' surroundings and create a computer-generated imagery (CGI) teacher and animations, which are then used to enhance the learning experience. In the following sections, Mondly will be analysed and evaluated in the domains of (A) *Educational Value*, (B) *Design*, (C) *Engagement*, and (D) *Language learning*.

6.2.1. Educational value

(A1) Thinking Skills. With regard to the cognitive demand placed on learners, Mondly primarily facilitates the use of thinking skills of level 1 (*DoK-1*) and level 2 (*DoK-2*) of Webb's (1997) *Depth of Knowledge* framework. A more comprehensive examination of Mondly's exercise design, as well as the underlying teaching and learning approaches, methods, and principles, can be found in section 6.2.4. In summary, language learning activities in Mondly involve tasks such as memorising, recalling, and reproducing information, as well as recognising, predicting, or inferring language patterns and grammatical structures. Consequently, Mondly receives a score of 3 points on dimension (A1).

(A2) 21st Century Skills. If learners utilise Mondly as a tool for advancing their language learning, they will naturally develop various 21st century skills. Firstly, by taking charge of their own learning time, establishing appropriate learning objectives, and working independently, learners will cultivate their initiative and ability to self-direct their learning. Secondly, by accessing, utilising, and assessing information provided through a mobile app, learners will enhance their information literacy, media literacy, and technology literacy, particularly when they effectively employ the technology to support their language learning. Mondly often provides support and guidance to learners, such as prompting them to set daily learning goals and sending reminders through push notifications. Additionally, Mondly's exercise design promotes effective reasoning, thus strengthening learners' critical thinking and problem-solving abilities. As a result, Mondly is awarded 5 points in dimension (A2).

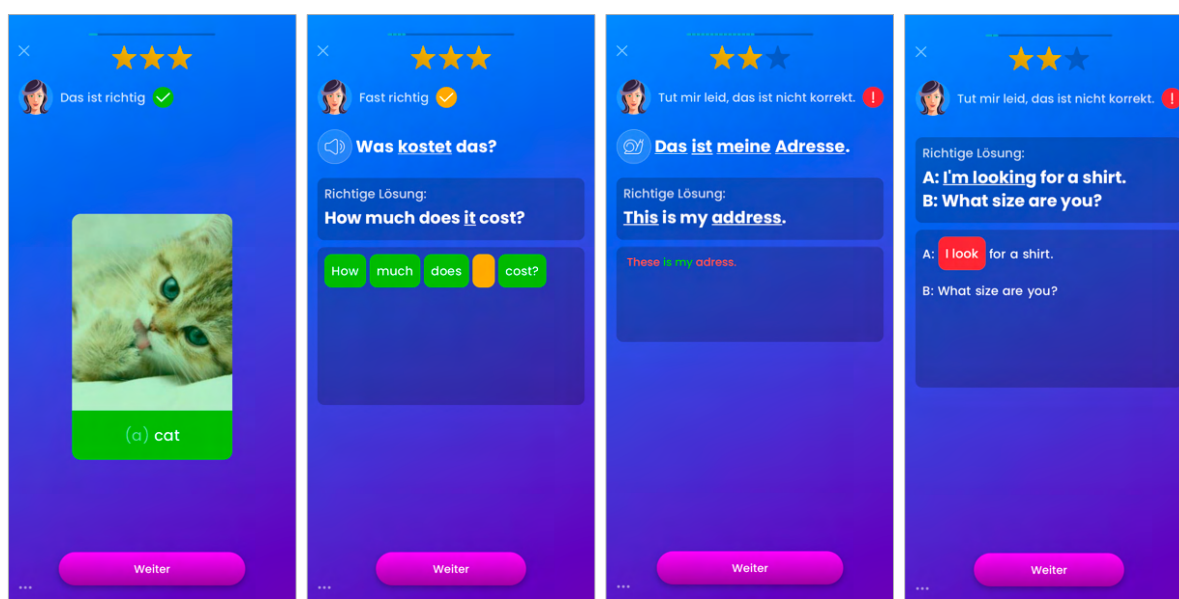


Figure 17 Examples of immediate feedback in Mondly

(A3) Value of Errors. Upon completing a language exercise, Mondly ensures that learners receive immediate and explicit feedback. If learners provide a correct answer, it will be highlighted in green, accompanied by the notification *This is correct*. In the case of a minor mistake, such as forgetting a preposition in a translation exercise, the missing word will be marked with an orange gap or box, and the response from the app will read *Almost correct*. If learners submit a completely incorrect answer, their response will be marked in red, and they will receive the notification *I'm sorry, this is not correct*. Additionally, Mondly will also provide the correct alternative for wrong answers. Examples of immediate feedback provided by Mondly are illustrated in Figure 17 above. Incorrect answers result in a deduction of stars awarded for the lesson. If learners lose all three stars, Mondly does not consider the lesson to be successfully completed. However, since there is no rigid learning path in Mondly, and all units and lessons are unlocked from the beginning, this does not affect learners' ability to proceed to subsequent lessons. Regarding speaking exercises, Mondly utilises voice recognition to analyse learners' spoken utterances and offer immediate feedback on pronunciation. However, it is important to note that the speech recognition feature is not fully developed and may not provide highly detailed feedback or accurately identify learners' pronunciation errors. At times, it may even fail to detect mistakes altogether. Unlike Duolingo, Mondly does not offer learners the opportunity to review previously answered language exercises that were marked incorrect at the end of each lesson. Therefore, Mondly receives 3 points on dimension (A3).

(A4) Feedback to Teachers. In contrast to Duolingo, Mondly does not provide a management layer for teachers where feedback regarding learners' achievements, mistakes, and progress is accessible. Consequently, teachers would need to have access to each learner's mobile phone and consult the app individually to gather information about their learners' progress. The app's statistics section offers an overall feedback on learners' progress, encompassing various information. This includes the learner's experience level (XP), the learning streak (indicating the consecutive days of completing learning activities), the amount of time spent on language exercises, the number of acquired lexical items, and the daily progress curve. However, since Mondly does not provide any direct feedback to teachers regarding learners' performance or progress, dimension (A4) cannot be applied in the evaluation of the app.

(A5) Cooperative Learning. Looking at cooperative learning, it is important to distinguish between the standard Mondly app and Mondly VR. In the standard language learning app, Mondly does not offer learners the option to connect and collaborate during the learning process or beyond. The only means of connection available is by following friends and observing each other's progress in the league ranking. However, Mondly VR introduces a multiplayer mode, which will be further discussed in section 6.2.4. This feature enables learners to meet peers, connect, and collaborate within the learning process. For the purpose of this evaluation, I consider Mondly VR to be a component of the Mondly learning package. Consequently, Mondly is awarded 4 points on dimension (A5).

(A6) Accommodation of Individual Differences. Mondly makes an effort to accommodate learners' individual differences and diverse learning styles. Many exercises in Mondly incorporate multiple modalities, combining written, spoken, and visual elements. This integration of different input modes serves as scaffolding and caters to various learning styles. In most exercises, Mondly combines text and sound by including recordings of words, phrases, or sentences that the exercise focuses on. These recordings are automatically played when the exercise is opened or can be accessed by tapping the speaker icon next to the item. Additionally, the recordings are available in two different speeds: a regular version and a slower version, which can be accessed by tapping the tortoise icon. Learners can also play the recordings multiple times. Mondly also offers 'hands-free' versions of its regular lessons, which primarily involve listen and repeat exercises and do not require learners to

interact by tapping buttons. With these differentiated instructional practices, Mondly receives 3 points for dimension (A6).

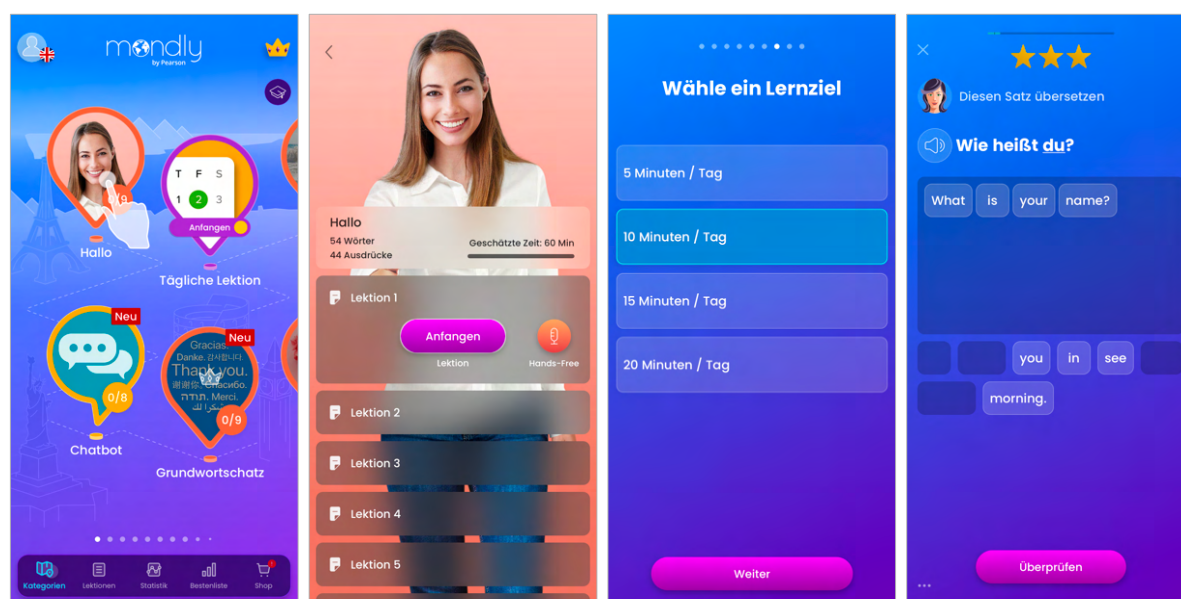
6.2.2. Design

(B1) Ability to Save Progress. Like Duolingo, Mondly only saves learners' progress if they create an account. If learners exit the app and go to their mobile phone's home screen while working on a lesson, they can resume from the exact point they left off upon reopening the app. However, if learners completely close the app while being in the midst of a lesson and then reopen it, they will need to restart the lesson from the beginning. Similarly, if learners exit a lesson by clicking the cross (X) icon, Mondly will warn them about the imminent loss of progress, necessitating a restart of the lesson. In terms of the capability to save progress, Mondly enables registered learners to automatically resume their last activity when they exit the app or manually choose their desired content. Consequently, Mondly receives a score of 5 points on dimension (B1).

(B2) Platform Integration. While Mondly does not officially and directly connect to other apps and independent websites, the app provides a rudimentary online community and employs e-mails to connect to its users. Mondly allows learners to follow their friends in the app and compete in league rankings. However, the app does not offer learners the space and tools for direct interaction and communication, e.g. through sending direct messages or posting on a message board. Mondly owns accounts on popular social media platforms such as *Facebook*, *Twitter*, *Instagram*, and *TikTok*, which are used to update learners on news and developments. In addition, if not deactivated in the settings menu, learners also receive newsletters on app updates or friendly reminders about set learning goals, challenges, etc. via e-mail. As Mondly provides a basic online community within the learning app and connects to its learners through e-mails, it receives 3 points on dimension (B2).

(B3) Screen Design. Overall, Mondly offers a visually appealing and well-organised graphical user interface (GUI). The app effectively integrates various design elements, including written text, photos, animations, shapes, colours, and sounds to enhance its language learning activities. Putting less emphasis on a colour-coded corporate identity, Mondly features a rather broad colour palette in its screen design. Vibrant colours, frequently complementary in nature, are employed to create a dynamic and energetic feel. While younger learners might be particularly attracted by the choice of colours, they may cause

eye strain during prolonged use. Rather than employing organic shapes, Mondly's GUI predominantly features rectangular, geometrical shapes with rounded edges. In addition, the app frequently incorporates high-quality photos that enhance the overall aesthetic. Besides the high quality of individual visual design aspects, Mondly manages to integrate all design constituents successfully, which enhances its language learning content. On dimension (B3), Mondly therefore, receives 5 points.



(a) Content map (b) Unit structure (c) Simple instructions (d) Interactive elements
Figure 18 Elements contributing to ease of use and navigation in Mondly

(B4) Ease of Use. Learners will be able to use Mondly without any guidance, as the app is designed to be intuitively accessible. When beginning a language course on Mondly, learners will be guided through a step-by-step onboarding process that involves selecting the target language, indicating their language proficiency level, and setting their daily learning objective. The task instructions and prompts found within the app are clearly articulated and utilise simple language. Learners are provided with progress bars for individual lessons and activities, which offer them a visual representation of their advancement. Mondly also relies on learners' intuitive understanding when it comes to elements that require their interaction. This is achieved through the use of drop shadows, which appear behind clickable buttons, and empty blue spaces, which indicate that an element has been successfully moved. These design cues serve to signal to learners that their engagement is necessary. Figure 18 above provides examples of some of these intuitive elements. As secondary level learners can access Mondly intuitively without the need for guidance, the app is awarded 5 points on dimension (B4).

(B5) Navigation. Mondly incorporates user-friendly features such as an intuitive navigation bar, a well-organised content map, and clearly structured units, all of which facilitate easy navigation within the app. These features are illustrated in Figure 18 above. Mondly's navigation bar, conveniently placed at the bottom of the screen, enables learners to easily move between different areas in the app, which are labelled and indicated by simple but expressive icons. Mondly's language courses are designed as content maps, offering learners a comprehensive overview of available units and additional activities. Furthermore, within each unit, the content is presented in a list format with numbered lessons. As Mondly provides learners with features that allow for easy navigation through the app's content, it receives 5 points on dimension (B5).

(B6) Information Presentation. Overall, Mondly presents its content in a logical manner. The content map, illustrated in Figure 19 below, provides learners with an overview of the available units. Mondly's English course comprises 36 regular units, 4 units aimed at younger children, and 15 units focused on English for Specific Purposes (ESP). In addition, the content map provides access to Mondly's chatbot function, described in more detail in section 6.2.4., and Mondly's daily lessons. Units comprise different topic areas (e.g. family, travelling, holiday activities, food & drinks, pets, shopping, etc.) and aim at specific learning objectives (e.g. booking a hotel room, asking for directions, etc.). They cover thematic vocabulary and grammar. Each unit encompasses 6 regular lessons, one conversation lesson, one vocabulary lesson, and a Pearson progress test. Each lesson, in turn, contains around 28 individual language exercises. Since Mondly does not feature a rigid learning path, lessons within units are unlocked from the beginning and can be completed in any order. Mondly takes a different approach compared to Duolingo by not relying on a machine learning model to offer personalised language exercises within lessons. Instead, the difficulty level of the language exercises is designed to gradually increase as learners move through the sequence of lessons. Additionally, while new vocabulary and grammar are introduced in the initial lessons of a unit, the subsequent lessons primarily focus on practicing and reinforcing these new elements. However, since all lessons within a unit, as well as all units themselves, are accessible to learners right from the beginning, it is highly recommended that they follow the suggested lesson sequence provided by the app. This ensures that learners can fully experience the planned progression in difficulty as they advance from one lesson to the next. To conclude, Mondly provides a good overview over its available content, organises its

course into well-structured units, and ensures a gradual increase in difficulty as learners progress from one lesson to the next. Consequently, Mondly receives 5 points on dimension (B6).



Figure 19 Examples of units accessible through Mondly's content map

(B7) Media Integration. Mondly achieves an educationally valuable and coherent learning experience by seamlessly integrating all of its media components. Written texts, graphics, photos, CGI objects, speech, and sounds are effectively combined and aligned. An analysis of Mondly's language course reveals a consistent correlation between written text and recorded speech, as well as written text and graphics. Discrepancies or disjointedness among the media components are not observed throughout the course. This consistent integration of media elements greatly enhances language activities and contributes to the overall quality of the app. Consequently, on dimension (B7), Mondly is awarded 5 points.

(B8) Cultural Sensitivity. Mondly only offers limited representation of diverse populations and cultures, but it does offer opportunities for learners to connect with peers from different communities around the world. Examples of diverse populations, as illustrated in Figure 20 below, are depicted in a few photos used in Mondly. In terms of linguistic diversity, Mondly provides two options. Learners can choose between British English and American English when selecting their preferred language course. Depending on their selection, learning activities then follow the spelling and pronunciation rules and conventions of either British English or American English. Similar to Duolingo, Mondly covers rather generic topics and language, however, there is a noticeable tendency towards Western culture. Exploring

different cultures explicitly through Mondly's content is not possible, and the app does not incorporate authentic pictures, videos, or texts from around the world. Nevertheless, Mondly VR's multiplayer mode enables learners to connect with peers worldwide, allowing for the sharing of ideas and the transfer of cultural knowledge. While Mondly scores 5 points on dimension (B8), indicating its efforts towards cultural sensitivity, there is room for improvement in terms of explicit cultural exploration and the use of authentic pictures, videos, or texts.

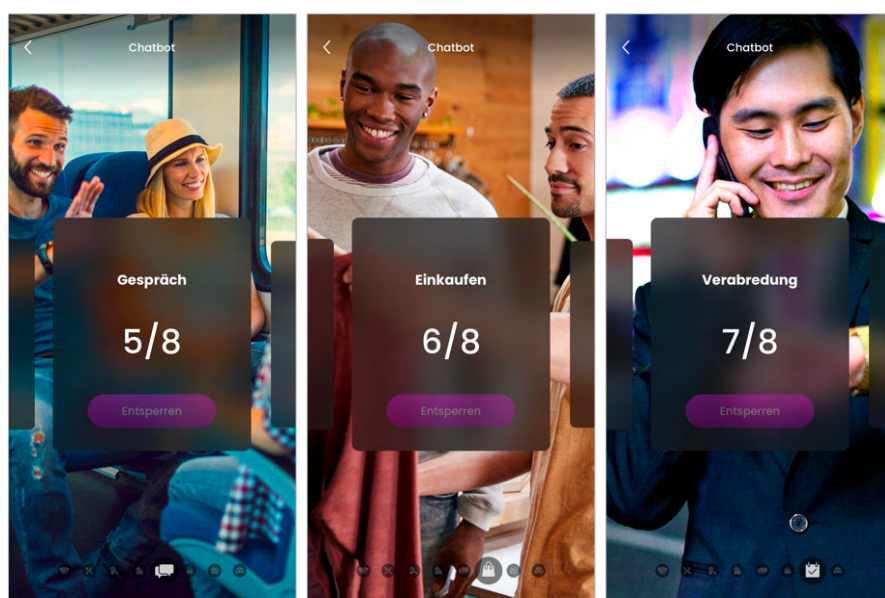


Figure 20 Examples of representation of diverse populations in Mondly

6.2.3. Engagement

(C1) Learner Control. Mondly allows learners to select the level where they will begin to engage with the app's content. During the onboarding process, learners are asked to specify their language proficiency level, with options including *beginner*, *intermediate learner*, and *advanced learner*. If learners discover that the selected difficulty level does not align with their actual language proficiency, they can adjust the difficulty level at any time, without losing their progress in the learning path. Once the onboarding process is completed, learners gain immediate access to all units available in Mondly's learning path. In contrast to Duolingo, Mondly does not impose a linear progression, where learners are required to complete units sequentially. Instead, learners are free to select and engage with any desired content, even if it is located further ahead in the learning path. Furthermore, within units, learners are not required to complete lessons in a fixed order. As Mondly gives learners

complete control over the level at which they will begin to engage the content, the app is awarded 5 points on dimension (C1).

(C2) Interactivity. Similar to Duolingo, Mondly follows a "learn by doing" approach (Lăpuşneanu 2021), which entails that the educational content covered by the app is predominantly delivered to learners through a variety of different language learning exercises. These will be further examined in section 6.2.4. As Mondly continuously encourages learners to engage with the content, the app receives 5 points on dimension (C2).

(C3) Pace. Mondly provides learners with complete control over their learning pace within the app's content. To access individual units, learners must tap on the respective unit, and to commence a lesson or learning activity, they actively click on the *start* button. Furthermore, learners are required to tap the *check* button to assess their answers and the *continue* button to progress to subsequent language exercises. In terms of settings, Mondly offers the option to enable 'auto check answers' and 'auto continue'. When these settings are activated, learners' answers are automatically evaluated without the need for manual button clicks, and they seamlessly transition to the next language exercise after completing the current one. Both of these settings are initially deactivated and learners must proactively enable them in the settings menu. However, even when both settings are turned on, the pacing remains suitable for secondary level learners. Moreover, Mondly does not incorporate any learning activities with time limitations, ensuring that learners can progress at their own preferred pace. As learners have full control over the pace of content presentation, Mondly is awarded 5 points on dimension (C3).

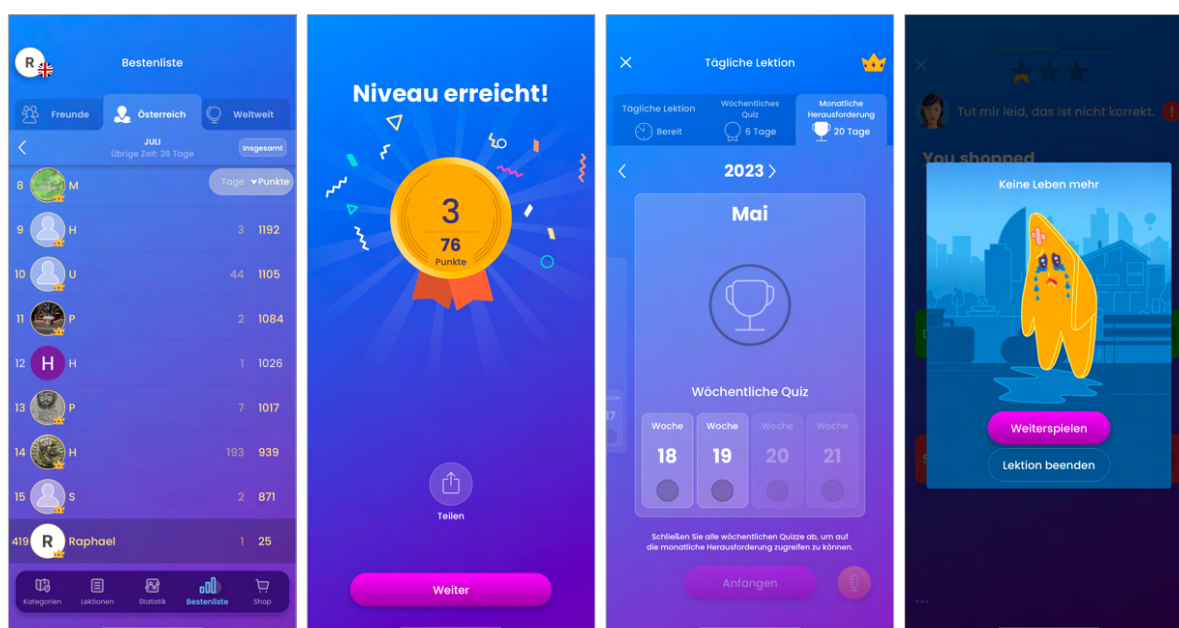
(C4) Personal Preferences. Mondly provides learners with limited opportunities for personalisation, primarily focused on settings rather than proper app features. With regard to proper app features, learners are only allowed to upload and change their profile picture (avatar). The app does not grant learners the option to customise other app elements, such as the background music, the colour scheme, the font type and size, the navigation menu, the interface language, and others. Nevertheless, Mondly does provide personalisation options within app settings. Learners are able to set daily learning reminders and switch sound effects and haptic feedback on or off. Additionally, they possess control over specific automated functionalities, such as the automatic reading of words, phrases, and sentences upon accessing learning exercises (auto play voice), the automatic evaluation of answers

without the need to tap the check button (auto check answer), and the automatic transition to subsequent learning exercises without requiring the tapping of the continue button (auto continue). Given that Mondly only permits learners to personalise one proper app feature, it is awarded 2 points on dimension (C4).

(C5) Interest. Mondly mostly offers language learning material that is designed to be interesting and engaging for secondary level learners. The app aims to cater to a relatively broad audience of learners, resulting in a wide range of topics and activities. It is worth noting that many of the topics covered in Mondly's regular units align with those found in approved traditional EFL textbooks used in Austria. Additionally, secondary level learners may find particular excitement in the innovative affordances and features provided by Mondly AR and Mondly VR. Taking these factors into consideration, Mondly receives 4 points on dimension (D5).

(C6) Aesthetics. Mondly's user interface positively contributes to learners' motivation and overall language learning experience because it is thoughtfully designed and visually appealing. What distinguishes Mondly from other language learning apps, such as Duolingo and its lively animations, is the app's frequent use of captivating high-resolution photos that seamlessly blend with the rest of the graphical user interface. These real-life photos play a prominent role in enriching language learning activities, such as translation or vocabulary exercises, and other areas of the app. Moreover, Mondly skillfully and effectively integrates a vibrant colour palette into its interface design, maintaining a high quality appearance without looking low-budget. Overall, Mondly's cohesive, appealing, effective, and efficient user interface positively impacts learners' motivation to engage with the app. Consequently, Mondly receives a score of 5 points for dimension (C6).

(C7) Gamification. In contrast to Duolingo, Mondly adopts a less gamified approach towards language learning. Nonetheless, Govendor & Arnedo-Moreno (2020: 676) identified twelve elements of gamification in Mondly's language learning programme. These gamified components, illustrated in Figure 21 below, are, for example, learner levels, lives (stars), experience points (XP), league rankings, learning streaks, an avatar, and others. Believed to increase learners' extrinsic motivation and engagement, Mondly's twelve elements of gamification earn the app a score of 5 points on dimension (C7).



(a) League ranking

(b) Learner levels

(c) Monthly challenge

(d) No lives left

Figure 21 Elements of gamification in Mondly

6.2.4. Language learning

(D1) Quality of Input. Although input in Mondly will be mostly comprehensible, the relevance of topics and language for secondary level learners demand questioning. As Mondly does not utilise a machine learning model to analyse learners' performances, and therefore cannot provide personalised language exercises tailored to each learner's language proficiency level, comprehensibility of language input partially depends on learners' ability to select the most fitting difficulty level. This may initially be challenging for younger and less experienced learners. However, the mostly generic and thus rather comprehensible content in Mondly's 36 regular units, as well as the possibility to change to a different difficulty level at all times without any loss of progress, balances out this issue. In addition, Mondly sometimes facilitates comprehensibility of language input by including text-accompanying images, illustrations, or animations. The topics covered in regular units (e.g. family, weather, travelling, sports, etc.) will likely activate parts of learners' background knowledge, which might also help to make language input more comprehensible.

Nevertheless, questions arise regarding the relevance of language input when we turn towards Mondly's English for Specific Purposes (ESP) content. Besides its 36 regular units and 4 units specifically aimed at younger children (topics: animals, fruit and food, colours and numbers, body parts), Mondly offers a total of 15 ESP units that aim at adult learners working in the corporate sector. These units cover issues such as business communication,

business meetings, business negotiations, marketing, human resources, etc. and will not be relevant to most secondary level learners. However, most of these units could be relevant to upper secondary level students attending higher vocational schools for business and commerce (i.e. Austrian *Handelsschulen* or *Handelsakademien*).

Another aspect influencing the input's comprehensibility and relevance is Mondly's strong focus on high-frequency vocabulary (Lăpuşneanu 2021). In accordance with Nation's proposal that teachers and learners should allocate enough time and attention to the "incidental learning, direct teaching, and direct learning" of high-frequency vocabulary (2022: 28), Mondly centers its language exercises around high-frequency words embedded in everyday phrases and sentences (e.g. *How much is the shirt?* or *I would like a tea, please!*). This strong focus on essential and relevant language input for everyday use distinguishes Mondly from its competitor Duolingo. The quantity of target language input could be increased in regular lessons, as words and phrases are sometimes presented in isolation without sufficient context. However, longer sequences of connected language input are provided in conversation lessons, chatbot conversations, and conversation activities in Mondly AR and Mondly VR. Considering the overall comprehensibility and relevance of Mondly's language input, the app is awarded 4 points on dimension (D1).

(D2) Opportunity to Produce Output. Although Mondly provides learners with ample opportunities to produce written and spoken output, the app's language exercises, analysed and evaluated in section (D4) *Skills*, impose restrictions on the length and freedom of learner output. Written output is always restricted to typing single words, phrases, or sentences, and never extends beyond the sentence level. While learners are allowed to produce spoken output more freely, the nature of this type of output is frequently confined to a predetermined set of responses. Spoken output reaching beyond the sentence level, especially during chatbot conversations, conversation lessons, and role-play activities in Mondly AR and Mondly VR is still expected to stay within a predefined conversation framework. The most promising opportunity for unrestricted and extended output is presented by Mondly VR and its multiplayer mode, where learners can gather in a virtual room and engage in target language conversations. As a result of the app's restrictions regarding learner output, Mondly receives 3 points on dimension (D2).

(D3) Level of Language Learning Material. Mondly offers three different difficulty levels (beginner, intermediate learner, advanced learner) for learners to choose from. These are said to correspond with CEFR levels A1-B2 (Mondly 2023a). However, an analysis of the app reveals that the increase in difficulty between these three levels is minimal. Moreover, Mondly does not specify how the difficulty levels align with CEFR levels. Unlike Duolingo, Mondly does not employ a machine learning model that can predict personalised and appropriate task difficulty levels based on learner's prior performances in language activities. Instead, Mondly's lessons are preprogrammed, fixed learning units, which do not allow for adapted, personalised language exercises according to learners' performance. Therefore, learners are left to select the appropriate difficulty level on their own, which is problematic due to the lack of information regarding the relationship between app difficulty level and CEFR level. This can particularly pose difficulties for younger learners. Consequently, Mondly receives 3 points on dimension (D3).

(D4) Skills. While Mondly covers all four language skills to some extent, the app does not equally prioritise all of them. In comparison to listening and writing skills, which are never holistically trained in language exercises, Mondly puts a stronger emphasis on developing learners' reading and speaking skills. Similar to Duolingo, not all aspects of a skill are necessarily covered comprehensively by the app's language exercises. For instance, many exercise types in Mondly include a TL recording of the word, phrase, or sentence that the activity is based on. Learners can use these recordings as scaffolding or support to successfully complete the task at hand, nonetheless, none of the activities provided explicitly train learners' listening skills in a more focused sense. Similar claims can be made regarding learners' writing skills. Along with the four skills, Mondly integrates the teaching of grammar and vocabulary knowledge to a high degree. On dimension (D4), Mondly therefore scores 3 points. To gain a better understanding of how each individual skill is trained by the app, the various types of language learning exercises shall be analysed in detail.

Types of exercises. As of June 2023, lessons in Mondly's English course encompass translation activities (a-f), language in use activities (g-h), vocabulary activities (i), and speaking activities (j). Translation activities and language in use activities are the most common types of exercises employed in lessons, while the other two types of exercises occur less frequently. Apart from regular lessons which comprise a mix of exercises, Mondly also offers focused lessons for each unit. The two types are vocabulary lessons and

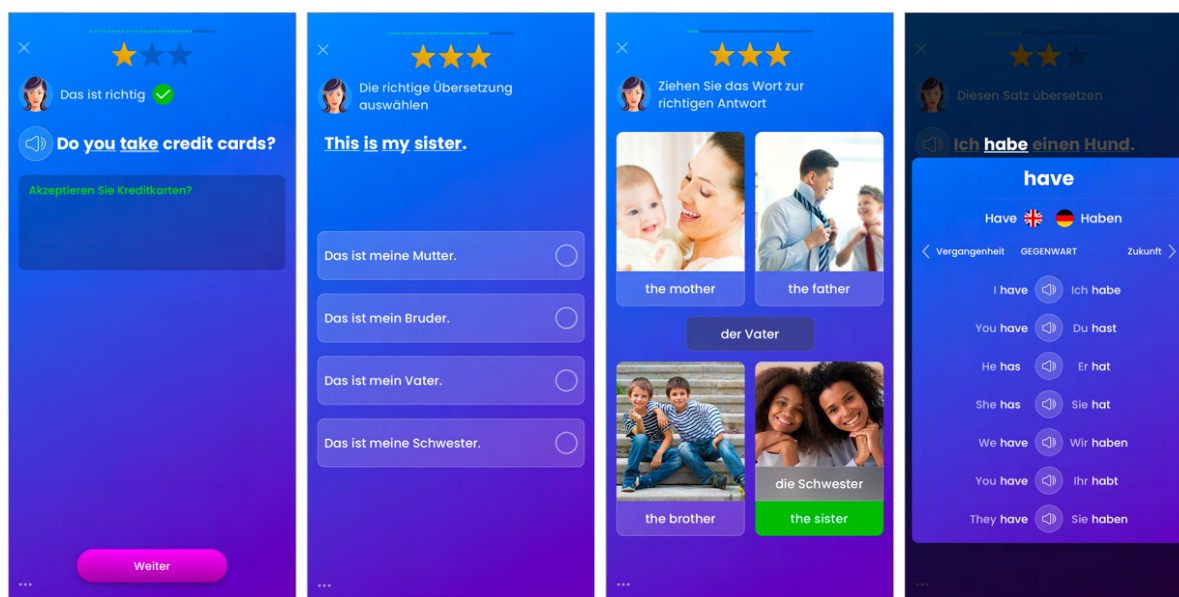
conversation lessons. In addition to its units, Mondly comprises a chatbot, which simulates human-like conversations with learners. A complete list of the 10 different types of activities, the two focused lessons, and the chatbot is provided in Table 29 below.

Table 29 Types of activities in Mondly

Translation activities	
(a) Translate the sentence	Learners are given a sentence in the L1 or TL and must translate it by typing it out or by arranging words to form the sentence.
(b) Complete the translation	Learners are given a sentence in the L1 or TL and must complete the translation. They might be asked to complete the blank by selecting the missing word, by arranging the letters to form the missing word, or by typing out the missing word.
(c) Choose the correct translation	Learners are given a word or sentence in the L1 or TL and must choose the correct translation out of four options.
(d) Translate what you hear	Learners listen to the recording of a sentence in the TL and must translate it by typing it out or by arranging words to form the sentence.
(e) Drag the word to the correct picture	Learners are given 4 words in the L1 and must drag them to the correct pictures. The pictures have the words' TL translation attached.
(f) Swipe up or down	Learners are given a flashcard comprising the L1 word and two pictures (top half and bottom half). TL translations are provided with the pictures. Learners are asked to swipe up or down to select the correct translation.
Language in use activities	
(g) Complete the sentence	Learners are given a single sentence or a two-line dialogue in the TL which contains a blank. Learners must select the missing word or phrase out of a few options, or arrange letters to correctly form the missing word or phrase.
(h) Form the sentence	Learners must arrange given words to form a correct sentence in the TL.
Vocabulary activities	
(i) Choose the correct word	Learners are given a definition in the TL and must choose the correct word out of 3 options.
Speaking activities	
(j) Repeat the sentence	Learners listen to the recording of a sentence in the TL and record themselves while they repeat it aloud.

Full lessons with special focus
Vocabulary lessons Learners complete (slightly altered) exercises of the type (f) swipe up and down.
Conversation lessons Learners listen to a conversation between two speakers and record themselves while they repeat each sentence aloud.
Activities outside of units
Chatbot Learners engage in a simulated human-like conversation with a bot. Learners' responses can be written or spoken.

Translation activities. Translation exercises, illustrated in Figure 22 below, are the most commonly used activity type in Mondly. Mondly follows a similar approach to Duolingo by blending principles of the Grammar Translation Method (GTM) and the Audiolingual Method (ALM). Following GTM principles, Mondly utilises learners' L1 for giving instructions or providing clarification (e.g. translations of TL words). Moreover, translating happens in both directions, which demands continual transfer of grammar knowledge between both languages, and usually focuses on isolated, single sentences. In addition to L1 translations for content words, which can be accessed by tapping the unknown item, Mondly also provides verb lists. Illustrated in Figure 22 below, these are shown when marked verbs are tapped and they offer learners an overview of the past, present, and future verb conjugations in the L1 and TL. In terms of accuracy of translations, Mondly is rather strict, as only one correct translation is listed for each content word in an exercise. Possible translations, however, are subject to change according to the contextual meaning of the word or phrase in the sentence. For instance, Mondly will provide different translations for homonyms according to the sentence context (e.g. *I need to address this issue.* – *ansprechen*, *Can you send me your address?* – *Adresse*). Compared to some of Duolingo's flawed translations, Mondly seems to be more consistent and accurate in this regard. In terms of spelling mistakes, Mondly is stricter than Duolingo. If learners mistype a word in a minor way (e.g. two letters are switched around, one letter is forgotten, etc.), Mondly rejects the answer, and learners lose (parts of) one of the three stars available for each lesson.



(a) Exercise type a

(b) Exercise type c

(c) Exercise type e

(d) verb list

Figure 22 Examples of Mondly's translation activities and verb lists

Applying ALM principles in translation activities, Mondly keeps grammatical explanations to a minimum. Even more than Duolingo, Mondly follows a mostly inductive and implicit approach: instead of providing to be memorised grammar rules or explanations in any form, learners are supposed to discover or deduce language patterns and structures from examples and apply these in later (translation) exercises. Translation exercises mostly follow typical ALM activity formats including replacement activities, inflection activities, and restoration activities. Furthermore, Mondly incorporates the ALM key principles of drill and repetition. Similar to Duolingo, Mondly incorporates "a unique spaced repetition system" (Lăpuşneanu 2021), meaning that the app presents new, more challenging words or phrases alongside previously learned lexical items. In contrast to Duolingo, Mondly does not disclose any specifics regarding the functionality of their spaced repetition algorithm. While the app provides immediate, explicit and corrective feedback when learners make a mistake, it is lacking positive reinforcement (e.g. verbal praise) after successful completion of language exercises. Instead of merely notifying learners that their answer is correct (*That is correct*), Mondly could provide words of praise (e.g. *Well done!*, *Good job!*, or others) to positively reinforce the formation of correct habits, as is advocated by the ALM.

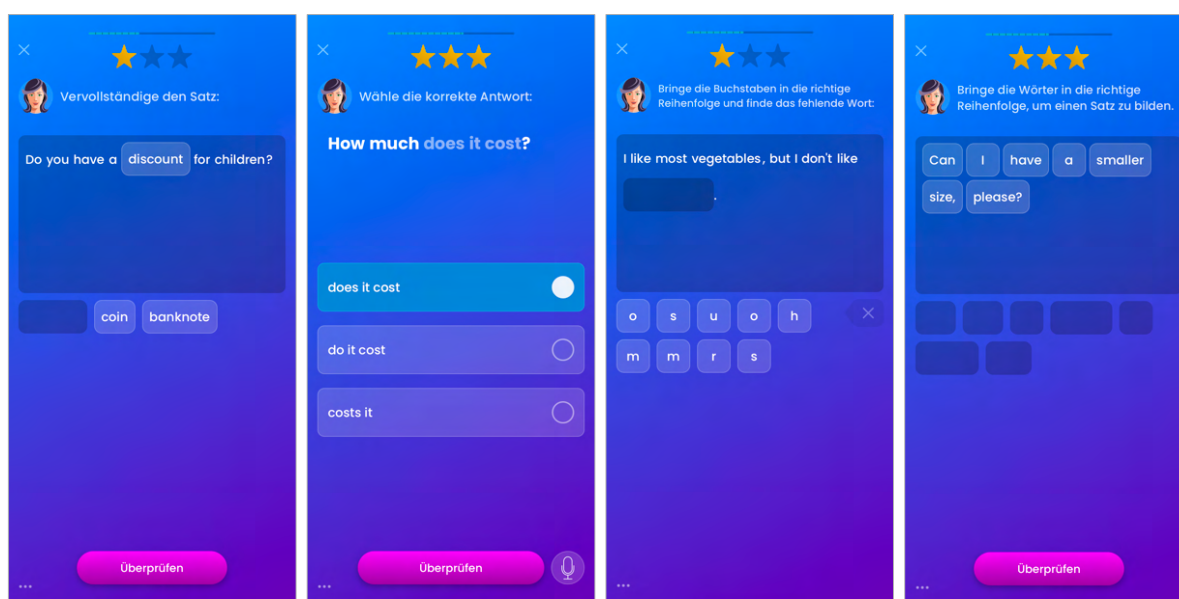


Figure 23 Examples of Mondly's language in use activities

Language in use activities. Mondly's language in use exercises (g-h), illustrated in Figure 23 above, bear resemblance to Duolingo's language in use activities, as they center around evaluating learners' understanding of vocabulary and grammar. These exercises are provided in two formats. Firstly, there are gap-fill exercises, which require learners to select the missing lexical item from a number of possible options or correctly arrange letters to form the missing word or phrase. Secondly, there are restoration exercises, which demand learners to rearrange words to form a correct and meaningful sentence. Mondly's language in use exercises assess learners' grammar knowledge about word class, inflection, tense, and word order as well as learner's overall vocabulary knowledge. In contrast to Duolingo, Mondly does not provide any scaffolding, e.g. through sentence-accompanying illustrations. In general, the primary objective of language in use exercises is to enhance learners' vocabulary and grammar proficiency through targeted practice opportunities.

Vocabulary activities. Vocabulary activities, illustrated in Figure 24 below, are sporadically integrated into Mondly's regular lessons. These exercises follow a word-definition matching format where learners are prompted to choose the corresponding lexical item for a given definition or select the appropriate definition for a given lexical item from a set of options. These exercises train learners' knowledge of vocabulary in terms of comprehending the connection between form and meaning of specific words and phrases.

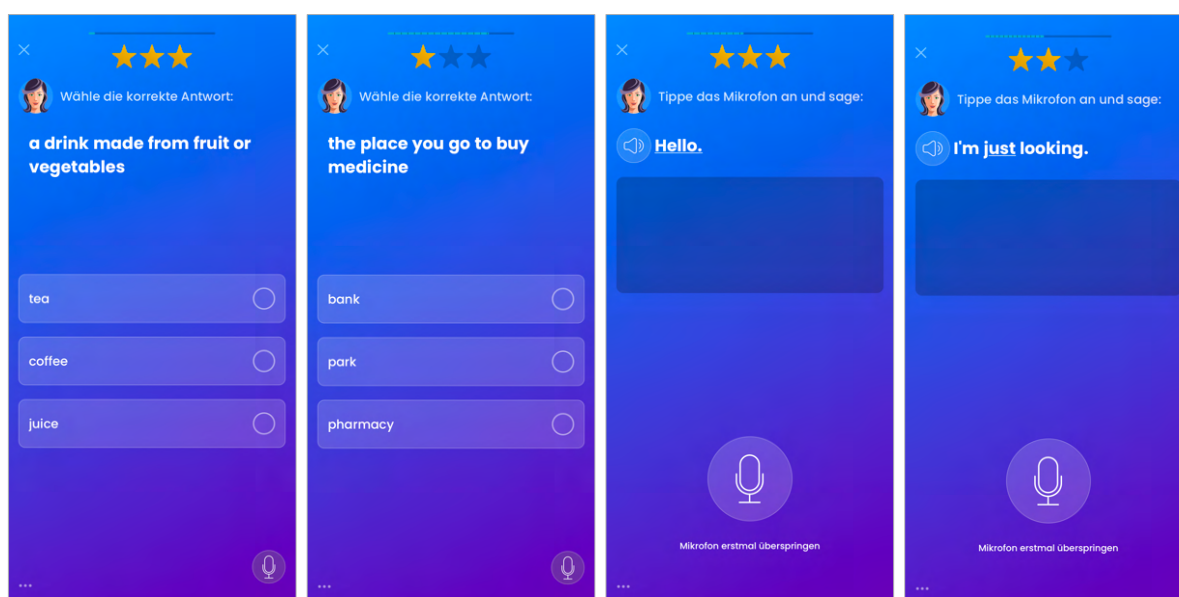


Figure 24 Examples of Mondly's vocabulary and speaking activities

Speaking activities. Within regular lessons, speaking activities, exemplified in Figure 19 above, follow a listen-and-repeat format, in which learners are asked to repeat a previously read aloud sentence while they record themselves. These exercises resemble imitative repetition drills used in the Audiolingual Method, and solely focus on training learners' pronunciation accuracy. Learners' pronunciation is assessed by a speech recognition tool, which Mondly does not provide specifics for. The feedback by the app is minimal, as it only indicates which words were pronounced (in)correctly. These exercises share similarities with Duolingo's speaking activities, but there is one key difference: learners are provided with a transcript of the target sentence, offering scaffolding from the start, and thus making the task easier. Moreover, the provision of a transcript interferes with the development of learners' listening skills since they are not required to comprehend the audio recording played in the exercise. Additionally, the exercises are limited to the sentence level, and the absence of free production by the learners means that spoken fluency, an essential aspect of real communicative competence, is not fostered at all.

Focused lessons. In addition to regular lessons, Mondly offers specialised lessons that focus on vocabulary knowledge and conversational skills. Each regular unit typically includes at least one vocabulary lessons and one conversation lesson, while special units that target English for Specific Purposes (ESP), such as business, may include two vocabulary and two conversation lessons. The vocabulary lessons in Mondly, illustrated in Figure 25 below, consist of a specific exercise type that is similar to exercise type (f) *swipe up and down* in terms of its format. Essentially, these exercises require learners to match L1 lexical items to

their translations in the TL. However, visual aids such as illustrations of the vocabulary items are not provided. Despite being labelled as vocabulary practice, these lessons primarily consist of translation exercises.

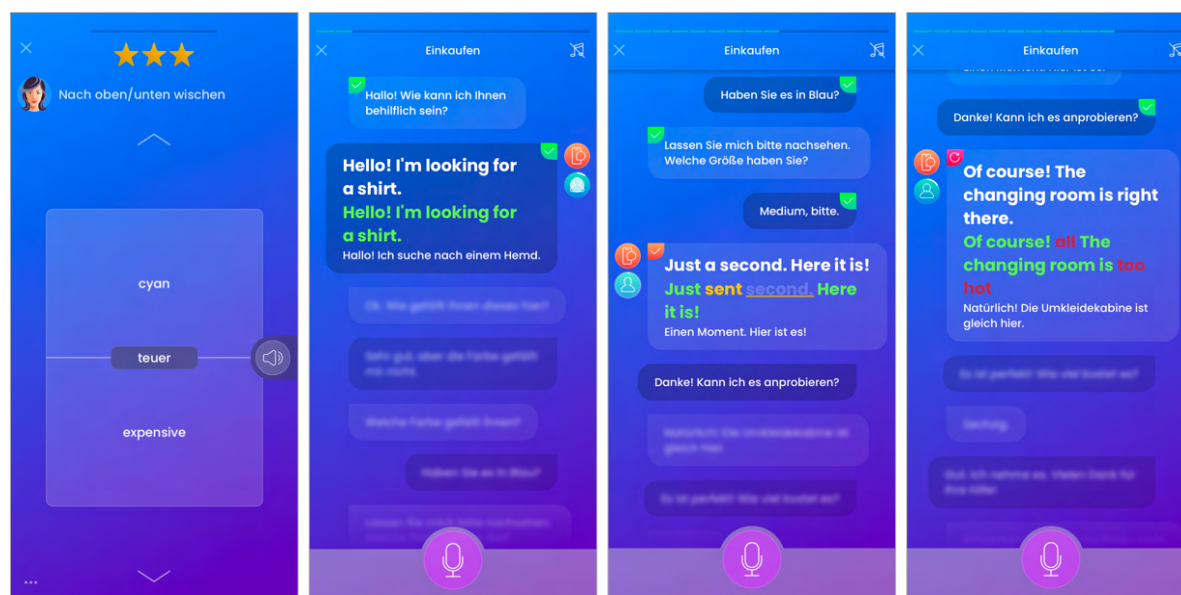


Figure 25 Examples of Mondly's vocabulary lessons and conversation lessons

The second type of focused lesson is the conversation lesson, as shown in Figure 25 above. These lessons involve a dialogue between two speakers on a specific topic related to the unit (e.g. going shopping). The written transcript of the dialogues is provided in both the learners' L1 and the TL. After each speaker finishes their part, learners are prompted to repeat the utterance aloud while recording themselves. They can listen to their own recordings at two different speeds (regular and slow speed). Mondly's speech recognition tool analyses the learners' recordings and provides feedback. Correctly spoken sentences allow the learners to progress in the dialogue, while partially or completely incorrect sentences need to be repeated. Mistakes made by learners are marked either in orange (indicating a wrong word but with relatively close pronunciation to the target word) or in red (indicating a wrong word). However, it should be noted that Mondly's speech recognition software is not always accurate and often fails to recognise incorrectly pronounced words unless they are significantly mispronounced. One peculiar aspect in conversation lessons is that learners are required to repeat the utterances of both interlocutors. In this sense, the activity resembles a typical ALM repetition drill, where language presentation is followed by drilling. Overall, the conversation lessons primarily focus on training learners' pronunciation skills and aspects of their understanding of pragmatics. However, the exercises seem to lack emphasis on developing learners' spoken fluency and overall communicative competence.

Chatbot. Outside of its units, Mondly provides additional conversation practice through its chatbot, illustrated in Figure 26 below. In general, chatbots can be divided into two types: rule-based or machine learning-based. Rule-based chatbots operate on the basis of a preprogrammed framework comprising specific questions and expected learner responses, whereas machine learning-based chatbots are able to recognise key words and patterns to comprehend the intention of the learner's input and consequently deliver a fitting response (Meshram et al. 2021: 940). While rule-based chatbots struggle to handle input that does not correspond to the programmed database, machine learning-based chatbots utilise techniques such as *machine learning* (ML), which refers to the ability to learn from and "generalize based on the examples it has seen before" (Meshram et al. 2021: 942), and *Natural Language Processing* (NLP), which "is a programming interface that [...] helps machine[s] to understand, interpret and manipulate human natural language" (Meshram et al. 2021: 942), to respond more flexibly to learners' input.



Figure 26 Mondly's chatbot

The analysis of Mondly's chatbot suggests that it blends elements of both rule-based and machine learning-based approaches. Although the provided conversations seem to follow a structured framework with predefined questions and suggested responses, learners can deviate to some extent from the proposed answers in certain instances, while Mondly is still able to maintain a meaningful conversation, as is exemplified in Figure 26 above. This indicates that Mondly's chatbot is based on a preprogrammed tree structure but also operates on a key phrase database, a fact noted by Fryer et al. (2020: 12). According to Fryer

(2020: 12), Mondly has been focusing on improving the machine learning capabilities of its chatbot in recent years to offer learners a more sophisticated and independent conversational experience. As a result of the chatbot's design and mechanics, conversation practice does not allow learners to be completely free in their responses. Nevertheless, some freedom is permitted, which is likely beneficial for the development of learners' spoken fluency and overall communicative competence. In terms of topics, Mondly's chatbot covers eight different conversation contexts, including greetings, eating out at restaurants, staying at hotels, purchasing tickets, small talk, going shopping, going on dates, and taking taxis. Each chatbot interaction involves topic-specific language. If learners cannot participate in a spoken conversation, Mondly offers the possibility to respond in written form. It is worth noting that the chatbot does not actively detect or mark learners' mistakes in terms of pronunciation or grammar during conversations.

Mondly AR. In addition to their standalone language learning app, Mondly also provides *augmented reality* (AR) and *virtual reality* (VR) apps. Mondly AR operates based on the concept of augmented reality, which "is a system that enhances the real world by superimposing computer-generated information on top of it" (Furht 2008a: 35). The purpose of AR is to provide learners with additional synthetic information, such as visualisations and audio, to assist them while they perform real world tasks. In Mondly AR, learners have the option to choose between lessons and conversation practice. After selecting their desired activity, the app prompts them to scan their surroundings using their mobile phone's camera. Mondly then generates a CGI (computer generated imagery) teacher and objects to enhance the learning experience.

AR lessons, which are illustrated in Figure 27 below, cover 10 different topics (introduction, space, instruments, animals, fruit, vegetables I & II, food, cutlery, and drinks). During these lessons, learners are exposed to CGI objects, such as different animals. Mondly AR utilises these objects to introduce and practice new lexical items. For instance, learners have the ability to approach these generated objects closely or walk around the animations to get a 360-degree view, in order to discover and learn new words related to the object. They are then prompted to repeat the presented word in the TL while they record themselves, allowing Mondly to check their pronunciation. Additionally, learners are also offered the opportunity to practice the newly learned words in sentences.

During AR conversations, illustrated in Figure 27 below, the app creates a simulated real life scenario, such as being at a restaurant, allowing learners to engage in role-plays with a computer generated character. Throughout the activity, learners are provided with a set of pre-formulated responses for each of their turns in the conversation. As observed in the chatbot conversations, learners are allowed to formulate their own, individual responses. However, Mondly's ability to process these freely produced utterances is, as of now, rather limited, which leads the app to potentially repeat its question or statement again. As AR conversations mimic real life situations, the main goal throughout the activity is completion of the task at hand (e.g. placing an order at a restaurant successfully). At the time of writing, Mondly AR provides three different scenarios for AR conversations including introducing oneself, ordering at a restaurant, and checking into a hotel. While Mondly AR stands out from other competitors by virtue of its novelty, the full potential of the innovative features has not been entirely realised yet. For younger, less advanced learners, Mondly AR can be an exciting opportunity, thanks to its lively animations and supportive conversation scaffolding. However, more advanced learners may find the limited freedom in terms of learner responses during the AR conversations somewhat monotonous and tiresome. Although Mondly AR may potentially improve learners' vocabulary knowledge and speaking skills, the interactions with animations cannot replace real life interactions where language skills can be practiced in a more authentic way.

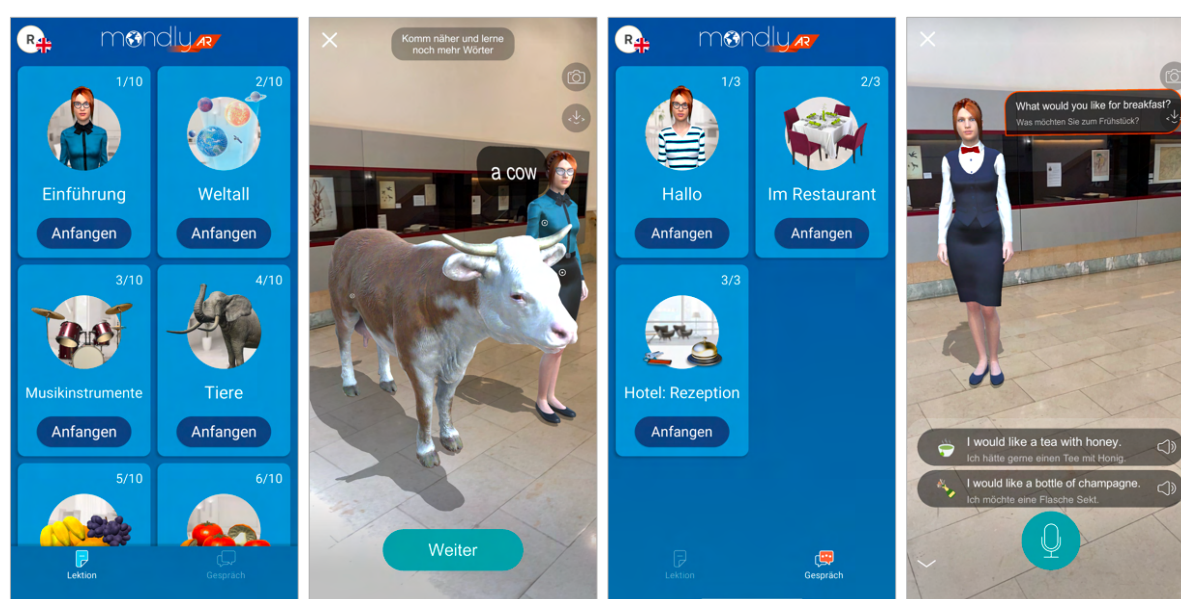


Figure 27 Lessons and conversation practice in Mondly AR

Mondly VR. Mondly VR utilises virtual reality technology, which "provides almost real and/or believable experiences in a synthetic or virtual way" (Furht 2008b: 968), in order to offer learners complete virtual immersion in a situated language learning environment. Unlike Mondly AR, which enhances learners' real world surroundings, Mondly VR creates a fully immersive virtual language learning experience where everyday situations, such as going grocery shopping, are simulated. To use Mondly VR, learners require a VR headset (e.g. *Oculus*). While the app is not available free of charge, the one-time-purchase fee is relatively affordable. Mondly VR offers the same two types of activities as Mondly AR: there are vocabulary lessons and conversation activities. In addition, Mondly VR also features a multiplayer mode, enabling learners to join a virtual room with up to eleven participants for one-on-one practice or collaborative learning (Mondly 2019). While the one-on-one conversations allow for more freedom in language production, the app also provides ready-made utterances as scaffolding to maintain the flow of the conversations. The multiplayer mode is currently available as beta version, i.e. it is in the second stage of development, undergoing testing by users. Mondly VR presents an innovative and promising approach to language learning, leveraging technology to enhance the learning experience. Particularly the multiplayer mode enables learners from all around the world to come together and collaborate in their language learning process, exchange ideas, and develop their communicative competence.

(D5) Exposure to the Target Language. Like Duolingo, Mondly adopts a bilingual approach to language instruction, which involves the frequent use of learners' L1 throughout the app. While all language learning activities in Mondly incorporate learners' first language, the extent to which it is used varies among different exercise types. Translation activities, in particular, feature learners' L1 most prominently. Conversation lessons, the chatbot, and simulated conversations in Mondly AR and Mondly VR, however, also strongly rely on the learners' L1. Task instructions are exclusively provided in the learners' first language, and all other areas outside of the language learning activities, such as the learner statistics, rankings, and the settings menu are dominated by the L1. As a result of the reliance on learners' first language, the app fails to maximise the amount of target language learners are exposed to, resulting in a score of 2 points on dimension (D5).

Evaluation results. Results of the evaluation of Mondly are shown in Table 30 below. It displays the app's individual scores awarded for each dimension, the app's tallied scores for each domain as well as the app's overall score.

Table 30 Evaluation results of Mondly

Duolingo Evaluation							
A: Educational Value		B: Design		C: Engagement		D: Language Learning	
A1 Thinking Skills	3	B1 Ability to Save Progress	5	C1 Learner Control	5	D1 Input	4
A2 21st Century Skills	5	B2 Integration	3	C2 Interactivity	5	D2 Output	3
A3 Value of Errors	3	B3 Screen Design	5	C3 Pace	5	D3 Level of Material	3
A4 Feedback to Teachers	N/A	B4 Ease of Use	5	C4 Personalisation	2	D4 Skills	3
A5 Cooperative Learning	4	B5 Navigation	5	C5 Interest	4	D5 Exposure to Target Language	2
A6 Individual Differences	3	B6 Information Presentation	5	C6 Aesthetics	5		
		B7 Media Integration	5	C7 Gamification	5		
		B8 Cultural Sensitivity	5				
Total score per domain	18/30		38/40		31/35		15/25
Total score	102 pts. / 130 pts.						

As Table 30 exhibits, Mondly receives 18 points for domain A Educational Value, 38 points for domain B Design, 31 points for domain C Engagement, and 15 points for domain D Language Learning, resulting in an overall score of 104 out of 130 points, which equals a percentage grade of about 78%. While Mondly shows strengths in terms of its design and overall capacity to engage learners, e.g. through its emphasis on learner control, its educational value and capacity to facilitate language learning show room for improvement and perform on average or below average on certain dimensions.

6.2.5. Conclusion

In conclusion, Mondly shares several similarities with Duolingo but also incorporates unique features that distinguish it from other competitors. While Mondly claims to help users "[s]peak a new language fluently fast" (Mondly 2023b), it is unlikely that the standard app alone can fulfill this ambitious promise. The language learning activities in Mondly primarily focus on building vocabulary, expanding grammar knowledge, and training receptive skills. Although the app offers valuable and relevant language input, most of the learning activities do not provide sufficient opportunities for learners to produce output beyond the sentence level. Mondly VR, however, stands out as it offers promising opportunities for learners to engage in authentic peer conversations. Relying solely on Mondly's standard language learning app, learners will struggle to develop real communicative competence. While

aspects of speaking skills, particularly pronunciation accuracy, are addressed throughout Mondly's language course, writing skills are neglected except for a focus on lexical item accuracy (i.e. spelling). Although Mondly VR introduces innovative features that enhance the language learning experience, and Mondly itself offers a solid language learning app with a gamified approach and an engaging user interface that may appeal to secondary level learners, using the app alone will only benefit specific aspects of language proficiency and may not effectively foster the development of true communicative competence.

7. Discussion of evaluation results

In this concluding chapter, insights gained from the evaluation of Duolingo and Mondly will be used to discuss possible strategies for integrating language learning apps into the EFL classroom.

7.1. Language learning apps in the EFL classroom

The evaluations of Duolingo and Mondly have yielded interesting insights into the design and mechanics of language learning apps. These findings can be used to draw conclusions about how an implementation of such apps might benefit secondary level learners in an EFL classroom setting. Since Duolingo and Mondly do not represent the full variety of language learning apps available on the market, these suggestions cannot be generalised for all language learning apps.

As discussed in sections 6.1.4 and 6.2.4., Duolingo and Mondly primarily train learners' vocabulary and grammar knowledge. Although both apps claim to provide holistic language learning courses, benefiting learners language proficiency overall, a thorough analysis shows that the proper training of receptive, but especially productive language skills is widely neglected throughout both apps. As both apps rely on outdated EFL teaching approaches, the Grammar Translation Method and the Audiolingual Method, learners will primarily engage in decontextualised grammar-translation exercises and audiolingual drills. Language learning apps like Duolingo and Mondly are therefore useful for developing decontextualised linguistic knowledge or training pronunciation accuracy. While these type of exercises can be valued for what they are, primarily being vessels for building vocabulary and grammar knowledge and training pronunciation and spelling accuracy, learners are likely to face challenges in applying the skills acquired through these exercises to real-life communication situations beyond the confines of the classroom (Richards & Rodgers 2014: 65) or even for authentic classroom interaction in the L2. Consequently, if we aim at written and spoken communicative competence as the overall objective of language learning in the EFL classroom, language learning apps, despite being developed in increasingly innovative ways, can only be considered as complementary learning tools to traditional EFL instruction.

Such complementary use can be found in terms of vocabulary and grammar training. The apps' capacities of building vocabulary and training grammar will likely benefit lower secondary level learners, especially beginners who learn English for the first time. This

assumption is underpinned by several efficacy studies that report a positive impact of language learning app implementation on learners' vocabulary acquisition and retrieval or grammar knowledge (e.g. Ajisoko 2020, Castañeda & Cho 2016, Dizon 2016, Li & Hegelheimer 2013, Wu 2014). By allocating a predefined amount of time in one of the weekly English lessons to vocabulary practice using a language learning app, teachers can rather effortlessly implement a novice tool in their EFL classroom. Alternatively, students could be asked to use any idle time during lessons to engage with the implemented language learning app and revise old or learn new vocabulary.

Teachers can potentially harness another advantage of language learning apps, which is their capability to stimulate and facilitate self-directed learning outside the language classroom. With their engaging gamified approaches to language learning, apps such as Duolingo or Mondly have the potential to motivate learners to continue their language learning journey beyond the traditional boundaries of the EFL classroom. By introducing language learning apps to their students, teachers can extend their educational influence to the students' homes, where the app serves as an extracurricular learning tool, acting as an agent and substitute for the teacher. Throughout my teaching experience over the past five years, I have frequently encountered requests from students and parents seeking recommendations for supplementary learning resources beyond traditional materials such as books, movies, podcasts, or videos in English. In response to these inquiries, I often suggested language learning apps to students, and the feedback I received from them was overwhelmingly positive. Many students expressed their enjoyment and satisfaction in using these apps as additional tools for language learning. While I cannot provide empirical evidence to support this claim, it is widely acknowledged among language educators that any additional exposure to the target language or opportunities for language learning can have a beneficial impact on learners' overall language proficiency.

Lastly, the innovative features offered by language learning apps, such as virtual reality (VR), augmented reality (AR), chatbots, and AI speech recognition, as discussed in sections 6.1.4 and 6.2.4, hold significant value for learning outside the EFL classroom and also introduce a new level of innovation within the classroom. Reflecting on Eli Broad's quote mentioned at the outset of this thesis, as language educators, it is imperative for us to embrace the changing times, stay abreast of technological advancements, and understand the digital lifestyle of our students, which includes ubiquitous mobile phone usage. By doing so, we can

establish a meaningful connection with our learners and provide language instruction that aligns with the demands of the 21st century and its technological progress.

8. References

- Abaunza, Geovanny; Martinez-Abad, Fernando; Conde-Rodriguez, Maria J. 2019. "Web applications in the EFL class in contexts rural school Colombian". In Conde Gonzales, Miguel A.; Rodriguez Sedano, Francisco J.; Fernandez Lllamas, Camino; Garcia-Penalvo, Francisco J. (eds.). *Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM 2019)*. New York, NY: Association for Computing Machinery, 613-618.
- Ajisoko, Pangkuh. 2020. "The use of Duolingo apps to improve English vocabulary learning". *International Journal of Emerging Technologies in Learning* 15(7), 149-155.
- Alfuhaid, Sultan R. 2021. "The utilisation of Duolingo to enhance the speaking proficiency of EFL secondary school students in Saudi Arabia". *English Language Teaching* 14(11), 9-15.
- Allen, Deborah; Tanner, Kimberly. 2006. "Rubrics: tools for making learning goals and evaluation criteria explicit for both teachers and learners". *CBE – Life Sciences Education* 5(3), 197-203.
- Armas Pesantez, Paul R.; Reinoso Espinosa, Ana G.; Macías Silva, Evelyn C.; Aquino Rojas, Marco A. 2019. "Incidencia de Duolingo en el desarrollo de las habilidades comunicacionales verbales del idioma inglés a nivel de educación superior [Duolingo's impact on the development of verbal communication skills in the English language at the higher education level]". *European Scientific Journal* 15(16), 29-44.
- Arnett, Katy. 2013. "Using the first language and code switching in second language classrooms". In Chapelle, Carol A. (ed.). *The encyclopedia of applied linguistics*. Hoboken, NJ: Wiley, 1-4. <https://doi-org.uaccess.univie.ac.at/10.1002/9781405198431.wbeal1240> (24 March 2023).
- Baran, Evrim. 2014. "A review of research on mobile learning in teacher education". *Educational Technology & Society* 17(4), 17-32.
- Bicknell, Clinton; Brust, Claire. 2020. "Learning how to help you learn: introducing Birdbrain!". *Duolingo Blog*. <https://blog.duolingo.com/learning-how-to-help-you-learn-introducing-birdbrain/> (7 June 2023).
- Blanco, Cindy. 2020. "Goldilocks and the CEFR levels: which proficiency level is just right". *Duolingo Blog*. <https://blog.duolingo.com/goldilocks-and-the-cefr-levels-which-proficiency-level-is-just-right/> (14 June 2023).
- Blanco, Cindy. 2022. "2022 Duolingo language report". *Duolingo Blog*. <https://blog.duolingo.com/2022-duolingo-language-report/> (17 April 2023).
- Bloom, Benjamin S. (ed.). 1956. *Taxonomy of educational objectives. The classification of educational goals. Handbook I: cognitive domain*. London: Longmans, Green & Co.

- Bowles, Melissa A. 2004. "L2 glossing: to CALL or not to CALL". *Hispania* 87(3), 541-552.
- Bryan-Gooden, Jahque; Hester, Megan; Peoples, Leah Q. 2019. *Culturally responsive curriculum scorecard*. New York, NY: NYU Metropolitan Center for Research on Equity and the Transformation of Schools. <https://steinhardt.nyu.edu/sites/default/files/2020-12/CRE%20Scorecard%20Revised%20Aug%202020.pdf> (7 February 2023).
- Buckler, Taylor; Peterson, Moya. 2012. "Is there an app for that? Developing an evaluation rubric for apps for use with adults with special needs". *The Journal of BSN Honors Research* 5(1), 19-32.
- Burston, Jack. 2015. "Twenty years of MALL project implementation: a meta-analysis of learning outcomes". *ReCALL* 27(1), 4-12.
- Bustillo, Judith; Rivera, Claribel; Guzmán, Juan G.; Ramos Acosta, Lizeth. 2017. "Benefits of using a mobile application in learning a foreign language". *Sistemas & Telemática* 15(40), 55-68.
- Cardwell, Ramsey L. 2021. "Why learn a made-up language?". *Duolingo blog*. <https://blog.duolingo.com/why-learn-a-made-up-language/> (20 April 2023).
- Castañeda, Daniel A.; Cho, Moon-Heum. 2016. "Use of a game-like application on a mobile device to improve accuracy in conjugating Spanish verbs". *Computer Assisted Language Learning* 29(7), 1195-1204.
- Center for Collaborative Education. 2017. *Fairness, bias, and cultural-responsiveness checklist*. Boston, MA: Center for Collaborative Education. <https://studentsatthecenterhub.org/wp-content/uploads/Fairness-Bias-and-Cultural-Responsiveness-Checklist.pdf> (7 February 2023).
- Chen, Xiaojun. 2016. "Evaluating language-learning mobile apps for second-language learner". *Journal of Educational Technology Development and Exchange* 9(2), 39-51.
- Cherner, Todd; Dix, Judy; Lee, Corey. 2014. "Cleaning up that mess: a framework for classifying educational apps". *Contemporary Technology Issues in Technology and Teacher Education* 14(2), 158-193.
- Chinnery, George M. 2006. "Emerging technologies: going to the MALL: mobile assisted language learning". *Language Learning and Technology* 10(1), 9-16.
- Chiu, Emily. 2021. "LGBTQIA+ representation in Duolingo stories and characters". *Duolingo blog*. <https://blog.duolingo.com/lgbtq-representation-in-duolingo-stories-and-characters/> (7 June 2023).
- Conway, Christopher M. 2020. "How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning". *Neuroscience and Biobehavioral Reviews* 112, 279-299.

- Corbalan, Gemma; Kester, Liesbeth; Merriënboer, Jeroen J. G. van. 2009. "Dynamic task selection: effects of feedback and learner control on efficiency and motivation". *Learning and Instruction* 19(6), 455-465.
- Cornillie, Frederik; Thorne, Steven L.; Desmet, Piet. 2012. "ReCALL special issue: digital games for language learning: challenges and opportunities". *ReCALL* 24(3), 243-256.
- Cowan, Pamela; Butler, Ryan. 2013. "Using activity theory to problematize the role of the teacher during mobile learning". *SAGE open* 3(4), 1-13.
- Curry, David. 2023. "Duolingo revenue and usage statistics (2023)". *Business of Apps*. <https://www.businessofapps.com/data/duolingo-statistics/> (20 April 2023).
- Deci, Edward L.; Vallerand, Robert J.; Pelletier, Luc G.; Ryan, Richard M. 1991. "Motivation and education: the self-determination perspective". *Educational Psychologist* 26(3-4), 325-346.
- Deterding, Sebastian; Dixon, Dan; Khaled, Rilla; Nacke, Lennart E. 2011. "From game design elements to gamefulness: defining 'gamification'". In Lugmayr, Artur; Franssila, Heljä; Safran, Christian; Hammouda, Imed (eds.). *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments (MindTrek '11)*. New York, NY: Association for Computing Machinery, 9-15.
- Diaz-Vera, Javier E. 2012. "Great expectations: formalizing and transforming mobile-assisted language learning". In Diaz-Vera, Javier E. (ed.). *Left to my own devices: learner autonomy and mobile-assisted language learning*. Bingley: Emerald, xi-xix.
- Dizon, Gilbert. 2016. "Quizlet in the EFL classroom: enhancing academic vocabulary acquisition of Japanese university students". *Teaching English with Technology* 16(2), 40-56.
- Duolingo. 2023a. *Duolingo for schools*. <https://schools.duolingo.com/> (31 May 2023).
- Duolingo. 2023b. "Color". *Duolingo Brand*. <https://design.duolingo.com/identity/color#core-brand-colors> (5 June 2023).
- Duolingo. 2023c. "Shape language". *Duolingo Brand*. <https://design.duolingo.com/illustration/shape-language#construction> (5 June 2023).
- Duolingo. 2023d. "Typography". *Duolingo Brand*. <https://design.duolingo.com/identity/typography#feather-bold> (5 June 2023).
- Duolingo. 2023e. "Can you become fluent with Duolingo?". *Duolingo Help Center*. <https://support.duolingo.com/hc/en-us/articles/360056797071-Can-you-become-fluent-with-Duolingo-> (6 June 2023).
- El-Hussein, Mohamed O.; Cronje, Johannes C. 2010. "Defining mobile learning in the higher education landscape". *Educational Technology & Society* 13(3), 12-21.

- Fox, Richard. 2001. "Constructivism examined". *Oxford Review of Education* 21(1), 23-35.
- Freeman, Cassie; Kittredge, Audrey; Wilson, Hope; Pajak, Bozena. 2023. "The Duolingo method for app-based teaching and learning". *Duolingo Research Report*, 1-19.
<https://duolingo-papers.s3.amazonaws.com/reports/duolingo-method-whitepaper.pdf>
 (24 May 2023).
- Fryer, Luke K.; Coniam, David; Carpenter, Rollo; Lăpușneanu, Diana. 2020. "Bots for language learning now: current and future directions". *Language Learning and Technology* 24(2), 8-22.
- Furht, Borko. 2008a. "Augmented reality". In Furht, Borko (ed.). *Encyclopedia of Multimedia*. (2nd edition). Boston, MA: Springer, 35-36.
- Furht, Borko. 2008b. "Virtual reality". In Furht, Borko (ed.). *Encyclopedia of Multimedia*. (2nd edition). Boston, MA: Springer, 968.
- Galitz, Wilbert O. 2007. *The essential guide to user interface design: an introduction to GUI design principles and techniques*. Indianapolis, IN: Wiley.
- García Mayo, Maria del Pilar. 2021. "Introduction". In Pilar García Mayo, Maria (ed.). *Working collaboratively in second/foreign language learning*. Berlin: De Gruyter Mouton, 1-12.
- Gay, Geneva. 2002. "Preparing for culturally responsive teaching". *Journal of Teacher Education* 53(2), 106-116.
- Glaserfeld, Ernst von. 1995. *Radical constructivism: a way of knowing and learning*. London: Routledge Falmer.
- Gogolin, Ingrid. 2008. *Der monolinguale Habitus der multilingualen Schule [The monolingual habitus of the multilingual school]*. Münster: Waxmann.
- Govender, Terence; Arnedo-Moreno, Joan. 2020. "A survey on gamification elements in mobile language-learning applications". In Garcia-Penalvo, Francisco J.; Garcia-Holgado, Alicia (eds.). *Proceedings of the Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM 2020)*. New York, NY: Association for Computing Machinery, 669-676.
- Guaqueta, Cesar A.; Castro-Garces, Angela Y. 2018. "The use of language learning apps as a didactic tool for EFL vocabulary building". *English Language Teaching* 11(2), 61-71.
- Hall, Graham; Cook, Guy. 2012. "Own-language use in language teaching and learning". *Language Teaching* 45(3), 271-308.
- Hattie, John; Clarke, Shirley. 2019. *Visible learning: feedback*. London: Routledge.

- Hattie, John; Timperley, Helen. 2007. "The power of feedback". *Review of Educational Research* 77(1), 81-112.
- Hirvela, Alan. 2013. "Teaching integrated skills". In Chappelle, Carol A. (ed.). *The encyclopedia of applied linguistics*. Hoboken, NJ: Wiley, 1-4. <https://doi-org.uaccess.univie.ac.at/10.1002/9781405198431.wbeal1161> (14 March 2023).
- Hu, Jane C. 2021. "Why are Duolingo's sentences so weird?". *Slate*. <https://slate.com/technology/2021/11/duolingo-weird-sentences-linguistics.html> (8 May 2023).
- Hutz, Matthias. 2021. "Focus on form – the lexico-grammar approach". In Surkamp, Carola; Viebrock, Britta (eds.). *Teaching English as a foreign language: an introduction*. Stuttgart: J.B. Metzler, 133-158.
- James, Kelsey K.; Mayer, Richard E. 2019. "Learning a second language by playing a game". *Applied Cognitive Psychology* 33(4), 669-674.
- Jiang, Xiangyin; Rollinson, Joseph; Plonsky, Luke; Gustafson, Erin; Pajak, Bozena. 2021. "Evaluating the reading and listening outcomes of beginning-level Duolingo courses". *Foreign Language Annals* 54(4), 974-1002.
- Jensen, Casper J.; Cadierno, Teresa. 2022. "Differences in mobile-assisted acquisition of receptive and productive vocabulary knowledge: a case study using Mondly". *The Language Learning Journal*, 1-16. <https://www.tandfonline.com/doi/full/10.1080/09571736.2022.2108123> (25 June 2023).
- Jonsson, Anders; Svingby, Gunilla. 2007. "The use of scoring rubrics: reliability, validity and educational consequences". *Educational Research Review* 2(2), 130-144.
- Kang, Sean H. 2016. "Spaced repetition promotes efficient and effective learning: policy implications for instruction". *Policy Insights from the Behavioral and Brain Sciences* 3(1), 12-19.
- Karim, Michael N.; Behrend, Tara S. 2013. "Reexamining the nature of learner control: dimensionality and effects on learning and training reactions". *Journal of Business and Psychology* 29(1), 87-99.
- Katz, Idit; Assor, Avi. 2007. "When choice motivates and when it does not". *Educational Psychology Review* 19(4), 429-442.
- Kaufman, Dorit. 2004. "Constructivist issues in language learning and teaching". *Annual Review of Applied Linguistics* 24, 303-319.
- Kim, Heyoung; Kwon, Yeonhee. 2012. "Exploring smartphone applications for effective mobile-assisted language learning". *Multimedia-assisted language learning* 15(1), 31-57.

- Kolb, Annika; Raith, Thomas. 2021. "Principles and methods – focus on learners, content and task". In Surkamp, Carola; Viebrock, Britta (eds.). *Teaching English as a foreign language: an introduction*. Stuttgart: J.B. Metzler, 195-209.
- Krapp, Andreas. 1999. "Interest, motivation, and learning: an educational, psychological perspective". *European Journal of Psychology of Education* 14(1), 23-40.
- Krashen, Stephen D. 1982. *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon Press. https://www.sdkrashen.com/content/books/principles_and_practice.pdf (24 March 2023).
- Krashen, Stephen D. 1985. *The input hypothesis: issues and implication*. London: Longman.
- Krashen, Stephen D. 2004. "Applying the comprehension hypothesis: some suggestions". Paper presented at the 13th *International Symposium and Book Fair on Language Teaching (English Teachers Association of the Republic of China)*, Taipei, Taiwan, 13 Nov. 2004. http://www.sdkrashen.com/content/articles/eta_paper.pdf (24 March 2023).
- Krashen, Stephen. 2014. "Does Duolingo 'trump' university-level language learning?". *The International Journal of Foreign Language Teaching* 9(1), 13-15. <https://sdkrashen.com/content/articles/krashen-does-duolingo-trump.pdf> (24 April 2023).
- Kukulska-Hulme, Agnes; Shield, Lesley. 2008. "An overview of mobile assisted language learning: from content delivery to supported collaboration and interaction". *ReCALL* 20(3), 271-289.
- Kukulska-Hulme, Agnes. 2013. "Mobile-assisted language learning". In Chapelle, Carol A. (ed.). *The encyclopedia of applied linguistics*. Hoboken, NJ: Wiley, 1-8. <https://doi-org.uaccess.univie.ac.at/10.1002/9781405198431.wbeal0768> (20 June 2023).
- Kukulska-Hulme, Agnes. 2020. "Mobile-assisted language learning". In Chapelle, Carol A. (ed.). *The encyclopedia of applied linguistics*. Hoboken, NJ: Wiley, 1-9. <https://doi-org.uaccess.univie.ac.at/10.1002/9781405198431.wbeal0768.pub2> (20 June 2023).
- Kukulska-Hulme, Agnes. 2009. "Will mobile learning change language learning?". *ReCALL* 21(2), 157-165.
- Lăpușneanu, Diana. 2021. "The learning method behind Mondly". <https://www.mondly.com/blog/learning-method-mondly/> (1 July 2023).
- Larsen-Freeman, Diane. 2000. *Techniques and principles in language teaching*. (2nd edition). Oxford: Oxford University Press.
- Lee, Cheng-Yuan; Cherner, Todd S. 2015. "A comprehensive evaluation rubric for assessing instructional apps". *Journal of Information Technology Education: Research* 14, 21-53.

- Leow, James; Bialostozky, Jacqueline; Blanco, Cindy. 2021. "How silly sentences can help you learn". *Duolingo Blog*. <https://blog.duolingo.com/how-silly-sentences-can-help-you-learn/> (14 June 2023).
- Li, Zhi; Hegelheimer, Volker. 2013. "Mobile-assisted grammar exercises: effects of self-editing in L2 writing". *Language Learning and Technology* 17(3), 135-156.
- Lidwell, William; Holden, Kritina; Butler, Jill. 2010. *Universal principles of design*. (2nd edition). Beverly, MA: Rockport.
- Loewen, Shawn; Crowther, Dustin; Isbell, Daniel R.; Kim, Kathy M.; Maloney, Jeffrey; Miller, Zachary F.; Rawal, Hima. 2019. "Mobile-assisted language learning: a duolingo case study". *ReCALL* 31(3), 293-311.
- Lucariello, Joan M.; Nastasi, Bonnie K.; Anderman, Eric M.; Dwyer, Carol; Ormiston, Heather; Skiba, Russell. 2016. "Science supports education: the behavioral research base for psychology's top 20 principles for enhancing teaching and learning". *Mind, Brain, and Education* 10(1), 55-67.
- Mayer, Richard E. 2021a. "Cognitive theory of multimedia learning". In Mayer, Richard E.; Logan, Fiorella (eds.). *The Cambridge handbook of multimedia learning*. Cambridge: Cambridge UP, 57-72.
- Mayer, Richard E. 2021b. *Multimedia learning*. (3rd edition). Cambridge: Cambridge UP.
- McQuiggan, Scott; Kosturko, Lucy; McQuiggan, Jamie; Sabourin, Jennifer. *Mobile learning: a handbook for developers, educators, and learners*. Hoboken, NJ: Wiley.
- Mertler, Craig A. 2000. "Designing scoring rubrics for your classroom". *Practical Assessment, Research, and Evaluation* 7(25), 1-8.
- Meshram, Siddhant; Naik, Namit; VR, Megha; More, Tanmay; Kharche, Shubhangi. 2021. "Conversational AI: chatbots". In Institute of Electrical and Electronics Engineers (ed.). *Proceedings of the 2021 International Conference on Intelligent Technologies (CONIT)*. Piscataway, NJ: IEEE, 940-945.
- Mihaylova, Mariela; Gorin, Simon; Reber, Thomas P.; Rothen, Nicolas. 2022. "A meta analysis on mobile-assisted language learning applications: benefits and risks". *Psychologica Belgica* 62(1), 252-271.
- Moline, Emily; Blanco, Cindy. 2020. "Covering all the bases: Duolingo's approach to speaking skills". *Duolingo Blog*. <https://blog.duolingo.com/covering-all-the-bases-duolingos-approach-to-speaking-skills/> (12 May 2023).
- Mondly. 2019. "Learn languages in virtual reality with Mondly VR multiplayer". <https://www.mondly.com/blog/take-a-glimpse-into-the-future-with-mondly-vr-multiplayer/> (5 July 2023).

- Mondly. 2021a. "Mondly celebrates its 7th anniversary with 75 million learners worldwide". <https://www.mondly.com/blog/mondly-timeline/> (25 June 2023).
- Mondly. 2021b. "Mondly wins Chairman's Award at 2021 WITSA Global ICT Excellence Awards". <https://www.mondly.com/blog/mondly-chairmans-award-witsa/> (26 June 2023).
- Mondly. 2023a. "Frequently asked questions". <https://www.mondly.com/support/appstore> (25 June 2023).
- Mondly. 2023b. "Speak a new language fluently fast". <https://www.mondly.com/app> (8 July 2023).
- Moreno, Roxana. 2010. *Educational psychology*. Hoboken, NJ: John Wiley & Sons.
- Nation, Ian S. 2022. *Learning vocabulary in another language*. (3rd edition). Cambridge: Cambridge University Press.
- Nicholls, John G. 1984. "Achievement motivation: conceptions of ability, subjective experience, task choice, and performance". *Psychological Review* 91(3), 328-346.
- Ok, Min W.; Kim, Min K.; Kang, Eun Y.; Bryant, Brian R. 2016. "How to find good apps: an evaluation rubric for instructional apps for teaching students with learning disabilities". *Intervention in School and Clinic* 51(4), 244-252.
- Olsen, Dwayne G. 1999. "Constructivist principles of learning and teaching methods". *Education* 120(2), 347-355.
- Oxford, Rebecca L. 1997. "Cooperative learning, collaborative learning, and interaction: three communicative strands in the language classroom". *The Modern Language Journal* 81(4), 443-456.
- Oxford, Rebecca L. 2001. "Integrated skills in the ESL/EFL classroom". *The Journal of TESOL France* 8, 5-12.
- Palalas, Agnieszka. 2011. "Mobile-assisted language learning: designing for your students". In Thouësny, Sylvie; Bradley, Linda (eds.). *Second language teaching and learning with technology: views of emergent researchers*. Dublin: Research-publishing.net, 71-94.
- Papadakis, Stamatios; Kalogiannakis, Michail; Zaranis, Nicholas. 2017. "Designing and creating an educational app rubric for preschool teachers". *Education and Information Technologies* 22(6), 3147-3165.
- Partnership for 21st Century Learning. 2019a. *Framework for 21st century learning*. Columbus, OH: Battelle for Kids. https://static.battelleforkids.org/documents/p21/P21_Framework_Brief.pdf (7 February 2023).

- Partnership for 21st Century Learning. 2019b. *Framework for 21st century learning definitions*. Columbus, OH: Battelle for Kids. https://static.battelleforkids.org/documents/p21/P21_Framework_DefinitionsBfK.pdf (7 February 2023).
- Patall, Erika A.; Cooper, Harris; Civey Robinson, Jorgianne. 2008. "The effects of choice on intrinsic motivation and related outcomes: a meta-analysis of research findings". *Psychological Bulletin* 134(2), 270-300.
- Pearson. 2022. "Pearson acquires leading online language learning platform". <https://plc.pearson.com/en-GB/news/pearson-acquires-leading-online-language-learning-platform> (25 June 2023).
- Persson, Veronica; Nouri, Jalal. 2018. "A systematic review of second language learning with mobile technologies". *International Journal of Emerging Technologies in Learning* 13(2), 188-210.
- Phillips, Christine; Chaparro, Barbara S. 2009. "Visual appeal vs. usability: which one influences user perceptions of a website more?" <https://researchinuserexperience.wordpress.com/2009/10/15/visual-appeal-vs-usability-which-one-influences-user-perceptions-of-a-website-more/> (27 February 2023).
- Piaget, Jean. 1971. *Biology and knowledge: an essay on the relations between organic regulations and cognitive processes*. Chicago, IL: Chicago University Press.
- Porto, Melina. 2010. "Culturally responsive L2 education: an awareness-raising proposal". *ELT Journal* 64(1), 45-53.
- Rachels, Jason R.; Rockinson-Szapkiw, Amanda J. 2018. "The effects of a mobile gamification app on elementary students' Spanish achievement and self-efficacy". *Computer Assisted Language Learning* 31(1-2), 72-89.
- Reeves, Thomas C.; Harmon, Stephen W. 1994. "Systematic evaluation procedures for interactive multimedia for education and training". In Reisman, Sorel (ed.). *Multimedia Computing. Preparing for the 21st century*. Harrisburg, PA: Idea Group, 472-505.
- Reyna, Jorge. 2013. "The importance of visual design and aesthetics in e-learning". *Training and Development* 40(5), 28-31.
- Richards, Jack C.; Rodgers, Theodore S. 2014. *Approaches and methods in language teaching*. (3rd edition). Cambridge: Cambridge University Press.
- Rosell-Aguilar, Fernando. 2017. "State of the app: a taxonomy and framework for evaluating language learning mobile applications". *CALICO Journal* 34(2), 243-258.
- Ryan, Richard M.; Deci, Edward L. 2000. "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being". *American Psychologist* 55(1), 68-78.

- Sailer, Michael; Homner, Lisa. 2020. "The gamification of learning: a meta-analysis". *Educational Psychology Review* 32(1), 77-112.
- Salcedo Salcedo, Sandra P.; Fernández Morales, Flavio H.; Duarte, Julio E. 2018. " Mejora de la habilidad de escritura en inglés en niños con Síndrome de Down con el apoyo de nuevas tecnologías [Improving the English language writing skills of children with Down syndrome supported by new technologies]". *Revista Espacios* 39(10), 18-31.
- Savvani, Stamatia. 2019. "State-of-the-art Duolingo features and applications". In Auer, Michael E.; Tsiatsos, Thrasyvoulos (eds.). *The challenges of the digital transformation in education: proceedings of the 21st international conference on interactive collaborative learning (ICL2018)*. Vol. 2. Cham: Springer Nature Switzerland, 139-148.
- Schiefele, Ulrich. 1991. "Interest, learning, and motivation". *Educational Psychologist* 26(3-4), 299-323.
- Schnackenberg, Heidi L.; Sullivan, Howard J. 2000. "Learner control over full and lean computer-based instruction under differing ability levels". *Educational Technology Research and Development* 48(2), 19-35.
- Seaborn, Katie; Fels, Deborah I. 2015. "Gamification in theory and action: a survey". *International Journal of Human-Computer Studies* 74, 14-31.
- Settles, Burr; Meeder, Brendan. 2016. "A trainable spaced repetition model for language learning". In Erk, Katrin; Smith, Noah A. (eds.). *Proceedings of the 54th annual meeting of the Association for Computational Linguistics*. Vol. 1. Berlin: Association for Computational Linguistics, 1848-1858.
- Shing, Yee L.; Brod, Garvin. 2016. "Effects of prior knowledge on memory: implications for education". *Mind, Brain, and Education* 10(3), 153-161.
- Shneiderman, Ben; Plaisant, Catherine. 2005. *Designing the user interface: strategies for effective human-computer interaction*. (4th edition). Boston, MA: Pearson/Addison Wesley.
- Shortt, Mitchell; Shantanu, Tilak; Kuznetcova, Irina; Martens, Bethany; Akinkuolie, Babatunde. 2023. "Gamification in mobile-assisted language learning: a systematic review of Duolingo literature from public release of 2012 to early 2020". *Computer Assisted Language Learning* 36(3), 517-554.
- Simsek, Ali. 2012. "Learner control". In Seel, Norbert M. (ed.). *Encyclopedia of the sciences of learning*. New York, NY: Springer, 1748-1750.
- Sorgenfrei, Christian; Smolnik, Stefan. 2016. "The effectiveness of e-learning systems: a review of the empirical literature on learner control". *Decision Sciences Journal of Innovative Education* 14(2), 154-184.

- Statista. 2022a. *Smartphone-Besitz bei Kindern und Jugendlichen in Deutschland im Jahr 2021 nach Altersgruppe*. <https://de.statista.com/statistik/daten/studie/1106/umfrage/handybesitz-bei-jugendlichen-nach-altersgruppen/> (26 June 2023).
- Statista. 2022b. *Number of available apps in the Apple App Store from 1st quarter 2015 to 3rd quarter 2022*. <https://www.statista.com/statistics/779768/number-of-available-apps-in-the-apple-app-store-quarter/> (30 July 2023).
- Statista. 2023a. *Number of available applications in the Google Play Store from December 2009 to June 2023*. <https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/> (30 July 2023).
- Statista. 2023b. *Leading language learning apps worldwide in February 2023, by downloads*. <https://www.statista.com/statistics/1239522/top-language-learning-apps-downloads/> (26 June 2023).
- Still, Brian; Crane, Kate. 2017. *Fundamentals of user-centered design: a practical approach*. Boca Raton, FL: CRC Press.
- Sung, Yao-Ting; Chang, Kuo-En; Liu, Tzu-Chien. 2016. "The effects of integrating mobile devices with teaching and learning on students' learning performance: a meta-analysis and research synthesis". *Computers & Education* 94, 252-275.
- Swain, Merrill. 1985. "Communicative competence: some roles of comprehensible input and comprehensible output in its development". In Gass, Susan M.; Madden, Carolyn G. (eds.). *Input in second language acquisition*. Rowley, MA: Newbury House, 235-253.
- Swain, Merrill. 1995. "Three functions of output in second language learning". In Cook, Guy; Seidlhofer, Barbara (eds.). *Principle and practice in applied linguistics: studies in honour of H. G. Widdowson*. Oxford: Oxford University Press, 125-144.
- Swain, Merrill. 2005. "The output hypothesis: theory and research". In Hinkel, Eli (ed.). *Handbook of research in second language teaching and learning*. Mahwah, NJ: Erlbaum, 471-483.
- Sweeney, Paul; Moore, Caroline. 2012. "Mobile apps for learning vocabulary: categories, evaluation and design criteria for teachers and developers". *International Journal of Computer-Assisted Language Learning and Teaching* 2(4), 1-16.
- Sweller, John. 1988. "Cognitive load during problem solving: effects on learning". *Cognitive Science* 12(2), 257-285.
- Sweller, John; Merriënboer, Jeroen J. G. van; Paas, Fred G. W. C. 1998. "Cognitive architecture and instructional design". *Educational Psychology Review* 10(3), 251-296.
- Sweller, John; Merriënboer, Jeroen J. G. van; Paas, Fred G. W. C. 2019. "Cognitive architecture and instructional design: 20 years later". *Educational Psychology Review* 31(2), 261-292.

- Teske, Kaitlyn. 2017. "Duolingo". *CALICO Journal* 34(3), 393-401.
- Thompson, Gregory L.; Harrison, Katie. 2014. "Language use in the foreign language classroom". *Foreign Language Annals* 47(2), 321-337.
- Tommerdahl, Jodi M.; Dragonflame, Chrystal S.; Olsen, Amanda A. 2022. "A systematic review examining the efficacy of commercially available foreign language learning mobile apps". *Computer Assisted Language Learning* (ahead of print), 1-30.
<https://www.tandfonline.com/doi/full/10.1080/09588221.2022.2035401> (8 July 2023).
- Traxler, John. 2007. "Defining, discussing, and evaluating mobile learning: the moving finger writes and having writ...". *International Review of Research in Open and Distance Learning* 8(2), 1-13.
- Turnbull, Miles. 2001. "There is a role for the L1 in second and foreign language teaching, but...". *The Canadian Modern Language Review* 57(4), 531-540.
- Vandewaetere, Mieke; Clarebout, Geraldine. 2011. "Can instruction as such affect learning? The case of learner control". *Computers & Education* 57(4), 2322-2332.
- Vavrus, Michael. 2008. "Culturally responsive teaching". In Good, Thomas L. (ed.). *21st century education: a reference handbook*. Vol. 2. Thousand Oaks, CA: Sage.
- Vesselinov, Roumen; Grego, John. 2012. "Duolingo effectiveness study: final report".
http://static.duolingo.com/s3/DuolingoReport_Final.pdf (20 June 2023).
- Vygotsky, Lev S.; Cole, Michael; John-Steiner, Vera; Scribner, Sylvia; Souberman, Ellen (eds.). 1978. *Mind in society: the development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Walker, Harry. 2011. "Evaluating the effectiveness of apps for mobile devices". *Journal of Special Education Technology* 26(4), 59-63.
- Walkington, Candace; Bernacki, Matthew L. 2014. "Motivating students by 'personalizing' learning around individual interests: a consideration of theory, design, and implementation issues". *Motivational Interventions* 18, 139-176.
- Webb, Norman L. 1997. *Criteria for Alignment of expectations and assessments in Mathematics and science education*. Madison, WI: University of Wisconsin-Madison.
- Webb, Norman L. 2002. *An analysis of the alignment between Mathematics standards and assessments for three states*. Madison, WI: University of Wisconsin-Madison.
- Whitton, Nicola. 2012. "Game-based learning". In Seel, Norbert M. (ed.). *Encyclopedia of the sciences of learning*. New York, NY: Springer, 1337-1340.
- Widdowson, Henry G. 1978. *Teaching language as communication*. Oxford: Oxford University Press.

- Winters, Niall. 2006. "What is mobile learning?". In Sharples, Mike (ed.). *Big issues in mobile learning: report of a workshop by the Kaleidoscope Network of Excellence Mobile Learning Initiative*. Nottingham: University of Nottingham, 5-9.
- Wu, Qun. 2014. "Learning ESL vocabulary with smartphones". *Procedia – Social and Behavioral Sciences* 143, 302-307.
- Wylie, Louise. 2023. "Education app revenue and usage (2023)". *Business of Apps*. <https://www.businessofapps.com/data/education-app-market/> (8 July 2023).
- Zhang, Yan. 2010. "Cooperative language learning and foreign language learnign and teaching". *Journal of Language Teaching and Research* 1(1), 81-83.

9. Appendix

9.1. Abstract English

With the rapid advancements in mobile technology over the past few decades, mobile language learning apps have increasingly taken center stage in the field of language education in recent years. Made possible by a high and growing percentage of children and teenagers who possess mobile devices such as smartphones and tablets, these apps offer a new approach to language learning, free from temporal, spatial, and educational constraints. However, the abundance of language learning apps available for download from app stores raises an important question for language teachers and learners alike: which language learning apps are of high quality and beneficial for language learning? To address this question, this study pursued two main objectives. Firstly, a theory-based evaluation rubric for language learning apps was developed, which underwent a peer-review. This rubric comprises four domains and twenty-six dimensions, serving as a tool to assess language learning apps in terms of their educational value, design, engagement potential, and ability to facilitate language learning. Secondly, this Evaluation Rubric for Language Learning Apps for Secondary Level Learners was employed to evaluate and compare two popular language learning apps, *Duolingo* and *Mondly*. The systematic-qualitative-interpretive analysis and evaluation aimed to delve into the mechanics of these popular language learning apps, identify the underlying principles and methods of EFL learning and teaching, and explore how these apps might enhance learners' language learning experience. The analysis revealed that both language learning apps, despite promising to improve learners' overall language proficiency through fostering accuracy and fluency, primarily focus on training vocabulary and grammar knowledge. Essential language skills such as reading, listening, speaking, and writing are only partially addressed or overlooked entirely by both apps. While language learning apps offer many innovative features and benefits for language learners, these insights suggest that both apps can only be utilised as complementary resources to traditional language learning methods in the EFL classroom and beyond.

9.2. Abstract German

Beruhend auf den schnellen Fortschritten in der mobilen Technologie in den letzten Jahrzehnten haben mobile Sprachlern-Apps in den letzten Jahren zunehmend eine zentrale Rolle im Bereich der Sprachbildung eingenommen. Ermöglicht durch einen hohen und wachsenden Prozentsatz von Kindern und Jugendlichen, die mobile Geräte wie Smartphones und Tablets besitzen, bieten diese Apps einen neuen Ansatz für das Sprachenlernen, frei von zeitlichen, räumlichen und pädagogischen Einschränkungen. Die Fülle an Sprachlern-Apps, die zum Herunterladen in App Stores verfügbar sind, wirft jedoch eine wichtige Frage für Sprachlehrer:innen und Lernende gleichermaßen auf: Welche Sprachlern-Apps sind von hoher Qualität und förderlich für das Sprachenlernen? Um diese Frage zu beantworten, verfolgte diese Studie zwei Hauptziele. Erstens wurde eine theoriebasierte Bewertungsrubrik für Sprachlern-Apps entwickelt, die einer Peer-Review unterzogen wurde. Diese Rubrik umfasst vier Domänen und sechsundzwanzig Dimensionen und dient als Instrument zur Bewertung von Sprachlern-Apps hinsichtlich ihres pädagogischen Wertes, Designs, Potenzials Lernende zu motivieren und ihrer Fähigkeit, das Sprachenlernen zu fördern. Zweitens wurde diese Bewertungsrubrik für Sprachlern-Apps für Schülerinnen und Schüler der Sekundarstufe verwendet, um zwei beliebte Sprachlern-Apps, *Duolingo* und *Mondly*, zu bewerten und zu vergleichen. Die systematische-qualitative-interpretative Analyse und Bewertung zielte darauf ab, die Mechanismen dieser beliebten Sprachlern-Apps zu untersuchen, die den Apps zugrunde liegenden Prinzipien und Methoden des Fremdsprachenlernens und -unterrichts zu identifizieren, und zu erforschen, wie diese Apps das Sprachenlernen verbessern könnten. Die Analyse ergab, dass beide Sprachlern-Apps trotz der Aussicht, die Sprachkompetenz der Lernenden durch Förderung von *accuracy* und *fluency* zu verbessern, hauptsächlich auf das Training von Wortschatz und Grammatikkenntnissen abzielen. Wesentliche Sprachfertigkeiten wie Lesen, Hören, Sprechen und Schreiben werden von beiden Apps nur teilweise trainiert oder sogar vollständig vernachlässigt. Obwohl Sprachlern-Apps viele innovative Funktionen und Vorteile für Sprachlernende bieten, deuten diese Erkenntnisse darauf hin, dass beide Apps nur als ergänzende Ressourcen zu traditionellen Fremdsprachenlernmethoden im Englischunterricht und darüber hinaus genutzt werden können.

9.3. Evaluation Rubric for LLAs for Secondary Level Learners

Evaluation Rubric for Language Learning Apps for Secondary Level Learners

Domain (A) Educational Value (Lee & Cherner 2015, adapted)					
Dimension (A1) Thinking Skills: Which level of thinking skills (Webb's <i>Depth of Knowledge</i>) must learners apply to meet the app's learning objectives?					
5	4	3	2	1	N/A
Learners are required to apply extended thinking skills (DoK-4) to meet the app's learning objectives. key words: <i>create, synthesise, design, analyse</i>	Learners are required to apply strategic thinking skills (DoK-3) to meet the app's learning objectives. key words: <i>assess, critique, formulate, hypothesise</i>	Learners are required to apply existing skills and conceptual knowledge (DoK-2) to meet the app's learning objectives. key words: <i>classify, compare, predict summarise</i>	Learners are required to recall and reproduce information (DoK-1) to meet the app's learning objectives. key words: <i>recall, recognise, match, memorise</i>	Learners are not required to apply any level of thinking skills because the app is geared more towards entertainment than education.	Not Applicable
Dimension (A2) 21st Century Skills: Does the app require learners to engage 21st century skills, such as <i>critical thinking</i> and <i>problem solving</i> (e.g. reasoning effectively, solving problems innovatively), <i>communication</i> and <i>collaboration</i> (e.g. communicating clearly and effectively, collaborating with others), <i>creativity</i> and <i>innovation</i> (e.g. thinking creatively, working creatively with others), <i>initiative</i> and <i>self-direction</i> (e.g. managing goals and time, working independently), and <i>information and media literacy</i> (e.g. accessing and evaluating information, applying technology effectively)?					
5	4	3	2	1	N/A
The app frequently requires learners to engage 21st century skills.	The app sometimes requires learners to engage 21st century skills.	The app rarely requires learners to engage 21st century skills.	The app does not require learners to engage 21st century skills.	The app promotes rote memorisation skills.	Not Applicable
Dimension (A3) Value of Errors: If learners answer a question incorrectly, does the app use the incorrect answer to advance learning?					
5	4	3	2	1	N/A
The app offers learners an explanation for their incorrect answer, allows learners to review the material related to the question, and then gives them an opportunity to answer the question again.	The app offers learners either an explanation for their incorrect answer or directs them to the material related to the question before giving learners an opportunity to answer the question again	The app offers learners an explanation for why their answer was incorrect, but the app does not give them an opportunity to answer the question again.	The app offers learners an opportunity to answer the question again, but the app does not provide information to explain why their answer was incorrect the first time.	The app does not give learners an opportunity to answer the question again or provide information about why their answer was incorrect.	Not Applicable

Dimension (A4) Feedback to Teachers: How does the app allow teachers to monitor their students' academic growth (language proficiency level), achievements (e.g. weekly challenges), and progress (advance in the language course)?					
5	The app tracks learners' academic growth, achievements, and progress made while using the app. Teachers can access the reports either through email or by visiting the app's website.	4	The app tracks learners' achievements or progress made when using the app, but it does not track learners' growth. Teachers can access the reports using email or by visiting the app's website.	3	The app tracks learners' achievements or progress made when using the app, but it does not track learners' growth. Also, learner progress can only be seen in the app and not by using email or a website.
				2	The app only tracks learners' achievements or progress when they are using the app. Once the app is exited, progress is not saved, and teachers do not have any way to access it.
				1	The app does not track students' academic growth, achievements, or progress made.
					N/A
Dimension (A5) Cooperative Learning: Does the app allow learners to communicate (e.g. send/receive messages and post to message boards), share learning artifacts (e.g. images, recordings, and documents), and collaborate (e.g. create presentations and larger projects) with other users?					
5	The app features an online community where learners can communicate freely with other users, share learning artifacts, and collaborate on projects.	4	The app features an online community where learners may be able to communicate with other users, share learning artifacts, or collaborate on projects.	3	The app features an online community where learners can communicate freely with other users, but they are not able to share learning artifacts or collaborate on projects.
				2	The app includes an online community, but learners have little or no control over how they communicate with other users. Also, the ability to share learning artifacts or collaborate on projects is not available.
				1	The app does not allow learners to communicate or interact with other users.
					N/A
Dimension (A6) Accommodation of Individual Differences: Does the presentation of the app's content recognise that learners learn in uniquely different ways (e.g. visual, auditory, and/or kinaesthetic)?					
5	All of the app's content recognises that learners learn in different ways and caters to differentiated instructional practices.	4	Most of the app's content recognises that learners learn in different ways and caters to differentiated instructional practices.	3	Some of the app's content recognises that learners learn in different ways and caters to differentiated instructional practices.
				2	Very little of the app's content recognises that learners learn in different ways and caters to differentiated instructional practices.
				1	The app's content is presented in a format that assumes all learners learn in the same way.
					N/A

Dimension (B5) Navigation: How easily can learners move through the app's content and options?					
5	4	3	2	1	N/A
Learners can move through the app's content and options fluidly.	Learners need to put forth a little effort to move through the app's content and options.	Learners need to put forth some effort to move through the app's content and options.	Learners are somewhat impeded from moving fluidly through the app's content and options because of its organisation.	Learners encounter substantial challenges when trying to move through the app's content and options because of its disjointed organisation.	Not Applicable
Dimension (B6) Information Presentation: Is the app's content presented in a logical manner? (e.g. the app's content grows increasingly rigorous as learners experience success, the app activates learners' background knowledge before presenting them new information, and/or the app provides an overview of its content before learners engage in specific tutorials or activities.)					
5	4	3	2	1	N/A
The app's content is presented in a logical manner.	The app's content is presented in a manner that is mostly logical.	The app's content is presented in a manner that is somewhat logical.	The app's content is presented in a manner that is somewhat illogical.	The app's content is presented in an illogical manner.	Not Applicable
Dimension (B7) Media Integration: Are the app's texts, graphics, videos, sounds, and speech integrated effectively so each of the app's media components complements each other and forms a cohesive program? (focus: learning content)					
5	4	3	2	1	N/A
The app's text, graphics, videos, sounds, and speech are integrated seamlessly to form a cohesive program.	The app's text, graphics, videos, sounds, and speech are integrated adequately to form a cohesive program.	The app's text, graphics, videos, sounds, and speech are integrated poorly, but they still form a mostly cohesive program.	The app's text, graphics, videos, sound, and speech are mismatched to the app's content, but they still form a somewhat cohesive program.	The app's text, graphics, videos, sound, and speech are mismatched to the app's content, and it does not form a cohesive program.	Not Applicable
Dimension (B8) Cultural Sensitivity: Does the app use culturally responsive teaching methods to represent diverse populations?					
5	4	3	2	1	N/A
The app allows learners to connect to and share ideas with people from diverse communities across the globe.	The app allows learners to explore diverse cultures from across the globe using authentic pictures, images, and/or texts.	The app makes general references to one or more cultures.	The app makes reference to diverse cultures, but the information contains some cultural biases.	The app presents stereotypical images and information about cultures.	Not Applicable

Dimension (B5) Navigation: How easily can users move through the app's content and options?					
5	4	3	2	1	N/A
Learners can move through the app's content and options fluidly.	Learners need to put forth a little effort to move through the app's content and options.	Learners need to put forth some effort to move through the app's content and options.	Learners are somewhat impeded from moving fluidly through the app's content and options because of its organisation.	Learners encounter substantial challenges when trying to move through the app's content and options because of its disjointed organisation.	Not Applicable
Dimension (B6) Information Presentation: Is the app's content presented in a logical manner? (e.g. the app's content grows increasingly rigorous as learners experience success, the app activates learners' background knowledge before presenting them new information, and/or the app provides an overview of its content before learners engage in specific tutorials or activities.)					
5	4	3	2	1	N/A
The app's content is presented in a logical manner.	The app's content is presented in a manner that is mostly logical.	The app's content is presented in a manner that is somewhat logical.	The app's content is presented in a manner that is somewhat illogical.	The app's content is presented in an illogical manner.	Not Applicable
Dimension (B7) Media Integration: Are the app's texts, graphics, videos, sounds, and speech integrated effectively so each of the app's media components complements each other and forms a cohesive program? (focus: learning content)					
5	4	3	2	1	N/A
The app's text, graphics, videos, sounds, and speech are integrated seamlessly to form a cohesive program.	The app's text, graphics, videos, sounds, and speech are integrated adequately to form a cohesive program.	The app's text, graphics, videos, sounds, and speech are integrated poorly, but they still form a mostly cohesive program.	The app's text, graphics, videos, sound, and speech are mismatched to the app's content, but they still form a somewhat cohesive program.	The app's text, graphics, videos, sound, and speech are mismatched to the app's content, and it does not form a cohesive program.	Not Applicable
Dimension (B8) Cultural Sensitivity: Does the app use culturally responsive teaching methods to represent diverse populations?					
5	4	3	2	1	N/A
The app allows learners to connect to and share ideas with people from diverse communities across the globe.	The app allows users to explore diverse cultures from across the globe using authentic pictures, images, and/or texts.	The app makes general references to one or more cultures.	The app makes reference to diverse cultures, but the information contains some cultural biases.	The app presents stereotypical images and information about cultures.	Not Applicable

Domain (C) Engagement (Lee & Cherner 2015, adapted)					
Dimension (C1) Learner Control:					
Does the app allow learners to select the level (e.g. grade level, age level, or difficulty level) where they will begin engaging with content?					
5	4	3	2	1	N/A
The app allows learners complete control related to the level at which they begin engaging with content.	The app allows learners significant control related to the level at which they begin engaging with content.	The app allows learners some control related to the level at which they begin engaging with content.	The app allows learners little control related to the level at which they begin engaging with content.	The app allows learners no control related to the level at which they begin engaging with content.	Not Applicable
Dimension (C2) Interactivity:					
Does the app only present content to learners, or does the app allow learners to engage with the content through activities?					
5	4	3	2	1	N/A
The app always requires learners to engage with its content through activities.	The app often requires learners to engage with its content through activities.	The app occasionally requires learners to engage with its content through activities	The app only allows learners to interact with it by searching for specific content.	The app only presents content to learners. Learners are never asked to engage in activities.	Not Applicable
Dimension (C3) Pace:					
Does the app allow learners to control the rate at which they are presented content? If the app does not allow learners to control the rate at which they move through the content, does the app present the content at an appropriate rate for secondary level learners?					
5	4	3	2	1	N/A
The app allows learners to completely control the rate at which they move through the app's content.	The app allows learners to mostly control the rate at which they move through the app's content.	The app does not allow learners to control the rate at which they move through the app's content, but the content is presented at an appropriate rate for secondary level learners.	The app does not allow learners to control the rate at which they move through the app's content, but the content is presented at a mostly appropriate rate for secondary level learners.	The app does not allow learners to control the rate at which they move through the app's content, and the content is presented at an inappropriate rate for secondary level learners.	Not Applicable
Dimension (C4) Personalization:					
Can learners personalise the app by setting individual preferences (e.g. background music, images, colour scheme, font style or size, navigation menu, app interface language or language style, avatar, etc.)?					
5	4	3	2	1	N/A
Learners can personalise four or more features of the app.	Learners can personalise three features of the app.	Learners can personalise two features of the app.	Learners can personalise one feature of the app.	Learners cannot personalise the app.	Not Applicable

Dimension (C5) Interest: Will the app's language learning material (i.e. topics, activities) likely appeal to secondary level learners?					
5	4	3	2	1	N/A
The app's language learning material might be particularly interesting and engaging to secondary level learners.	Most of the app's language learning material might be interesting and engaging to secondary level learners.	Some of the app's language learning material might be interesting and engaging to secondary level learners.	The app's language learning material might not be interesting and engaging to secondary level learners.	The app's language learning material will likely deter secondary level learners from engaging with it.	Not Applicable
Dimension (C6) Aesthetics: Will the app's graphics and interface likely motivate learners to engage with it?					
5	4	3	2	1	N/A
The app's graphics and interface will motivate learners to engage with it.	The app's graphics and interface will most likely motivate learners to engage with it.	The app's graphics and interface may possibly motivate learners to engage with it.	The app's graphics and interface will likely have no impact on whether or not learners are motivated to engage with it.	The app's graphics and interface will likely deter learners from engaging with it.	Not Applicable
Dimension (C7) Gamification: Does the app use game-like elements (e.g. prizes, points, ranking, avatars, etc.) to boost learner engagement?					
5	4	3	2	1	N/A
The app uses four or more game-like elements which will potentially boost learner engagement.	The app uses three game-like elements which will potentially boost learner engagement.	The app uses two game-like elements which will potentially boost learner engagement.	The app uses one game-like element which will potentially boost learner engagement.	The app does not use game-like elements.	Not Applicable

Domain (D) Language Learning					
Dimension (D1) Quality of Input:					
Does the app provide comprehensible and relevant input (language and topics) for secondary level learners?					
5	4	3	2	1	N/A
The app continuously provides comprehensible input. It covers language that is highly relevant to secondary level learners.	The app provides mostly comprehensible input. It covers language that is mostly relevant to secondary level learners.	The app provides rather comprehensible input. It covers language that is rather relevant to secondary level learners.	The app sometimes provides input that is incomprehensible. It covers language that is aimed at the general public.	The app often provides input that is incomprehensible. It covers language that is aimed at a specific audience (e.g. people working in business).	Not Applicable
Dimension (D2) Opportunity to Produce Output:					
Does the app provide learners with the opportunity to produce written or spoken output?					
5	4	3	2	1	N/A
Learners are offered the opportunity to produce written or spoken output freely. Their output is not limited to predefined answers. The output goes beyond the sentence level.	Learners are offered the opportunity to produce written or spoken output rather freely. Their output is sometimes limited to predefined answers. The output is limited to the sentence level.	Learners are offered the opportunity to produce predefined written or spoken output. The output is limited to the sentence level.	Learners are offered the opportunity to produce predefined written or spoken output. The output is limited to the word level.	Learners are not offered the opportunity to produce any output when engaging the app. The app merely provides content for reception.	Not Applicable
Dimension (D3) Appropriateness of Language Learning Material:					
Is the language learning material (i.e. language, activities, topics) for specific language proficiency levels (e.g. A1) appropriate for the intended target learners (e.g. secondary level learners on level A1)?					
5	4	3	2	1	N/A
The level of the language learning material is always matched appropriately to the learners' level.	The level of the language learning material is usually matched appropriately to the learners' level.	The level of the language learning material is rarely matched appropriately to the learners' level.	The level of the language learning material is not matched appropriately to the learners' level.	The app's language learning material is inappropriate for language learning purposes.	Not Applicable

Dimension (D4) Skills: How many of the four core skills (reading, listening, writing, speaking) does the app train?					
5	4	3	2	1	N/A
The app integrates all four skills. Vocabulary or grammar practice might be included in the learning activities.	The app integrates three of the four skills. Vocabulary or grammar practice might be included in the learning activities.	The app integrates two of the four skills. Vocabulary or grammar practice might be included in the language activities.	The app trains at least one of the four skills. Vocabulary or grammar practice might be included in the language activities.	The app only provides vocabulary or grammar practice. None of the four skills are trained explicitly.	Not Applicable
Dimension (D5) Exposure to the Target Language: Does the app maximise exposure to the target language?					
5	4	3	2	1	N/A
The app only uses the target language in learning activities. No translations in the native language are provided, or they are only provided upon request. Certain app features (e.g. settings menu) might make use of the native language.	The app mostly uses the target language in learning activities. Some translations in the native language are provided. Certain app features might make use of the native language.	The app uses the target language and translations in the native language in learning activities to the same extent. Certain app features might make use of only the native language.	The app provides some exposure to the target language in learning activities. However, translations are very frequent. Other areas of the app are dominated by the native language.	The app only provides minimal exposure to the target language. The chosen native language dominates all areas of the app (e.g. learning activities, option menu, user interface).	Not Applicable

9.4. Transcripts of the think-aloud protocols and follow-up interviews

Transcription convention

I	Interviewer
P1	Participant 1
...	Short pause, long pause; within or between utterances
No-	Hyphens indicate a sudden cut-off
YES	Capitalised words indicate emphasis on an utterance
[laughs]	Comment about action, including non-verbal action
(xxx)	Indicates a stretch of unintelligible talk
(points)	Indicates explanatory note

Note: These transcripts display participants' trains of thought. Any German or English errors have not been corrected. Utterances in English have been transcribed according to English capitalisation rules. To improve readability, notes have been added.

Think-aloud protocol 01, 5th February 2021

I: Im Zuge dieser Think-Aloud-Session wirst du eine language learning app mithilfe des Evaluation Rubric for Language Learning Apps for Secondary Level Learners evaluieren. Der rubric deckt insgesamt 26 Dimensionen ab. Für jede Dimension sollst du der App zwischen einem Punkt und fünf Punkten geben. Kannst du eine Dimension bei einer App überhaupt nicht anwenden, dann wählst du auf der Skala die Option not applicable. Versuche bitte während des Evaluationsprozesses laut zu denken. Sage während du evaluierst bitte alles, das dir hinsichtlich des rubrics durch den Kopf geht. Besonders interessieren mich deine Gedanken hinsichtlich der Formulierungen, der Sinnhaftigkeit und Anwendbarkeit der einzelnen Domänen und Dimensionen. Während du evaluierst und dabei laut denkst, werde ich Notizen machen. Gibt es von deiner Seite aus noch offene Fragen?

P1: Think-aloud auf Deutsch oder auf Englisch?

I: Wenn du möchtest, dann auf Deutsch. Falls es keinen weiteren Fragen gibt, würde ich dich nun bitten mit dem Evaluationsprozess zu beginnen.

[Participant 1 commences with the evaluation]

[A1] Gut, dimension A1, ähm... da würde ich sogar not applicable werten, weil man bei dieser App eigentlich nur nachsprechen muss. Da geht es primär um die Aussprache, deswegen würde ich meinen, dass da nichts mit create oder analyse oder compare (passt), das würde für mich nicht passen, ...

[A2] ... das Gleiche für die 21st Century Skills (A2), die würde ich auch not applicable werten.

[A3] Dimension A3, ähm... da würde ich, das kommt auf die Aktivität an, weil es gibt viele Aktivitäten mit den incorrect (answers), es gibt zwar explanation aber die opportunity das nochmal zu machen kommt eigentlich ganz spät, wobei es ein paar Aktivitäten gibt wo man es sofort korrigieren kann.

[A4] Dimension A4, achso, da habe ich jetzt keine Punkte gegeben (referring back to A3). Dimension A3, da würde ich vier Punkte geben. Dann bei dimension A4, the app tracks learners academic growth, achievement, progress, das passt bei fünf (Punkte) aber auch bei drei (Punkte), the learners' progress can only be seen in the app and not by using, also es geht ein Teil von der Punkteanzahl fünf und von der Punkteanzahl drei, das heißt dazwischen, vier Punkte für dimension A4.

[A5] Dimension A5, einen Punkt, does not allow learners to communicate, einfach weil es nur zum Nachsprechen ist, weil es primär um die Aussprache geht, also gibt es da keine interaction.

[A6] Dimension A6, äh, Accomodation of Individual Differences, ja, jein, some of the app's content recognises that learners learn in different ways, aber daher das es eigentlich nur zum Nachsprechen ist, ja, zwei, drei Punkte.

[B1] Design (domain B), ähm... genau ob man die App wieder öffnen kann (B1), dann muss man sie (die Aufgaben) aber wieder neu auswählen, das ist bei drei (Punkte), exiting an app and then manually select.

[B2] B2 enhance by how it connects with other apps, other communities, learners' email. Also da geht es, da bin ich mir nicht ganz sicher, also connecten man kann sich, über Facebook anmelden (kann man sich), aber indem man seinen Account erstellt über Facebook, aber ob sie wirklich enhanced würde ich nicht meinen. Also ein Punkt.

[B3] The app's texts, graphics, videos (xxx) well organised (B3). Ja, würde ich schon sagen, usually organised, jetzt nichts... es würde jetzt nicht enhance den content aber es passt eigentlich, sind vier Punkte bei B3.

[B4] Bei B4, ähm, minimal guidance, es gibt keine instructions, aber es sollten die Schüler bei der Nutzung einfach so nutzen können.

[B5] How easily can users move through the app's content and options. B5, ja sollte ganz easy sein. Can move through the app's content and options fluidly, fünf (Punkte).

[B6] Is the app's content presented in a logical manner (B6), ja, ist presented in a logical manner, es geht immer von leicht bis schwer.

[B7] Are the app's texts, graphics, visuals, integrated effectively so each of the app's components complement each other and form a cohesive program (B7). The apps... ja, are integrated adequately to form a cohesive program, sollte passen.

[B8] Does the app use culturally responsive teaching methods to represent culturally diverse populations, ja vier (Punkte), also allows learners to explore diverse cultures, da war was mit Chinese New Year oder so, aber mit connect, share ideas with people from diverse communities geht bei dieser App nicht, deswegen nur vier (Punkte) bei B8.

[C1] Does the app allow learners to select the level where they will begin engaging content (C1). Does the app allow learners to select the level... ja, sie können verschiedene Levels auswählen, aber auch nur in der Pro-Version. Das heißt, man muss sich eigentlich extra anmelden. Aber ansonsten generell ja. Five (points), the app allows learners complete control related ... (xxx), fünf (Punkte).

[C2] Does the app only present content to learners or does the app allow learners to engage the content through activities. (C2). Ähm... eigentlich kein engagement, ich weiß nicht ob das Nachsprechen als engagement gilt. Ähm ... und ob das Nach... oder dieses präsentieren von den ganzen Wörtern und Sätzen einfach nicht nur content ist. Oder ist das Nachsprechen auch engaging, ich weiß es nicht. Ähm... deswegen, ja, schwierig... vielleicht not applicable in this case... maybe, I don't know. Probably not applicable.

[C3] Dimension C3 Pace. Does the app allow learners to control the rate at which they are presented content. The rate... wie schnell man durch (xxx). Does the app present the content at an appropriate rate for lower secondary learners... Naja, die rate ist vorgegeben, glaube ich... die kann man nicht kontrollieren, aber für lower secondary learners sollte es passen. Also drei Punkte für C3.

[C4] C4, can learners personalise the app by setting individual preferences. PERSONALISE... background, music, images, avatars, nein kann man nicht personalisieren, man kann ein Bild von seinem Foto, also ein Foto von seinem Profil hochladen, aber sonst kann man es nicht personalisieren, so one point.

[C5] Ähm... will the app's language learning material likely appeal to lower secondary learners (C5). Ja, most of the app's (language learning material) not all (of the) app's (language learning material). Es sind verschiedene Themen, viele oder die meisten sind für die lower secondary passend. Four (points).

[C6] C6, graphics and interface are fine. Die sind sehr ansprechend und würden sicher motivieren. So five points for C6.

[C7] C7 Gamification. Does the app use game-like elements. Yes, prizes, points, ranking, avatars, alles mögliche ist dabei, so five (points).

[D1] Language learning. Gut, (domain) D. Does the app provide comprehensible, valuable or relevant input... yes, it provides mostly comprehensible, valuable input, weil es erklärt ja wo die Fehler sind, aber es ist nicht unbedingt relevant für lower secondary sondern eher specific audience würde ich sogar sagen. Also für die Leute die unbedingt ihre pronunciation lernen oder verbessern möchten. Also von vier, the app provides mostly comprehensible input, aber von eins, it is aimed at a specific audience wenn nicht sogar general public. Ja, zwei Punkte wahrscheinlich.

[D2] D2, does the app provide learners with the opportunity to produce written or spoken output. Ja, ähm, but not freely...predefined, produce predefined und at the sentence level, limited to the sentence level, also Wörter und sentences, ja genau.

[D3] Is the app's language learning material for specific grade-levels appropriate for its target audience (D3). Ja... for specific grade levels, appropriate for its target audience [pauses and thinks]. Kann ich bei einer pronunciation app nicht so beurteilen... Weiß ich nicht. Also es würd schon passen für jeden, aber ich weiß ich nicht was ich jetzt für diese App geben würde. Usually matched, ja, usually matched appropriate to its intended learners. Also four (points).

[D4] How many skills does the app train. Naja, nur pronunciation, ist nicht wirklich einer von den four skills. All of the four skills, three of the four skills. Part of the speaking skills, but not really training the speaking skills. Das kommt darauf an ob man das jetzt als speaking gelten lassen kann. Entweder zwei (Punkte), at least one of the four skills, oder eben not applicable, weil es ja eben nur um das... wobei doch man hört ja auch Konversationen, also zwei (Punkte), at least one of the four skills and vocabulary and grammar practice might be integrated, genau.

[D5] And does the app maximise exposure to the target language. Yes, only uses target language. Kein... also hätte ich beim drüber schauen jetzt eigentlich nicht gemerkt, dass die native language irgendwo dabei wäre. Ja, only uses the target language.

[Participant 1 finishes the evaluation]

[The interviewer commences the follow-up interview]

I: Alright, bist du mit deiner Evaluation fertig?

P1: Yes.

I: Gut, ähm... dann habe ich mir jetzt zwei dimensions aufgeschrieben bei denen ich gemerkt habe, dass du etwas steckst, oder das die etwas schwierig zu verstehen sind.

P1: Ja.

I: Und zwar sind beide aus D, aus der language learning domain, einmal D3 und D4. Bei D3-

P1: Ja.

I: Ist es für dich da verständlich was mit der dimension wirklich gemeint ist wenn du dir das durchliest?

P1: Ich habe nicht ganz, also ich verstehe nicht ganz was mit specific grade-levels gemeint ist.

I: Mhm.

P1: Language learning materials for specific grade levels-

I: Wie würdest du das interpretieren. Wonach wird da gefragt?

P1: Specific... is the app's language learning material for specific grade-levels appropriate for its target audience. Ob die activities passend sind für verschiedene Schulstufen nehme ich mal an... in der App.

I: Okay, so würdest du das jetzt interpretieren?

P1: Genau.

I: Okay. Im Grunde ist damit gemeint wenn in den Apps direkt gekennzeichnet ist okay das ist A1, also Level A1, oder das wäre eine speaking für fifth grade oder sixth grade-

P1: Ja.

I: Und da ist dann gemeint ob die dann auch wirklich für das level appropriate sind.

P1: Achso, aha.

I: Das ist bei deiner App jetzt ein bisschen gemein, weil die activities nicht wirklich gekennzeichnet sind mit grade-levels, sondern es ist ja alles irgendwie zusammengemischt.

P1: Ja es ist nur easy, medium und hard oder difficult.

I: Okay, ja, aber es ist sehr interessant für mich, weil ich dadurch merke, dass bei dieser Dimension das anders formuliert werden müsste, vielleicht so, dass gleich klarer ist was damit gemeint ist. Gut das war jetzt mal die D3. Und bei der D4-

P1: Ja.

I: Hast du dann gemeint du würdest pronunciation nicht als einen skills bezeichnen und sagen ein skill wird trainiert, das würde darunter fallen unter das training speaking skills, aber es ist noch nicht ein kompletter skill.

P1: Nein, weil wenn hier steht four skills, at least one of the four skills, dann denke ich mir bei four skills, reading, speaking, listening und writing.

I: Mhm.

P1: Also an diese four major categories würde ich eher denken. Also so habe ich das dann verstanden, dass wenn diese App die Aussprache trainiert, dass das zwar unter speaking skills fällt, aber es nicht einer von den vier skills ist, weil speaking ja mehr als nur Aussprache ist.

I: Mhm. Ja, okay. Da gibt es ja beim einen Punkt bei der scoring scale the app only provides vocabulary or grammar practice, none of the four skills are trained explicitly.

P1: Mhm.

I: Würdest du dann sagen, dass es vielleicht darunter fällt, und dass man dann den descriptor ein bisschen umformulieren müsste?

P1: Naja, wenn da steht none of the four, none of the four skills are trained explicitly, würde ich auch nicht sagen, weil es ist ja ein Teil von speaking ist, und irgendwie wird es schon teilweise trainiert, weil du ja auch so phrases für Konversationen hast.

I: Okay, das heißt der letzte Satz, none of the four skills, den müsste man vielleicht ein bisschen umformulieren damit dann eindeutig ersichtlich ist, okay es wird zwar nur ein Teilbereich dieser skills trainiert, aber das wäre trotzdem ein Punkt wert.

P1: Genau, ja.

I: Okay, mhm. Bei den anderen domains bei A, B und C habe ich jetzt nicht so viel mitgeschrieben, weil ich nicht das Gefühl hatte, dass da große Schwierigkeiten geherrscht haben. Bei den Formulierungen auch. Ich würde jetzt noch, wenn es für dich okay ist, ein paar allgemeine Fragen zum rubric stellen.

P1: Mhm. Ja.

I: Die erste Frage meinerseits ist: inwiefern würdest du diesen rubric in deiner persönlichen Praxis einsetzen?

P1: Ich persönlich für meinen Beruf?

I: Mhm.

P1: Puh. Ähm... nur wenn ich muss, also weil es ist schon sehr viel und ähm... muss aber ehrlich sagen, dass ich mit Apps wenig arbeite, ähm... aber vielleicht in einer verkürzten Form, aber das wär schon sehr, aber es ist mehr der Umfang, es ist nicht der rubric an sich, der ist nicht schlecht, aber es ist wirklich sehr detailliert und es werden viele Aspekte angesprochen, ähm... aber es ist mehr der Umfang der sehr, also es ist mehr so aus zeitlichen Gründen wenn da so viel ist.

I: Das schreckt dich ein bisschen ab höre ich da raus?

P1: Ja.

I: Okay, also du findest ihn zu ausführlich. Und wenn es eine Lightversion geben würde bei der, zum Beispiel, von jeder dimension die drei Wichtigsten herausgenommen werden, dann hat man insgesamt zwölf, wäre es dann attraktiver, dass man den rubric wirklich einsetzt wenn man mit Apps arbeitet?

P1: Ja, schon eher, weil die... die... für jede... also es ist wirklich sehr detailliert und für jeden Punkt den du vergeben möchtest sind die Texte ja auch ziemlich lang, also schon recht viel. Aber eine Lightversion wäre sicher gut.

I: Okay. Ähm... weil du jetzt gerade die Formulierungen angesprochen hast von den descriptors-

P1: Ja.

I: Gibt es Formulierungen der dimensions, also die einzelnen dimensions A1, A2, A3 und so weiter, hast du da eine gehabt wo du wirklich konkret Schwierigkeiten hattest zu verstehen was damit gemeint ist?

P1: Ähm... nur das was wir besprochen haben mit specific grade-levels, das war mir, wie gesagt, nicht so ganz klar. Ähm... das mit dem does the app allow learners to control the rate (C3) at which they are presented content and if not-

I: C3. Pace, oder?

P1: Genau. Da habe ich mir auch zuerst überlegen müssen was ist gemeint mit pace, ob die pace vorgegeben ist von der App oder meine eigene pace, ähm... ja. Ansonsten ist es eigentlich ganz klar, und viele unterscheiden sich sicher nur durch ein Wort wie little, some, significant, complete, was man eh von den anderen rubrics kennt.

I: Okay, würdest du bei den Adverbien, wenn zum Beispiel der Unterschied zwischen fünf (Punkte) und vier (Punkte) nur ein Wort ist, also bei fünf (Punkte) zum Beispiel frequently und bei vier (Punkte) dann nur often, würdest du dir jetzt schwer tun da wirklich eine Entscheidung zu treffen zwischen vier (Punkte) und fünf (Punkte) wenn es nur von einem Wort abhängt?

P1: Ja, das kommt auf das Adverb an, weil bei B7 zum Beispiel war es bei mir klar seamlessly und adequately, also für mich, ich verstehe als seamlessly perfekt und adequately passend, in dem Sinne es ist jetzt nicht perfekt aber es passt. Also es kommt glaube ich auf das Adverb an. Aber wenn jetzt so etwas wie frequently und often zu nahe aneinander ist, dann würde ich mir schon schwer tun. Aber wenn es wirklich klar formuliert ist, dann sollte es eigentlich kein Problem sein.

I: Okay. Wenn man jetzt weggeht von der Formulierung, findest du den rubric im Großen und Ganzen passend von den Formulierungen her und vor allem von den Aspekten die durch den rubric überprüft werden sollen. Oder findest du es gibt eine oder mehrere dimensions die man rauswerfen könnte, die überhaupt nicht passen um eine App zu evaluieren?

P1: Nein, also ich finde die domains eigentlich recht gut. Also educational value finde ich sehr interessant, also als domain. Ich meine language learning sowieso, weil das ist eine language learning app. Engagement das passt auch. Design auch, das hat auch einen wichtigen Einfluss. An educational value hätte ich persönlich vorher nie gedacht, also so als extra domain, aber ich finde eigentlich alle passend.

I: Okay. Wir haben jetzt schon sehr viele Fragen besprochen. Was mich jetzt noch interessieren würde, ähm... die fünfteilige Evaluationsskala, was denkst du über die? Wenn du jetzt mit dem evaluation rubric evaluierst, findest du es okay, dass du da eine fünfteilige Skala hast, findest du es vielleicht angenehmer weil es ein bisschen differenzierter sein kann, oder würdest du mit einer vierteiligen Skala auch leben können?

P1: Das hängt von den descriptors ab. Also das hängt davon ab wie die einzelnen descriptors beschrieben sind, also wenn das klar dasteht was vier (Punkte) bedeutet, was drei (Punkte) bedeutet, dann komme ich mit vier (Punkten) auch aus.

I: Mhm.

P1: Mit fünf... also... ob es jetzt vier oder fünf (Punkte) sind, ist jetzt glaube ich nicht so das Ausschlaggebende, sondern mehr wie es beschrieben ist, also von mir aus sechs, sieben, aber wenn es klar dasteht was sieben bedeutet, was fünf bedeutet, ist es eigentlich egal. Also... ja.

I: Und die not applicable Option die man da gibt, glaubst es könnte bei gewissen Dimensionen dazu verleiten, dass man es einfach auswählt obwohl man eins bis fünf (Punkte) vergeben könnte?

P1: Nein, also das kann schon sein, aber ich glaube, dass das nicht so schlecht ist, weil es regt dich dann doch irgendwie an genauer nachzudenken, okay... es passt nicht, aber vielleicht passt doch irgendwas von den fünf anderen descriptors. Also ich finde es grundsätzlich nicht schlecht. Aber natürlich kenne ich das, dass man sagt keine Lust, weiß nicht was da steht, aber wenn man sich es genauer anschaut, dann würde ich meinen, denkt man dann doch eher genauer nach ob nicht vielleicht doch einer der descriptors passt.

I: Mhm. War es für dich aus der Arbeit mit dem rubric ersichtlich, dass das eigentlich für language learning apps produziert oder formuliert ist die auf secondary, also lower secondary learners abzielen?

P1: Nur auf lower secondary?

I: Mhm.

P1: Hätte ich jetzt nicht gesagt, nicht nur auf lower secondary. Also das ist erst wo es erwähnt worden ist mit... wo war das [looks up dimension C3], also bei C3 zum Beispiel da fängst das erst an mit lower secondary learners. Und ich hätte, wenn ich von A bis C von Anfang an durchgegangen wäre, nicht gedacht, dass es nur für ähm... lower secondary gestaltet ist. Also es hätte schon für upper secondary auch gehen können, weil eben auch dieses Depth of Knowledge mit den extended thinking skills vorkommt, create, synthesize, compare, das sind ja so Skills die ich mir eher nicht so sehr in der Unterstufe vorstelle, kann man schon machen und so, aber... wäre jetzt nicht... auch dieses 21st Century Skills, critical thinking, communication, collaboration, self direction, also da hätte ich schon gedacht, dass das nicht nur für die lower secondary-

I: Ja, gut. Das wäre meine Folgefrage gewesen. Ob du A1 und A2 eher in die Oberstufe verortest, wenn man da den rubric einsetzt um eine App zu evaluieren, oder ob es für die Unterstufe auch irgendwie sinnvoll ist die zwei dimensions da drin zu haben. Dann bleibe ich jetzt noch kurz bei den 21st century skills hängen. Wenn du jetzt den rubric anschaut, dann steht da unten in einer Fußnote wo man die abrufen kann, wo man die ansehen kann. Wenn da jetzt steht the app engages 21st century skills, wüsstest du dann was damit gemeint ist?

P1: Ohne jetzt da drauf zu klicken oder wie?

I: Ohne darauf zu klicken-

P1: Würd es mir-

I: Und auch das was in Klammer steht, würdest du wissen was das ist oder wäre das für dich so unbekannt?

P1: Ich hab es mir jetzt nur so zusammengereimt mit den ganzen... eben was da dann stand mit critical thinking, communication also... so circa weiß ich schon was gemeint ist, aber jetzt ganz genau-

I: Mhm.

P1: Hab aber jetzt auch nicht da nachgeschaut bei der Website, aber wenn da jetzt diese explanation in der Klammer nicht dabeigestanden wäre, dann hätte ich es wahrscheinlich nicht gewusst oder nur geraten. Ich meine ich wäre sicher eh auch in die Richtung gegangen, aber ich hätte es jetzt nicht ganz genau gewusst.

I: Mhm. Okay. Gut. Gibt es noch irgendetwas, das du zu dem rubric sagen möchtest?

P1: Nein. Nothing to add.

Think-aloud protocol 02, 5th February, 2021

I: Im Zuge dieser Think-Aloud-Session wirst du eine language learning App mithilfe des Evaluation Rubric for Language Learning Apps for Secondary Level Learners evaluieren. Der rubric deckt insgesamt 26 Dimensionen ab. Für jede Dimension sollst du der App zwischen einem Punkt und fünf Punkten geben. Kannst du eine Dimension bei einer App überhaupt nicht anwenden, dann wählst du auf der Skala die Option not applicable. Versuche bitte während des Evaluationsprozesses laut zu denken. Sage während du evaluierst bitte alles, das dir hinsichtlich des rubrics durch den Kopf geht. Besonders interessieren mich deine Gedanken hinsichtlich der Formulierungen, der Sinnhaftigkeit und Anwendbarkeit der einzelnen Domänen und Dimensionen. Während du evaluierst und dabei laut denkst, werde ich Notizen machen. Gibt es von deiner Seite aus noch offene Fragen?

P2: Nein.

I: Dann würde ich dich nun bitten mit dem Evaluationsprozess zu beginnen.

[Participant 2 commences with the evaluation]

[A1] Gut, also ich evaluiere die App Mondly und fange einfach mit der dimension A1 Thinking Skills an. Und bei diesem, bei diesen vier ähm... thinking Stufen ist halt die bei der Mondly nur ähm... würde ich die zwei (Punkte) nehmen, the apps content best aligns to level one. Wobei ich mir da nicht sicher bin ähm, weil ja in der Stufe eins the app doesn't align to Webb's Depth of Knowledge ähm... ob das eben nicht das Gleiche wäre wie not applicable. Which level of Webb's Depth of Knowledge must learners engage to meet the apps learning objectives, da wäre die Nummer eins fast ähnlich wie not applicable, weil eben keines dieser vier Levels dann gegeben wäre. Aber auf jeden Fall wäre die App Mondly eindeutig Level zwei, weil in Klammer diese Wiederholung der Skills hilft, und das ist recall, recognise, match and memorise. Die Sachen in Klammer helfen da sehr diesen rubric anzuwenden.

[A2] Dimension A2 21st Century Skills, ähm... the app sometimes requires learners to engage 21st century skills, wobei da aber eben diese 21st century skills (xxx), creativity and collaboration wird nicht gefördert, aber hauptsächlich dann media literacy, also apply technology effectively, weil es eben so einen Chatbot gibt der SMS imitiert, und einen augmented reality Teil der App. Also es ist hauptsächlich media literacy und dieses work independently und communication, collaboration das alles ist eigentlich nicht dabei, das ähm... 21st century skills sind halt sehr weit gefasst und in dem Fall ist es nur eines davon, media literacy, aber das ziemlich häufig.

[A3] Gut, dimension A3 Value of Errors. Ähm... da war es ziemlich gut ausformuliert, also hier wäre es in dem Fall number two, the app offers learners an opportunity to answer the question again, but the app does not provide information to explain why the answer was

incorrect the first time. Es ist halt oft schon markiert was falsch ist, aber eben nicht, ähm, wieso es falsch ist. Aber es ist auch oft schwer, vor allem bei grammatikalischen Sachen, wenn es heißt es ist grammatikalisch falsch zu sagen wieso. Auf jeden Fall wird halt trotzdem keine Antwortmöglichkeit gegeben oder nicht der Grund gegeben wieso etwas falsch ist. Ich glaube das ist oft auch schwer zu sagen, wenn es falsch ist, ist es halt falsch.

[A4] Dann dimension A4 Feedback to Teachers, da hätte ich fast gesagt ähm... not applicable, weil es eben keine App ist für den Unterricht sondern für Selbstlernende, wenn man es sich selbst beibringen möchte, so kommt es mir vor. Es gibt also keine Klasse die man einrichten kann, soweit ich es gemerkt habe, und man kann sich selbst mit anderen vergleichen, aber es gibt keinen Administrator oder Lehrer der da irgendetwas anschauen kann, und deswegen hätte ich gesagt how does the app allow teachers to monitor the learners' progress is not applicable. Weil es bei dieser App eben keine solche Klassenfunktion meines Wissens nach gibt... aber die rubrics und die Abstufungen sind eigentlich klar formuliert, und es ist halt nur so, dass eben der Prozess nur für die Lerner, für die Lernenden einsehbar ist und nicht für jemanden anderen. Es gibt wohl learner progress den man anschauen kann, aber eben nicht als Lehrer sondern nur der Schüler selbst. Also vielleicht könnte man das noch einbauen, dass nicht nur der Lehrer sondern auch die Schüler, weil die Schüler sehen es ja schon, dass die dann eben sich selbst beobachten können, aber die Lehrer können es nicht machen.

[A5] A5 Cooperative Learning, ähm... da habe ich immer three (points), the app features an online community where learners can communicate freely with other learners. They are not able to share learning artifacts or collaborate on artifacts or projects, also es gibt nur so eine Ranking Liste, man kann Level aufsteigen und das kann man teilen auf Facebook oder auf anderen, also es gibt dann schon eine online community, es gibt ein Ranking, aber man kann nicht zusammenarbeiten soweit ich das gesehen habe, aber da ist die (dimension) eigentlich auch ziemlich eindeutig und ja...

[A6] Dimension A6 Accomodation of Individual Differences, ähm... da habe ich auch nur three (points), some of the app's content recognises that learners learn in different ways and caters to differentiated instruction practices. Da ist es halt auch oft schwer auf einem Handy die verschiedensten Lerntypen einzubauen, weil es halt doch auf einem Bildschirm ist, und da versucht die App das eben visual, teilweise mit augmented reality, auditory... man kann sich immer, es läuft viel über das Gehör und kinaesthetic eben, dass man aufstehen muss und um die Katze herumgehen muss damit man das Wort tail dann bekommt. Aber trotzdem gibt es halt Einschränkungen bei einer App. Gut, irgendetwas wollte ich noch sagen... es ist auch immer gut, dass man das Mikro auswählen kann und Antworten mündlich auch geben kann. Das wäre vielleicht öfter notwendig, das ist nur selten verpflichtend, dass man wirklich was sagt, aber hin und wieder kommt es schon vor, aber man kann aber immer statt zu schreiben etwas sagen. Und die rubric ist eigentlich da schon ziemlich eindeutig, dass entweder all of the apps, most of the apps, some of the apps, das ist eigentlich eine eindeutige Einteilung und das kann man dann schon so bewerten.

[B1] Domain B Design, dimension B1 Ability to Save Progress... das ähm... war für mich nicht gleich bewertbar, weil wenn man aussteigt, dann wird man gefragt ob man aussteigen will, und trotzdem, und eben dann geht das verloren was man bis dahin gemacht hat, also wenn man gerade mitten in einer Lektion ist. Und deswegen ist es halt dann nicht das was man

zuletzt gemacht hat, weil man muss auch etwas wiederholen, dann wäre es so wie bei einem Spiel bei einem Speicherpunkt. In close proximity to where they were when they logged in, aber ich muss das nicht weitermachen ich kann woanders weitermachen. Also die App, da muss man, es gibt wohl so Speicherpunkte eben wenn man eine neue Lektion anfängt, und ich war mir da eben nicht sicher ob ich jetzt ähm... drei oder vier (Punkte) nehmen sollte. Weil man eben zwar nicht da ankommt wo man stehen geblieben ist, wenn man davor sagt OK ich möchte es nicht speichern weil ich muss jetzt öfter üben, was eigentlich in close proximity ist, aber ich werde ja vorher gefragt, also ist es dann meine Schuld, und wenn ich die Lektion abgeschlossen habe, dann kann ich gleich wieder anfangen wo ich dann weitermache, also da wäre es dann eher drei (Punkte) und wenn ich nicht speichere wäre es eher vier (Punkte).

[B2] Gut, dimension B2 Integration... it connects to other apps, online communities, independent websites and learners' email, ähm... ich habe da halt nur drei (Punkte), weil es mit einer online community interacted, weil es eben ähm auf Facebook, man kann über Facebook Freunde einladen und auch über Email Freunde einladen, andere Apps kommen nicht vor, online communities schon und independent websites nicht wirklich, kommen nicht wirklich vor, deswegen habe ich drei (Punkte) genommen. Der rubric war eigentlich klar durch diese vier Punkte, wobei mir eben eins the app does not integrate with any of the listed connection wäre für mich das gleiche wie not applicable, (xxx) wieder dass es nur vier Möglichkeiten hier oben gibt und nicht fünf. Für mich wäre eins und not applicable hier fast das Gleiche, aber ja... die, vielleicht dass man noch eins hinzufügt oder vielleicht nur vier macht, und wenn not applicable immer ein extra Punkt sein soll dann passt das schon so.

[B3] Dimension B3 Screen Design... Ja es war... normalerweise bei einer App kann ich mir nicht vorstellen, zumindest bei einer App die man bezahlen muss sowie es diese ist, dass da irgendwas unorganisiert ist oder verwirrend, aber ähm... es ist halt sehr viel hineingepackt in diese dimension, weil eben graphics, videos, sound and speech, man könnte es vielleicht trennen, dass man sagt graphics and videos und sound and speech. Weil eben die graphics dann doch ähm... wichtig sind, videos gibt es in der App keine, aber so das Graphische vielleicht von dem Sprachlichen oder von dem content auch trennen, weil graphisch ist oft auch nur der Hintergrund, und wenn der Hintergrund dann ein bisschen verwirrt weil er so übermäßig schön gestaltet ist, dass er ablenkt, dann hat das nichts mit sound und speech zu tun. Vielleicht könnte man das noch aufteilen ein bisschen, dass man das trennt irgendwie graphics und sound zum Beispiel als zwei extra dimensions macht.

[B4] Dann dimension B4 Ease of Use. Is the app intuitive and are learners able to engage it with minimal guidance. Hier habe ich four (points) learners are able to engage the app immediately with minimal guidance. Da ist die Frage ob die guidance von der App kommen muss oder von einem Lehrer oder von einem Außenstehenden. Auf jeden Fall hat die App einen introductory course und zeigt einem eben wie die App funktioniert. Wenn das als guidance zählt, brauche ich das schon zuerst weil es eben, aber ich glaube es ist verpflichtend. Vielleicht kann man es auch überspringen, aber ob eben guidance hier von der App selbst kommt oder von einem Lehrer (xxx). Guidance von der App brauche ich schon durch diesen introductory course, aber von dem Lehrer eben oder von dem Außenstehenden brauche ich überhaupt nichts, es erklärt mir alles die App, und deswegen the app is intuitive because it explains everything, es erklärt wirklich alles ähm... vorher bevor man mit der ersten Lektion anfängt, und so kompliziert ist es auch nicht, das schafft

schon jeder intuitiv, aber guidance von einem Außenstehenden braucht man nicht. Vielleicht kann man noch dazuschreiben, dass guidance by the teacher oder irgendjemandem halt oder der app gemeint ist.

[B5] Dimension B5 Navigation. How easily can learners move through the app's content and options. Da habe ich mir glaube ich auch schwerer getan, ähm... ich bin nicht, wenn jetzt gemeint ist multiple clicks or swipes da kann ja auch Werbung gemeint sein. Dass man fünf Sekunden warten muss und die Werbung anschauen muss das man weiterkommt, und das ist auf jeden Fall bei der App nicht der Fall, es gibt sehr viele Lektionen und die sind halt auf verschiedene Screens aufgeteilt, deswegen kann es schon sein, dass man um zu einer bestimmten Lektion zu kommen wieder zwei-, dreimal weiterklicken muss bis man sie gefunden hat. Deswegen, ähm... es ist eigentlich schon so, dass man flüssig wechseln kann, aber dass es eben eine breite Masse an Angebot gibt, dann muss man halt auch viel wechseln, und man braucht dann halt mehrere Klicks, aber es geht trotzdem flüssig. Trotzdem ist es keine disjointed organisation weil es eigentlich in Themengebiete geteilt ist, aber die Themengebiete muss man sich halt alle anschauen bis man weiß welche Lektion man dann als nächstes machen muss, also man braucht dazu schon mehrere Klicks, aber das hindert jetzt nicht unbedingt, sondern das braucht man einfach damit der Bildschirm übersichtlich bleibt, und damit man nicht 40 Lektionen auf einem Bildschirm haben muss. Gut, das mehrere Klicks die Organisation auch verbessern könnten, und das hat mich bei dieser Dimension ein bisschen irritiert, das es für diese App zum Beispiel übersichtlicher macht dadurch dass man öfter klicken muss.

[B6] Gut, dimension B6 Information Presentation. Is the apps content presented in a logical manner. OK, das passt zum Oberen dazu. Da ist die Dimension, der rubric eigentlich auch klar wieder mit is presented, is mostly logical, somewhat logical, somewhat illogical, also da ist er ziemlich klar der rubric. Da habe ich mich auch leicht für die vier (Punkte) entscheiden können weil nicht alles ganz logisch ist. Man kann viele Lektionen vorziehen, aber man muss nicht die Reihenfolge der App befolgen, deshalb ist es schon logisch, aber die Themengebiete sind oft ein bisschen willkürlich gewählt scheint mir. Deswegen habe ich mostly logical gewählt, aber der rubric ist eigentlich klar.

[B7] Dimension B7 Media Integration. The app's text, graphics, sound and speech integrate effectively... (xxx). Mmh... puh... da habe ich auch vier (Punkte) genommen, are integrated adequately, ähm... und form a cohesive programme. Ich wüsste da eben jetzt nicht was, vielleicht kenne ich keine App die nicht cohesive ist, aber das mit cohesive programme... zusammenhängende App das ist ein bisschen, also eine nicht zusammenhängende App ist für mich nur schwer, in dem Fall hängt es anscheinend gut zusammen und es ist jetzt nicht so, dass es ganz viele verschiedene Funktionen gibt, dass die da nicht cohesive wäre. Also das Einzige das ist, dass man hier eine zweite App herunterladen muss für einen Teil, das ist fast ein Viertel der App, die wie augmented reality funktioniert, und deswegen ist es eigentlich kein cohesive programme sondern es sind zwei cohesive programme, weil man muss die andere App zuerst herunterladen und dann einsteigen damit man die Lektionen dort machen kann. Also deswegen auch nur adequately, es ist nicht nur ein Programm sondern es sind zwei Apps die man herunterladen muss. Aber text, graphics, videos, sound and speech sind eigentlich integrated also bei dieser dimension wüsste ich jetzt eben nicht wie eine App aussieht wo das nicht der Fall ist.

[B8] Gut, dimension B8 Cultural Sensitivity. Does the app use culturally responsive teaching methods, da habe ich not applicable gewählt, weil bei dieser dimension B8 macht die App da eigentlich keine ähm... keine Anmerkungen zu verschiedenen Kulturen aber auch keine zu typischen Anmerkungen. OK, vielleicht ist das was ich gemacht habe ziemlich einfach. OK, es gibt ein Schlafzimmer, OK, vielleicht gibt es nicht in allen Kulturen ein Schlafzimmer, aber ähm es ist nicht so als wär es ähm wirklich stark stereotypisch von einer Kultur beeinflusst. Das was vielleicht schon dazugehört ist, dass man zwischen ähm amerikanischem Englisch und britischem Englisch entscheiden kann, das gehört vielleicht auch zu culturally responsive teaching, dass man zumindest diesen Unterschied macht zwischen der Sprache dieser zwei Kulturen sozusagen, aber ich konnte keinen cultural bias finden da es doch sehr einfache ähm Themen sind wie animals und housing und travel. Also vielleicht höchstens beim restaurant, dass die Gerichte eher europäisch angehaucht sind, also stimmt der, also beim restaurant muss man was bestellen und da sind dann Gerichte wie grilled fish and spaghetti und sowas. Das kann man dann schon sagen dass sie vielleicht, ich wüsste jetzt nicht zu welchem Land der bias wäre weil es doch sehr viele verschiedene Gerichte von Spaghetti bis Burger gibt, aber es ist zumindest nicht so, dass es sehr verschiedene Kulturen gibt sondern doch eher eurozentrisch wirkt. Aber ich habe jetzt trotzdem not applicable gewählt weil eben für mich ähm... das keiner speziellen Kultur zugehört, also die Gerichte oder die Begrüßungen nicht spezifisch auf eine Kultur, sonst wäre mir nichts aufgefallen was kulturspezifisch ist. Auf jeden Fall was es eben nicht ist, dass sie aktiv kultursensitiv arbeitet, und deswegen hab ich eben nicht eins aber auch nicht fünf (Punkte), aber eins auch nicht weil es nicht so stereotypische Bilder zeigt wie amerikanisches Englisch nur Burger, sondern dann sind wohl Essen die man in Europa genauso wie in Amerika findet, und ich glaube nicht, dass sich die Gerichte je nachdem welche Sprache man mit dieser App lernt (ändern), und ich glaube da sind überall die gleichen Gerichte und deswegen (xxx). Also da ist es weder stereotypical für mich also weder eins noch fünf (Punkte) und deswegen habe ich not applicable gewählt.

[C1] Domain C Engagement, dimension C1 Learner Control. The app allows learners significant control over the level which they begin engaging its content, also man kann die Lektionen aussuchen, und deswegen kann man ähm sehr kontrollieren was man lernt. Das Level im Prinzip kann man nicht so gut kontrollieren aber eben die Themen. Für mich ist das Thema Hotel anspruchsvoller als die Begrüßung, und der rubric ist eigentlich auch klar mit complete control, significant control, some control, da ist der rubric eigentlich ziemlich klar, und grade, age level und difficulty level das kann man am Anfang alles eingeben. Also da ist der rubric ziemlich eindeutig, da hatte ich keine Probleme etwas auszusuchen.

[C2] Dimension C2 Interactivity, da war auch wieder der rubric eindeutig mit frequently und often, ähm... ja vielleicht zwischen vier (Punkte) und fünf (Punkte), also vielleicht bei fünf (Punkte) always und nicht frequently nehmen, weil frequently und often sind fast das gleiche. Auf jeden Fall bis auf das war der rubric eindeutig. Die App war auch eindeutig eine fünf, da der ganze Inhalt durch Aktivitäten beigebracht wird. Da gibt es keine Lektion wo man selbst nicht aktiv mitarbeiten muss, und deswegen hätte ich statt frequently always geschrieben. The app always requires its learners to engage content through ist activities. Die app besteht fat nur aus Aufgaben, eigentlich nur aus Aufgaben und deswegen, auch der Input ist über Aufgaben, wenn man sich dann entscheiden kann was man sagt. Man kann sich entscheiden, es ist alles sehr learner centered und sehr interaktiv. Die Kategorie fünf würde ich vielleicht always statt frequently schreiben.

[C3] Dimension C3 Pace. Da habe ich Nummer drei gewählt, the app does not allow learners to control the rate at which they move through the app's content. Das man die einzelnen Lektionen zu verschiedenen Themen machen kann, war das für mich ähm eher, die App war für mich eher nicht so auf lower secondary learners ausgelegt, sondern für jeden, auch für jemanden der wirklich in der Erwachsenenbildung oder zumindest jetzt als Erwachsener sich selbst wieder Englisch beibringen will, und so ist es nicht unbedingt eine App die man in der Schule verwenden muss aber kann, aber können sicher, aber es ist auch ohne Schule schaffbar, leicht schaffbar die App zu bewältigen. Deswegen appropriate rate for lower secondary learners, das betrifft halt alle Sprachenlerner und nicht nur die lower secondary learners sondern auch Erwachsene die von neu anfangen wollen oder die es in der Schule schon gehabt haben und wieder neu lernen wollen, würde diese App helfen. Vielleicht sollte man lower secondary learners mit etwas anderem ersetzen, vielleicht auch was für Erwachsene ähm... betrifft es vielleicht nur ähm... die ihr Sprachenlevel statt lower secondary learners angemessenes level A2 to A2 learners. Weil es ist hier A OK vielleicht A2, A1 to A2. Das kann man als Erwachsener auch machen, muss nicht in der Unterstufe sein also diese Formulierung hat mich ein bisschen gestört, und ansonsten war das eigentlich gut, also da man wirklich Lektionen auswählen kann war es für mich auch eindeutig eine drei, weil die Lektionen immer nacheinander sind, also eine Lektion war immer ein Themen und die sind nacheinander, und die kann man leicht durchmachen nacheinander, also es ist nicht zu schwer und deswegen ist es ähm... kann man die Geschwindigkeit nicht verschnellern, weil es sind oft gleiche Aufgaben und deswegen haben ich da learners control the rate weil es oft sehr langwierig ist, wenn man es schon gelernt hat, aber the content ist presented at an appropriate rate für lower secondary learners, das ist wahr, aber für mich war es eindeutig nur drei, trotzdem war es ein bisschen schwierig weil für mich die App für Erwachsene war, aber A1 to A2 also lower secondary level stimmt schon auch, aber von den Themengebieten her nicht nur für schüler sondern auch für andere Berufsgruppen.

[C4] Dimension C4 Personalisation. Da war, das war bei mir auch nicht so leicht diese Dimension, weil zählt es jetzt wenn man den Avatar verlinken kann mit seinem Facebook Profil wo dann das Foto auftaucht, ist es dann one feature weil automatisch das Bild und der Name personalisiert ist, oder zählt das dann wirklich als zwei features weil das Bild ist ja das von Facebook, aber der Name ist auch von Facebook also sind es eigentlich zwei features. Aber das macht man mit nur einem Klick, also sobald man es mit Facebook verbunden hat, ist es personalisiert und etwas anderes kann man nicht machen. Also ich hab nicht herausgefunden wie es geht selbst ein Bild zu machen und das reinzustellen. Es geht glaube ich wirklich nur über das Facebook Profil, und was ist jetzt ein feature. Ist der Name ein feature oder zählt das schon als feature, viele sind auch nur als guest angegeben, also wahrscheinlich schon. Und wenn ich das nur mit meinem Facebook Account verlinken aber nicht selbst gestalten kann, kein eigenes Bild hochladen kann, zählt das dann auch als personalisation, also als one feature oder als three features weil Bild, Alter ist vielleicht auch noch dabei. Auf jeden Fall war das Wort feature für mich ungenau, vielleicht noch ein paar Beispiele mehr wären gut gewesen, weil Avatars (xxx), weil Avatar, eben dieser Avatar besteht aus vielen features und was alles zählt dazu, oder ob der Avatar jetzt nur ein feature ist, das vielleicht ein bisschen spezifizieren. Und ich glaube das kann man leicht machen mit ein paar Beispielen oder mit images und avatars. Ich weiß nicht ob ein Avatar wenn man die Augen, Frisur und Mund richten kann, sind es dann three features oder ist der ganze Avatar nur ein feature, deshalb war es für mich nicht ganz klar was mit feature gemeint ist, wie groß oder wie klein soll man es fassen.

[C5] Dimension C5 ähm... da war eigentlich der rubric klar. Might be particularly interesting, might be interesting, most of the app's, some of the app's, also da war der rubric eigentlich klar, und trotzdem ist es halt dann schwer zwischen zwei zu entscheiden, zwischen zwei (Punkte) und drei (Punkte) weil es sehr knapp ist, ist es jetzt some oder, also might be particularly interesting oder might be interesting ist halt schwer zu entscheiden, aber der rubric glaube ich nicht dass man es da machen könnte, ich denke es ist schon klar. Aber es passiert halt oft, dass es zwischen zwei hängt, und ja, sonst waren die Fragestellung und die Antworten eigentlich klar.

[C6] Dimension C6 Aesthetics das war auch klar, the app's graphics, interface motivate learners to engage it. Da ist der rubric klar, und graphics und interface ist eigentlich auch klar was gemeint ist, da gibt es keine Kommentare. Das ist bei Mondly eindeutig eine Vier, weil man kann natürlich vor allem was junge Lerner betrifft immer noch mehr machen, aber das meiste ist schon, so wie eben der Chatbot oder augmented reality, das ist schon sehr viel Mühe die man sich macht um Lernende zu motivieren.

[C7] Gamification, das war für mich auch nicht klar was sind game-like elements, was zählt da dazu. Ist so etwas wie eine tägliche Lektion, also jeden tag eine neue Lektion, ist das jetzt schon etwas Spielerisches, weil das gibt es auch schon oft bei Spielen, dass man jeden Tag ein neues level machen kann. Auf jede fall habe ich jetzt, man bekommt points man bekommt ein level und ein ranking und deswegen habe ich jetzt three game-like elements genommen. Aber es ist auch bei dieser app nicht zu schwer, weil ich glaube game-like elements... ja eigentlich fast doch klar, vor allem wenn man sich die Beispiele anschaut, wird es einem doch klar. Ich nehme das zurück was ich vorher gesagt habe. Dadurch das Beispiele dabei stehen, prizes, points, ranking, avatars, also das hilft dann, die Beispiele helfen dann das game-like elements zu definieren, und da war es bei mir dann eine Vier.

[D1] Und Domain D, Quality of Input, dimension D1. Does the app provide comprehensible, valuable and relevant input. Ähm... Da ist der rubric klar mit continuously, mostly, rather, und diese Abstufungen machen es eigentlich meistens ziemlich klar. Da ist es einfach eine kontinuierliche Abstufung von fünf bis eins, und diese Wörter helfen dann wenn man sich entscheiden muss. Also da ist es eindeutig eine Fünf beziehungsweise eine Vier, eben weil wieder relevant to lower secondary learners, und da hätte ich gesagt es ist mehr eine App wo es viel um Berufsfelder geht, eher schon für, auch für Selbstlerner oder Erwachsene und nicht für lower secondary, aber trotzdem A1, A2 level, aber vielleicht nicht lower secondary learners sondern eher Erwachsene, weil es um viele Berufsfelder geht und um reisen. Nicht alles ist eben so wichtig für lower secondary learners was für andere A1, A2 learners wichtig ist.

[D2] Dimension D2 Opportunity to produce output. Ich glaube da ist halt sehr viel, output ist schwer bei so einer App, also produce output rather freely ist eher schwer bei einer App zu machen, aber ist eben dann nur eins bis drei zumindest bei den Apps die ich kenne, trotzdem müsste der rubric eigentlich klar sein, aber ich glaube eben es ist schwer, dass man fünf (Punkte) erreicht. Dass der output limited, predefined answers ist, vielleicht gibt es sowas auch wenn man die persönliche Betreuung bei einer App hat, aber ähm... ja ich kenne eben nur Apps die von eins bis drei gehen, eben nicht freie output Möglichkeiten geben wo man mehr als einen Satz schreibt. Ansonsten ist das eigentlich auch klar der rubric.

[D3] Dimension D3 Appropriateness of Language Learning Material. Eben da waren wir bei diesem lower secondary learners was bei dieser App nicht der Fall ist, deswegen weiß ich nicht, intended learners wäre für mich eher Selbstaufrischung entweder für Reisen oder ein eben Berufsveränderung, wenn man jetzt viel auf einmal im Hotelbetrieb arbeite und deshalb jetzt Phrasen braucht oder im Restaurant arbeitet und deswegen Phrasen braucht die man im Restaurant braucht oder eben in ein englischsprachiges Land reist, dazu ist die App glaube ich sehr gut, und da sind die intended learners eher nicht lower secondary students, could be, aber es ist halt eher für die, vor allem darauf aufbaut dass es um Reisen geht da es sehr joblastig und berufslastig ist, vielleicht auch einen neuen Job wo man Englisch braucht bekommt und das eine Auffrischung dafür ist. Also level of apps hätte ich vier (Punkte) genommen, weil es usual matched weil ich nicht weiß was sind die intended learners, weil es steht nirgends ob die App für Erwachsene ist oder für students. Also es steht in der app jetzt nicht für wen die app gedacht ist, deshalb ist da jetzt vielleicht ein bisschen schwer zu beantworten wer sind die intended learners und wer ist die target audience.

[D4] Dimension D4 Skills. Ähm...da sollte man, könnte man vielleicht noch einmal die four skills aufzählen. Ich glaube wenn man das vielleicht noch in Klammern dazuschreibt, dann weiß jeder welche skills gemeint sind, nur zur Erinnerung, aber sonst wäre es three of the four skills weil eben ja, writing ist nicht sehr, schreiben muss man schon, eben nur auf Wort- oder Satzlevel schreiben, aber sonst eher nicht. Deswegen habe ich nicht alle vier gegeben. Und eben vielleicht noch als Erinnerung die four skills dazugeben.

[D5] Dimension D5 Exposure to the Target Language. Does the app maximize exposure to the target language. Ich habe number four the app mostly uses target language in the learning activities, some translations in the native language are provided gewählt. Also bei der App kann man auswählen ob man die Übersetzung haben will oder nicht, und deshalb habe ich die Nummer Vier genommen. Oh, ich hätte ich Nummer Fünf nehmen sollen. OK, Nummer Fünf. Wenn ich es durchlese, weil option menu ähm... man kann die native language eben einstellen, aber man muss es nicht tun, und der rubric ist da eigentlich klar, dass sie eben ähm mostly und dann some und only, minimal exposure, also da ist der rubric eigentlich klar bei dimension D5. Und sonst war es das glaube ich jetzt, oder?

[Participant 2 finishes the evaluation]

[The interviewer commences the follow-up interview]

I: Yes, danke. Was du zweimal erwähnst hast, war, dass der score eins mit der not applicable option zu ähnlich ist, dass man da oft bei zwei dimensions schwer unterscheiden kann soll man score eins auswählen oder das not applicable. Wie hast du diese not applicable option wahrgenommen? War das für dich OK, das soll ich überhaupt nicht verwenden, oder war es für dich schon so OK, ich merke, ich weiß bei dieser dimension nicht was ich verteilen sollen deswegen nehmen ich not applicable weil es sich anbietet?

P2: Ja, also wenn es gar nicht vorgekommen ist in der einser Spalte, da steht dann es kommt gar nicht vor, dann würde, finde ich, halt not applicable auch passen, weil wenn es gar nicht vorkommt, dass man vielleicht den einser descriptor sagt es kommt ganz selten vor. Not applicable wäre ja dann es kommt gar nicht vor. Vielleicht muss das dann in einem rubric

drin sein, es ist Standard. Ich habe jetzt eben wenn etwas gar nicht vorgekommen ist, ob ich eben dann auswähle es kommt gar nicht vor oder es ist not applicable weil es gar nicht vorkommt, deswegen war mir das unklar.

I: OK, ich verstehe was du meinst. Zwischen dem einser score und dem not applicable gibt es ein paar dimensions. Im Grunde genommen ist das not applicable dafür vorgesehen wenn du zum Beispiel eine gewisse Eigenschaft bei einer App nicht hast weil man die nicht anwenden kann beim Evaluieren. Gut, dann die zweite dimension die 21st Century Skills. Wenn jetzt bei der Frage nur dastehen würde does the app engage learners to engage 21st century skills ohne die Klammer, würdest du wissen um was es sich da handelt?

P2: Da hätte ich halt eher nur an media literacy gedacht, also nicht an das collaborative writing oder collaborate, cooperation und das Ganze, also diese ähm... da hätte ich auch nur an media literacy gedacht und das ist auch das Einzige das bei mir in der app vorkommt, weil kooperativ ist da bei mir nicht viel und eben die 21st Century Skills, dass das mit kooperativen Arbeiten auch zu tun wäre mit jetzt nicht gleich in den Sinn gekommen, sondern eher halt mit media literacy.

I: Ok. Also wäre es dann vielleicht besser, oder wie würdest du es empfinden wenn jetzt zum Beispiel statt diesen zwei 21st century skills, wenn man das irgendwie zusammenkürzt und vielleicht auf das fokussiert was in der Klammer steht was drunter verstanden wird, als Erleichterung?

P2: Ja, also aussuchen was du willst, ob du media literacy willst, obwohl dieses kooperieren dann eh bei den späteren Fragen dabei ist, wie kann man sich mit den anderen unterhalten, aber dieses media literacy ist halt schon auch ein wichtiger Punkt bei Apps, und ich glaube das kommt auch bei meiner App viel vor, und dieses kooperative dann dafür sehr wenig, und deswegen weiß ich jetzt nicht soll ich es stark werten oder wenig werten, also entweder aufteilen und beide einzeln fragen. Oder eins von den zwei spezifizieren, so dass nur eines von den Sachen die in der Klammer stehen gefragt ist.

I: Okay, next one, das war B1, da hast du gesagt du bist dir nicht sicher ob du drei oder vier (Punkte) wählen sollst, weil bei Mondly, bei der App ein Teil von vier passt aber auch ein Teil von drei. Ich glaube, dass ist eine sehr gute Beobachtung, weil bei den descriptors könnte es insofern problematisch werden, weil eben teilweise was vom vierer score zutrifft und dann teilweise auch vom dreier score. Vielleicht muss man da aufpassen, dass da irgendwie, vielleicht kannst du ja irgendwie-

P2: Wobei es teilweise schwierig ist weil die Apps so unterschiedlich aufgebaut sind. Es ist halt so wie bei einem Spiel, du kannst speichern und kannst sagen der Fortschritt geht verloren. Wenn du mitten in einer Lektion bist, und dann geht es halt eher, dass du in Nähe wo du aufgehört hast, weil es steht ja in proximity to where they were when the logged off. Und wenn man aber zum Beispiel eine Lektion fertig hat und dann aussteigt, dann kann man sofort mit der nächsten Lektion anfangen.

I: Mhmh. Okay. Dann zu B3. Da hast du gemeint du würdest keine App kennen die nicht well organised ist, also wo diese ganzen Bauteile nicht gut integriert sind. Würdest du dann sagen, dass die dimension im Grunde genommen ausgelassen werden kann? Weil das wäre dann in weiterer Folge auch ein Frage welche dimensions man vielleicht auslassen könnte, aber wäre B3 dann eine vom den dimensions wo du sagst das könnte man jetzt vielleicht auslassen, weil es eigentlich keine App gibt die nicht well-organised ist. Oder findest du sie vielleicht nicht so wichtig wie andere dimensions?

P2: Ja, also allgemein eine Frage zum Aufbau sollte schon drin sein, aber es sind halt dann auch andere, wie so ein logischer Aufbau ist und sowas, also sind ja dann bei, ich weiß nicht

wo das ist, aber irgendwo steht ja auch, dass nur logical so eben und, ich weiß halt nicht und das mit not enhance, oder detract... Was heißt denn detract?

I: Ähm. Das es die Schüler-

P2: Ablenkt?

I: Ja, genau.

P2: Ja, es gibt wahrscheinlich schon Beispiele, aber in dem Fall ist es eine App die etwas kostet nach sieben Tagen Gratisperiode. Also ich glaube schon, dass es sowas gibt, also ich glaube schon dass es hineingehört wie ist die App organisiert. Mir kommt vor das kam woanders im rubric auch vor. Das ist ja how easily can learners move through content das gehört ja auch dazu, und dann mit the cohesive programme das wäre eher vielleicht was da eher gehört. Eher die B7 hätte ich weggelassen als B3 obwohl es eigentlich ähnlich ist.

I: Okay. Dann kommen wir zum nächsten Punkt. Wenn du jetzt den rubric so vor dir liegen hast, hättest du dann aus dem rubric aus den dimensions heraus erkannt, dass man damit eigentlich Apps evaluieren soll die für lower secondary gedacht sind. Oder nicht?

P2: Ja, aber nur weil es zweimal gestanden ist mit lower secondary learners, aber sonst eigentlich nicht. Vielleicht eher die Frage nach der age group oder dem learning level, jetzt steht es halt mit intended learners, zum Schluss steht jetzt nichts mit lower secondary. Ich finde diese App ist halt eher so wie Babel für Erwachsene gedacht die gerade Zeit haben oder auf Reisen nach Amerika gehen wollen oder irgendwohin in den Urlaub fahren und deshalb wollen sie ihr Englisch ausbessern, so ist es mir vorgekommen.

I: Mhm.

P2: Und deswegen hätte ich dann eher gefragt für welche Schulstufe ist die App, ist es klar für welche Schulstufe die App ist oder kann man die App in jeder Schulstufe anwenden oder so. Sagen wir wenn es wirklich für lower secondary sein soll, dass man wirklich irgendetwas mit Schule hineinschreibst, das man es in der Schule anwenden kann oder außerhalb von der Schule oder als Selbstlernangebot oder das man es wirklich verpflichtend macht. Mir fällt bis auf die zweimal wo es stand eigentlich, es ist eine normale App, eine Sprachlernapp.

I: Im Grunde kannst du mit dem rubric jede Sprachlernapp evaluieren aber es zielt eigentlich schon auf lower secondary ab. Gut, dann kommen jetzt noch ein paar allgemeinere Fragen. Inwiefern würdest du den rubric in deiner Arbeitspraxis anwenden, vor allem im Hinblick auf die Ausführlichkeit und auf die Länge mit den 26 Dimensionen?

P2: Puh... ja, ich bin jetzt allgemein von den rubrics nicht gewohnt, so viel wie man es im Studium macht, mache ich es jetzt auch nicht mehr, dass man so viele rubrics durchliest und damit arbeitet. Wenn man es gewohnt ist geht es sicher schneller weil es eh immer das gleiche Schema ist, aber trotzdem bei manchen ist halt doch nicht das Schema wo ganz links bei fünf oft steht und ganz rechts wenig, sondern dass sind dann schon ein bisschen kompliziertere oder genauere Entscheidungen zu treffen, dann braucht man schon ein bisschen Zeit, also für mich selbst glaube ich nicht dass ich das irgendwie so wenn ich eine App im Unterricht einsetze, dass ich das vorher durchgehe, hätte ich jetzt nicht gemacht. Aber ich arbeite zurzeit eh wenig mit rubrics. Vielleicht wenn man reinkommt oder drin ist oder allgemein öfter machen muss so wie ich das machen sollte eigentlich, dann geht es vielleicht schneller, aber es ist halt dann doch, ich habe mir gedacht die App habe ich gestern ausprobiert und heute noch ein bisschen und jetzt habe ich mich 15 Minuten vorher hingesetzt und fülle den rubric aus. Das ist sich halt nicht ausgegangen, ich bin nicht ganz fertig geworden, weil es schon länger dauert, weil die descriptors sind schon genau. Manchmal ist es einfach wenn es eine kontinuierliche Skala mit rechts viel und links wenig ist, aber es ist halt nicht bei jeder Dimension so, dass man so schnell drüber gehen kann und sagen ja das ist ein Dreier oder ein Vierer. Da muss man dann schon genau überlegen und

die descriptors durchlesen. Deswegen glaube ich, dass ich den rubric in meiner Praxis nicht verwenden würde. Vielleicht aber wenn man ein Feedback oder eine Wertung haben will für eine App im Playstore, kann man sowas gern hineingeben. Wenn man sagt von 26 Punkten hat sie so viele Punkte erreicht.

I: Okay, und wenn man davon ausgeht wir haben den rubric mit den 26 dimensions der sehr aufgebläht ist , ähm... eine Lightversion davon, zum Beispiel du nimmst aus jeder domain aus allen vier domains jeweils drei, maximal vier dimensions heraus und hast dann 12 bis 16 dimensions, wäre dann der Anreiz größer, dass man das wirklich einsetzt? Was hältst du von so einer Lightversion?

P2: Für was wäre die dann gedacht?

I: Der volle rubric ist im Grunde genommen für die genaue Evaluation von Apps vorgesehen. Die Lightversion wäre dann mehr oder weniger das Produkt mit dem LehrerInnen im Alltag wirklich arbeiten können, weil er übersichtlicher ist aber man sich die wichtigsten Dimensionen herausgeholt hat.

P2: Ja, vielleicht führt man sich vor Augen, dass so etwas wie target language klar ist, dass es vorkommen muss, aber ich weiß nicht ob ich jetzt, also wo ich das Ergebnis jetzt vorweisen würde. Nur für mich alleine, ich habe eh, manche Sachen sind mir dann wichtiger, manche domains, und vielleicht ist dann mir persönlich in meinem Unterricht die cultural sensitivity wichtiger als andere Aspekte, und dann merke ich eh ob die App mit gefällt oder nicht. Natürlich um es dann professionell zu bewerten müsste ich dann den rubric verwenden, aber nur für mich, dass ich die App aussuche, dass ich es verwende brauche ich ja jetzt den rubric nicht, nur wenn ich es halt wirklich objektiv machen will, oder weiterempfehlen will so als Bewertungssystem kann ich mir den rubric schon vorstellen. Wenn ich jetzt eine App im Unterricht verwende, werde ich auch keinen rubric light verwenden glaube ich, außer ich muss weiterempfehlen oder ich muss es argumentieren wieso ich die App verwenden.

I: Okay.

P2: Wenn ich vielleicht dem Schulleiter den Kauf einer App vorschlagen muss für alle unsere Schüler dann wäre es vielleicht fein wenn ich sagen kann da hat die App so viele und so viele Punkte, aber so für mich selbst glaube ich nicht dass ich eine Light-Variante verwenden will, weil jedem andere domains wichtiger sind.

I: Mhm. Okay. Danke für deine Zeit und dein Feedback.

Think-aloud protocol 03, 8th February, 2021

I: Im Zuge dieser Think-Aloud-Session wirst du eine language learning App mithilfe des Evaluation Rubric for Language Learning Apps for Secondary Level Learners evaluieren. Der rubric deckt insgesamt 26 Dimensionen ab. Für jede Dimension sollst du der App zwischen einem Punkt und fünf Punkten geben. Kannst du eine Dimension bei einer App überhaupt nicht anwenden, dann wählst du auf der Skala die Option not applicable. Versuche bitte während des Evaluationsprozesses laut zu denken. Sage während du evaluierst bitte alles, das dir hinsichtlich des rubrics durch den Kopf geht. Besonders interessieren mich deine Gedanken hinsichtlich der Formulierungen, der Sinnhaftigkeit und Anwendbarkeit der einzelnen Domänen und Dimensionen. Während du evaluierst und dabei laut denkst, werde ich Notizen machen. Gibt es von deiner Seite aus noch offene Fragen?

P3: Nein.

I: Dann würde ich dich nun bitten mit dem Evaluationsprozess zu beginnen.

[Participant 3 commences with the evaluation]

[A1] Gut, also ich beginne mit der Dimension A1 Thinking Skills und da stellt sich bei mir die Frage, wenn ich die Frage lese, was überhaupt das Level of Webb's Depth of Knowledge ist, weil ich das eigentlich nicht kenne. Wenn das aktiv als erste Frage dort steht, habe ich nicht wirklich eine Ahnung was das ist. Es wird aber das Ganze ein bisschen klarer durch die Antwortgebung, weil in den fünf Dimensionen dann Beispiele gegeben werden, eben beispielsweise, create, synthesise, analyse, design. Also ich glaube, dass das dann wirklich auf diese Skills fokussiert ist und dass ich mich auf diese Skills auch in der Antwortsuche fokussieren muss. Ähm, das heißt, wenn ich diese Rubriken so durchlese... finde ich es ein bisschen irreführend, dass dann verschiedene Skills gegeben sind. Ich habe das zum Beispiel extended thinking skills, dann habe ich strategic thinking skills und da weiß ich halt nicht ob diese Beispiele die dann in der Klammer gegeben sind wirklich ausnahmslos nur DIESE Beispiele sind oder ob es auch noch weitere Beispiele gibt. Also das zum Beispiel auch compare in der fünften Skala vorhanden sein könnte. Das heißt die erste Frage finde ich tatsächlich ein bisschen verwirrend. Wenn ich es aber für diese pronunciation App beantworten müsste, springt mir halt der Punkt zwei ins Auge, wo es um recall und reproduction geht, weil es bei Elsa hauptsächlich darum geht dass Kinder sich, generell Menschen sich sounds merken, sounds reproduzieren und wirklich sich anhören und aktiv reproduzieren. Deshalb würde ich da zwei Punkte hergeben, aber wie gesagt mit Vorbehalt weil ich die Frage nicht hundertprozentig verstehe.

[A2] Dann bezüglich der dimension A2 21st Century Skills finde ich ganz gut erklärt weil da zusätzlich noch ein Link angegeben ist, wo man sich informieren kann. Was aber natürlich dann auch fraglich ist, ob das Sinn macht wenn ich eine App-Skala entwerfe und dann zusätzlich noch einen Link dazu habe wo die Person sich dann erst wieder darüber informieren muss was da eigentlich beim Link dabeisteht. Würde ich vielleicht ein bisschen anders lösen. Und ich finde es ist zusätzlich ja noch in Klammern erklärt. Man hat dann beispielsweise critical thinking, communication, collaboration, creativity usw. Es ist fraglich ob man die Frage da ein bisschen anders formulieren könnte, damit die 21st century skills gleich erklärt werden, dass sie teilweise in Klammer sind. Wenn ich die Antwortmöglichkeiten anschau, ist es wieder notwendig dass ich überhaupt weiß was die 21st century skills sind. Ich würde aber sagen, je nachdem wie ich das verstanden habe, würde ich der App definitiv fünf Punkte geben, da es sehr viel um Kommunikation geht, also um richtige Aussprache, ähm... teilweise auch self-direction weil Kinder selbstständig arbeiten können, media literacy, es geht halt wirklich um die Verwendung einer App im Zuge des Lernens. Deshalb finde ich würde das für mich in diese Kategorie fünf fallen, wieder mit Vorbehalt das eben die 21st century skills vielleicht detaillierter erklärt gehören.

[A3] Bei der dimension A3 Value of Errors, ähm... ist die Frage klar formuliert finde ich. Also mir ist eindeutig klar was gesucht wird. Hier würde ich vier Punkte hergeben, weil die Kinder ähm wirklich mit einer explanation, das heißt auch mit der richtigen Aussprache von dem Wort ähm, provided werden, und die Kinder auch die Möglichkeit, die Lerner die Möglichkeit haben die Frage noch einmal zu beantworten. Hier finde ich eben auch die Kategorien sehr klar gewählt, wird immer weniger, das heißt die fünfte Kategorie ist am ausführlichsten und ja... Ich würde sagen ich würde hier vier Punkte hergeben.

[A4] Bezüglich Feedback to Teachers (A4). How does the app allow teachers to monitor the learners' progress. Die Frage ist wieder klar formuliert. Die Kategorien muss ich jetzt noch einmal kurz durchgehen... . Finde ich auch gut gewählt, da eben beim Ersten durchwegs auf zusätzliche achievements, academic growth eingegangen wird und das dann im Zuge dessen, also Punkt vier, Punkt drei usw. immer weniger wird. Also ich finde das ist eine gut gewählte Skala. Ich würde tatsächlich wieder fünf Punkte hergeben, weil man durchaus nach Feedback fragen kann. Also da gibt es ja diese communities wo man beispielsweise eine Klasse als community createn kann und dann gezielt nach Feedback bezüglich der oral production fragen kann, deshalb würde ich da fünf Punkte hergeben. Es ist wie gesagt eine klare Frage und auch eine klare Antwortstruktur.

[A5] Bezüglich Cooperative Learning (A5)... Klare Formulierung, wieder keine Unklarheiten meinerseits. Ähm, communicate, share learning artifacts and collaborate finde ich sehr gut gelöst in der Fragestellung, weil dann noch die Beispiele in Klammer gegeben sind auf was man sich da bei der Analyse der App fokussieren muss. Und auch die Antwortmöglichkeiten finde ich gut gewählt, weil beispielsweise bei Punkt fünf noch der factor of communicate freely dabei ist, was ich bei dieser App nicht so unbedingt finde, weil man durchaus nur in Gruppen kommunizieren kann. Deshalb würde ich hier auch vier Punkte hergeben. Ich finde von der Struktur her bauen eins bis fünf wieder aufeinander auf. Wie gesagt das freely in der fünften Kategorie kann ich bei dieser App jetzt nicht unbedingt feststellen, weil man auf die Gruppen fokussiert arbeitet.

[A6] Dann bezüglich dimension A6... Die Fragestellung ist wieder klar, finde ich. Hier kann ich drauf ein, also da ist es auch von der Antwortgebung klar, dass bei der fünften Kategorie auf alles eingegangen wird, auf visual, auditory and/or kinesthetic. Und von der Antwortmöglichkeit würde ich da auf die dritte Rubrik zugreifen, weil es... jetzt nicht ausschließlich different ways sind sondern die learners wirklich spezialisierter auf gewisse Aussprache, Aussprachearten und auch Intonationsarten arbeiten, das heißt learn in different ways, wobei hm... würde ich Kategorie drei nehmen. Ich habe mich jetzt gerade verlesen, aber all of the app's content und some of the app's content würde ich sagen. Some of the app's content weil nicht der ganze Inhalt der App auf alle möglichen Lernarten der Schülerinnen und Schüler eingeht.

[B1] Gut, dann Design, das ist dann dimension B1 Ability to Save Progress. Die Fragestellung für die Rubrik ist für mich wieder klar, das heißt kann ich zurück zu den Inhalten gehen die ich bereits gemacht habe, ähm... die Antworten... finde ich alles in allem wieder gut aufgebaut. Das heißt hier würde ich auch... hier würd eich eigentlich auch fünf Punkte hergeben. Nachdem ich wirklich gezielt zurückkehren kann zu was ich gemacht habe und das Ganze auch aufeinander aufbauend ist. Was ich bei der App nicht durchschaut habe ist, ob ich das dann noch einmal machen kann, das heißt da bin ich jetzt wahrscheinlich dann im Graubereich. Es steht aber immer wieder wenn man die App neu öffnet, dass man zu today's experience zurückkehren und dort weitermachen kann wo man aufgehört hat. Deshalb ist es für mich schon die fünfte Kategorie. Mmh... wobei sie das nicht manually selecten können. Dieses automatically return to the content and manually select the content ist vielleicht bei den Kategorien ein bisschen verwirrend. Weil es eigentlich in meinen Augen besser wäre wenn die Schüler und SchülerInnen manually zu dem content zurückkehren könnten und nicht automatically. Weil wenn ich etwas nicht verstanden habe, kann ich manually

zurückgehen. Deswegen würde ich da eventuell sogar die Kategorien ein bisschen verdrehen. Ja das ist eine subjektive Meinung dazu.

[B2] Dann Integration (B3). Wie ist die App mit anderen Apps und so weiter verbunden? Ich finde die Antwortmöglichkeiten klar formuliert, da gibt es nichts auszusetzen. Da muss man eigentlich nur bis vier zählen können und schauen ob alles integriert ist. Hier würde ich der App zwei Punkte geben, weil es nicht wirklich online communities sind, für mich wirklich außenstehende online communities. Da ist die Formulierung wieder etwas schwammig, weil man sagen könnte, dass auch die community die es in der App gibt eine online community ist. Da weiß ich jetzt aber nicht ob das gemeint ist. Ich gehe davon aus, dass es nicht gemeint ist. Deshalb würde ich zwei Punkte hergeben, weil dezidiert kann der learner nur auf seine Email zugreifen.

[B3] Dann dimension B3... klare Frage... finde ich von den Antwortmöglichkeiten gut gewählt, von den Rubriken, weil es Abstufungen gibt. Always organised, usually organised, organised in a way, does not distract, not well organised and confusing. Ich finde es sehr gut strukturiert von der Antwortmöglichkeit, deswegen ist es für mich auch einfach zu beantworten. Ich würde hier fünf Punkte hergeben, da es für mich easy and user-friendly ist.

[B4] Nächste Frage, is the app intuitive. Are learners able to engage with minimal guidance. Hier müsste man für manche wahrscheinlich den Begriff minimal guidance erklären. Was wird unter minimal guidance verstanden, wie weit geht minimal guidance überhaupt, ähm... Das ist tatsächlich eine Problematik bei den Antwortmöglichkeiten, weil wenn ich nicht verstehe was minimal guidance ist oder some guidance, wenn ich das nicht genau definiert habe, fällt es mir als Lehrer vielleicht schwer die Frage zu beantworten. Was für den einen minimal guidance ist, ist für den anderen vielleicht schon viel guidance. Das kommt auch ganz auf die Schüler an. Ich würde deshalb wahrscheinlich den Mittelweg gehen und sagen are able to engage the app with some guidance weil ich der Meinung bin, dass Schüler durchaus wissen müssen was ein spelling alphabet ist, what is the aim of using the app and what are the different categories. Also was können sie überhaupt mit dieser App lernen. Für meine Definition von minimal guidance, vielleicht könnte man das noch ein bisschen klarer formulieren.

[B5] Then Navigation (B5). Klare Formulierung der Frage. Ich finde auch die Antwortmöglichkeit wieder gut aufgebaut, das heißt das Erste wieder fluidly, dann some effort, dann multiple clicks, also es ist auch wieder gut abgestuft von den Antwortmöglichkeiten. Ich finde die Struktur der Frage gut. Ich müsste wieder fünf Punkte hergeben, weil es durchaus sehr einfach ist sich zurechtzufinden. Die App ist klar strukturiert, es gibt vier verschiedene Kategorien mit denen sich Schüler beschäftigen können, es gibt diese daily practice, chronological order also ich fange jetzt nicht bei Tag zehn an sondern an Tag eins. Ich finde es leicht durch die App zu navigieren und auch leicht verständlich.

[B6] Dann Information Presentation (B6)... ich finde gut, dass in der Klammer noch in a logical manner noch einmal erklärt was mit der logical manner gemeint ist, weil es sonst eventuell auch wieder Fragen aufwerfen kann. Dieses Klammersystem finde ich eigentlich ganz gut, da es zusätzlich Input zur Frage gibt. Das könnte man vielleicht auch bei anderen Fragen machen, wie zum Beispiel beim minimal guidance. Ich muss sagen die

Antwortmöglichkeiten sind gut abgestuft, logical manner, mostly logical, somewhat logical, also wieder eine klar strukturierte Rubrik. Ich würde fünf Punkte hergeben, weil das gut funktioniert mit diesem assessment test, das heißt die Schüler werden gut eingestuft und dann aufbauend auf dem Einstufungstest ganz gut durch die App geführt. Auch bei den sounds erkennt man so eine gewisse Tendenz, dass man von easy sounds to difficult ones fortschreitet. Das finde ich gut gemacht.

[B7] Bei Media Integration (B7)... ich finde die Frage ist ähnlich wie wir sie schon zuvor hatten mit der dimension B3 Screen Design. Is it well organised, does it form a cohesive programme sind für mich sehr ähnliche Fragen, because a cohesive programme has to be a well organised programme. Deshalb gebe ich die gleiche Antwort wie bei der anderen Frage. Auch hier würde ich fünf Punkte hergeben. Die App ist gut strukturiert und macht Sinn. Da ist die Frage ob man tatsächlich nicht einen von den zwei dimensions streichen könnte, weil sie für mich, wie gesagt, auf etwas sehr ähnliches abzielen.

[B8] Die dimension B8 Cultural Sensitivity kann ich nicht auf diese App anwenden. Es gibt zwar durchaus Themen die sich auch mit Kultur beschäftigen, aber das ist jetzt nicht die Sinnggebung der App. Es geht halt hauptsächlich um pronunciation und intonation. Da Einzige was man hier vielleicht noch erwähnen muss, ist, dass nicht alle Arten von Englisch abgedeckt werden, was vielleicht ein Problem sein kann wenn der Lehrer jetzt hauptsächlich British English spricht und der Schüler da jetzt American English lernt. Von den Kategorien her... verstehe ich nicht ganz warum bei der vierten kategorie using authentic picture, images and or text dabei ist und bei der Kategorie fünf nicht. ... Weil das für mich eigentlich noch intensiver ist wenn ich authentic pictures, images and/or text habe wenn es um teaching cultures geht. Deshalb würde ich wahrscheinlich diese beiden Kategorien miteinander vertauschen und die Kategorie fünf vielleicht ein bisschen abstufen, weil für mich kategorie vier eigentlich fünf Punkte wert ist. Sonst... ja, finde ich es gut gewählt.

[C1] Dann Engagement, ähm... Does the app allow learners to select the level. Ich finde da ist die Frageformulierung ein bisschen tricky weil sie es nicht direkt selecten können bei dieser App. Ähm, das würde ich vielleicht eher so formulieren ob es die App ermöglicht den Schülern ein Level zuzuordnen, und nicht dass die Schüler ein Level auswählen. Das macht ja keinen Sinn dass ein learner ein Level auswählt. Der learner sollte ja eher durch den Lehrer oder durch einen Test wissen wo er steht. Ich würde sagen... okay, das ist jetzt tricky weil vielleicht habe ich die Frage auch falsch verstanden... Wenn ich die Frage so lese wie sie dasteht hat für mich der learners selbst nicht wirklich control, weil der learner ja eigentlich durch einen assessment test eingestuft werden soll. Vielleicht würde ich deshalb doch drei Punkte hergeben. Wenn ich diesen assessment test nicht mache, ähm, kann ich eigentlich bei Null beginnen. Das heißt in some ways hat der learner control über sein Level. Ich würde tatsächlich aber die Frage umformulieren. Also da würde ich echt sagen wird ein Lerner kategorisiert, hat er die Möglichkeit in eine Level eingeteilt zu werden, weil selbst das Level auszuwählen macht für mich wenig Sinn. Besonders wenn man mit Schülern in der Schule arbeitet. Von den Strukturen her, beispielsweise little control bei der Kategorie 2 no control finde ich wieder gut abgestuft, aber wie gesagt die Frageformulierung würde ich anders wählen.

[C2] Dann Interactivity (C2)... hmm... passt von der Frageformulierung wieder ganz gut. Passt auch von den Abstufungen her ganz gut mit dem frequently, wobei man vielleicht statt often

bei der zweiten Rubrik etwas nehmen könnte das sich jetzt mehr von frequently unterscheidet. Bei der Kategorie fünf könnte man auch sagen the app ALWAYS requires its learners to engage its content through activities, weil frequently und often doch sehr nahe beieinander liegen. Mmh... trotzdem würde ich auch wieder fünf Punkte hergeben weil es immer so ist, dass der Lerner aktiv selbst Sprache reproduzieren muss und auch dazu Feedback bekommt. Deshalb auf jeden Fall fünf Punkte von mir.

[C3] Dann Pace (C3). Ich finde es klar formuliert. Ich finde es auch gut, dass eine zweite Frage dabeisteht. Also das mit the app does not allow learners to control the rate, does the app present the content at an appropriate rate finde ich gut gewählt. Dass man da auch diese extra Möglichkeit dazugibt. Ich würde der App wieder, ich finde die Rubriken wieder gut eingeteilt mit completely control, mostly control, die Abstufungen sind da klar gemacht. Man hat da keine Schwierigkeiten das einzuordnen. Die App bekommt von mir fünf Punkte und zwar hauptsächlich dadurch weil man sich diese daily practice Zeit selbst einteilen kann. Ich habe das auch auf meinem Handy eingestellt, man kann sich erinnern lassen wie man lernt, wie lange man lernt. Das finde ich ganz gut. Die App erinnert einen auch immer wieder selbstständig daran dass man lernen muss. Es ist ein guter pace würde ich sagen.

[C4] Dann Personalisation (C4). Can learners personalise the app by setting individual preferences. Ich finde es gut, dass wieder in Klammer dabeisteht was man mit den individual preferences meint. Die Kategorien sind ganz klar formuliert. Wobei ich sagen muss, es sind in der Fragestellung drei Kategorien in Klammer und bei der Antwortmöglichkeit fünf steht dann can personalise four or more features. Das könnte vielleicht für manche die Beantwortung der Frage erschweren, weil ich ja nur drei Auswahlmöglichkeiten für die individual preferences in der Klammer habe. Vielleicht würde ich da sogar ein bisschen mehr in die Klammer hineingeben, damit man besser versteht was mit individual preferences gemeint ist. So weiß ich nämlich jetzt nur drei und die anderen muss ich mir selbst denken. Ich würde sagen was in Klammer ist können sie ja kontrollieren. Sie können auf die background music eingehen, sie können den Avatar verändern, was sie nicht kontrollieren können sind beispielsweise die images die da im Hintergrund ablaufen, sound effects können sie aber einstellen. Ich würde sagen sie können, es ist schwierig einzuteilen. Ich würde sagen sie können, ihren user name können sie eigentlich auch umstellen. Ich würde sagen sie können drei features auf jeden Fall ändern, deshalb würde ich vier Punkte hergeben. Wie gesagt mit Vorbehalt, da ich die Klammern noch ein bisschen genauer ausführen würde nach individual preferences.

[C5] Dann Interest (C5). Will the app's language learning material likely appeal to lower secondary learners? Klare Frage. Man könnte sich natürlich jetzt fragen warum nur nach lower secondary learners gefragt wird, und ob man die Rubrik nicht auch auf ältere SchülerInnen anwenden könnte. Hmm... ich finde das aber tatsächlich eine spannende Frage, weil mir als Lehrerin nicht eingehen würde warum ich mich jetzt nur auf lower secondary learners spezialisieren soll. Man müsste das vielleicht eher so formulieren, da fehlt mir tatsächlich dann die Gegenfrage ob es nicht eher für upper secondary learners auch passend wäre. Vielleicht könnte man die Frage dahingehend ändern. Aber tatsächlich muss ich jetzt sagen, da ich auch an einer lower secondary school unterrichte, passt das von learning materials ganz gut, weil es eine wide range of topics abdeckt. Es werden beispielsweise Themen wie food oder cultures oder cities über die pronunciation activities sehr gut abgedeckt. Ich würde definitiv fünf Punkte hergeben. Ja, von den Abstufungen her

passt die Formulierung sonst ganz gut. Bei mir entsteht das Problem eher bei der Frage warum particularly to lower secondary learners.

[C6] Aesthetics (C6). Will the app's graphics and interface likely motivate learners to engage it? Hier habe ich mit der Kategorie drei und vier eine Problem, weil may motivate learners to engage it und may possibly motivate learners to engage it ist doch sehr ähnlich formuliert. Ich würde die Kategorie fünf ein bisschen aufwerten und würde sagen the app's graphics and interface motivate learners, most likely motivate ist für mich dann schon eher Stufe vier. Ich würde trotzdem sagen für mich ist es Kategorie vier, weil es, durch diese Planetenstrukturen, durchaus spannend sein kann für Schüler, aber auch nicht unbedingt, weil mir fehlen die Sachen für Jugendliche, also mehr verschiedene Bilder, Emojis oder solche Dinge könnte man vielleicht noch in der App einbauen. Und man könnte noch ein bisschen genauer spezifizieren was unter graphics verstanden wird. Sind das Tabellen, sind das Zeichnungen sind das bestimmte Bilder.

[C7] Dann Gamification (C7). Die Frage ist gut gestellt. Die Ausführung in der Klammer ist gut, es ist eine klare Formulierung. Die Antwortfindung ist für mich persönlich wieder ein bisschen schwierig, weil ich mir mit dem four or more, three, two, one, does not use any schwer tue. Ich finde das könnte man durchaus anders formulieren. Zum Beispiel mainly uses game-like elements, sometimes uses them, rarely uses them. Ich würde generell, glaube ich, oft bei dem rubric weggehen von diesen wirklich klaren Zahlen und eher auf so Strukturen wie particularly interesting, might be interesting usw. zurückgreifen. Manchmal ist es mit den Zahlen ein bisschen verwirrend. Ich würde da, wenn ich jetzt nach den Zahlen gehe, die Rubrik vier wählen, weil Punkte vergeben werden können, man hat auch so ein community ranking, man kann zwar keine avatars wählen, aber trotzdem fühlt sich der Lerner wahrscheinlich engaged. Wie bereits gesagt, würde ich bei der Formulierung von den Zahlen weggehen und auf Adjektive oder Adverbien zurückgreifen.

[D1] Dann Quality of Input (D1). Was ist valuable input? And relevant input? Das kann jetzt wieder alles sein. Comprehensible ist klar. Bei valuable und relevant hat wahrscheinlich jeder eine andere Definition, das kommt auch auf die Schulstufe an. Auf die App bezogen ist die Frage dann tatsächlich aber doch ganz gut beantwortbar. Auch die Formulierungen der Rubriken sind eigentlich ganz gut. Für mich liegt oftmals das Problem bei der Frageformulierung nicht wirklich bei den Rubriken. Ich finde die App fördert durchwegs comprehensible and valuable input. Hier würde ich also fünf Punkte hergeben. Auch mit den Formulierungen der Rubriken bin ich wieder einverstanden. Für mich gehört in der Frage aber geklärt what is comprehensible, valuable and relevant and what level.

[D2] Dann Opportunity to Produce Output (D2), passt von der Frage her... mhm... die Antwortstrukturen finde ich gut aufgeteilt, man muss sich das halt wirklich genau durchdenken wenn man sich die App anschaut. Ich würde sagen, dass es tatsächlich predefined written or spoken output ist und nicht freely, deswegen finde ich besonders hier die Kategorien der Rubrik sehr gut gewählt, weil es von dem freien Sprechen auf das Satz- und dann auf das Wortlevel geht. Ich würde sagen im Fall der App ist es auf das Satzlevel fokussiert, weil es doch auch conversation tasks gibt wo man ganze Sätze reproduzieren muss, deswegen würd ich drei Punkte hergeben. Diese Rubriken finde ich sehr gut.

[D3] Appropriateness of Language Learnign Material (D3). Is the app's language learning material (xxx) for target audience... ist mit den Abstufungen gut formuliert. Die Antworten sind klar, die Frage ist für mich auch klar. Ich würde sagen es ist always matched appropriately to the target audience, eben weil man am Anfang den assessment test machen kann. Rubriken und die Frage sind für mich klar, da gibt es nichts auszusetzen.

[D4] How many skills does the app train? (D4) Da würde ich für jeden Lehrerlaien hinzufügen welche skills man damit meint. Also wir wissen, dass damit listening, reading, speaking, vocabulary gemeint sind, und die würde ich aber wieder in Klammern in der Frage dazuschreiben. Und natürlich auch, na wobei da könnte man eventuell sogar diese four skills lassen, weil es ja im learning aktiven Englisch auch die four skills sind. Da würde ich der App vier Punkte geben, weil es keinen Fokus auf Grammar gibt sondern man muss zuhören, man muss auch lesen können, man muss reden können, wobei schreiben muss man eigentlich auch nicht können. Hmm... ja ich würde trotzdem vier Punkte hergeben und die skills noch näher definieren, sonst passt das eigentlich von den Antwortmöglichkeiten.

[D5] Does the app maximise exposure to the target language? (D5) Gute Rubriken, es gibt wieder diese Abstufung mit no translations, some translations, translations in the native language to the same extent. Die Rubriken finde ich sehr gut gewählt. Für die App kann ich fünf Punkte hergeben, weil eben rein muttersprachlicher Fokus vorhanden ist und der ganze Input und Output in der Muttersprache zu erfolgen hat. Genau, also Fazit, wie gesagt, weniger Probleme mit der Formulierung der Rubriken, nur bei Rubriken die tatsächlich auf four, three, two, one usw. aufbauend sind, das heißt, dass ich da wirklich Zahlen angeben muss, finde ich teilweise ein bisschen verwirren. Es gibt eher manchmal Probleme mit der Formulierung der Fragen. Ich finde ein paar Dinge gehören hier noch mehr spezifiziert.

[Participant 3 finishes the evaluation]

[The interviewer commences the follow-up interview]

I: Mhm. Danke für deine Evaluation. Ich habe mir jetzt einige Notizen gemacht und hätte im Hinblick auf zwei Dimensionen noch ein paar Fragen.

P3: Ja.

I: Und zwar ist das D1, das war comprehensible, valuable and relevant input. Ich verstehe was du damit meinst, dass valuable and relevant sehr subjektiv zu definieren sind. Würde es in der Fragestellung klarer werden wenn man Ende hinzufügt, dass es valuable and relevant for lower secondary level learners sein sollte.

P3: Lower secondary level?

I: Nicht unbedingt nur lower secondary, sondern wenn man es allgemein haltet für secondary level learners.

P3: Ich finde dadurch würde es definitiv ein bisschen klarer werden.

I: Das ist dann schon der nächste Punkt auf den du gerade eingegangen bist, der kam auch bei den vorherigen Think-aloud sessions zur Sprache. Das Thema lower secondary learners. Du hast das eben auch angemerkt, das zeigt mir, dass auch du findest, dass man den rubric auch für die upper secondary anwenden könnte.

P3: Mhm, auf jeden Fall.

I: Die zweite dimension auf die ich zurückkommen möchte ist D4, also die Skills. Da hast du gesagt du würdest drei oder vier Punkte hergeben. Würdest du jetzt das was du in der App

Elsa Speak wirklich machst, also die pronunciation practice als speaking skill so hinnehmen oder wäre das nur ein Teilbereich des speaking skill?

P3: Ich würde sagen es wäre ein Teilbereich des speaking skill. Aber was bei der App auch gemacht wird, ist, dass wirklich gezielt auf diese ähm speaking exams hingearbeitet wird, und da auch teilweise Phrasen, ganze Sätze eingelernt werden wie man sein Auftreten mit speaking verbessern kann. Deshalb finde ich durchaus, dass es da ein bisschen weitergeht, dass es nicht nur um pronunciation geht sondern wirklich auch um vocabulary training und Erweiterung des Wortschatzes im Sprechen und das halt wirklich auch aktiv in den conversation tasks anzuwenden.

I: Okay, und deswegen würdest du auch sagen es wird auch listening durch die App trainiert?

P3: Das würde ich auch sagen. Also das ist ein bisschen ein zweischneidiges Schwert, weil im Endeffekt kann ich die Frage auch lesen und muss sie nicht unbedingt anhören. Aber listening wird für mich auch trainiert weil ich gezielt auf die sounds und die intonation hören muss, und das ist für mich schon auch ein großer Teil von listening. Das ich das was ich höre dann auch wirklich reproduzieren kann. Da geht es nicht nur um das Verständnis sondern wirklich auch um das pronunciation level.

I: Okay. Ich hätte noch zwei, drei allgemeine übergreifende Fragen. Wen du an das Ausmaß und die Anzahl der dimensions denkst, es sind insgesamt 26, inwiefern würdest du den rubric so wie er vor dir liegt in deiner täglichen Praxis als Lehrerin anwenden?

P3: Niemals. [Interviewer and participant 3 laugh] Ich muss ehrlich sagen ich würde mich nicht hinsetzen und eine App im Hinblick auf 26 Dimensionen analysieren. Ich finde es gibt manche dimensions die eindeutig relevanter sind, zum Beispiel eben wie sieht es aus mit Feedback, welches learning feedback bekommt der Schüler. Oder wie sieht es aus mit den individual differences, auf welche Art und Weise kann der learner lernen. Manche Kategorien fand ich wichtiger als andere. Ich persönlich fange mit Kategorien wie zum Beispiel engage 21st century skills nicht viel an, weil na klar setzt eine App auf 21st century skills, da ich ja schon alleine mit dem Handy arbeite. Manche Kategorien könnte man fusionieren, beziehungsweise halt... manche könnte man tatsächlich auch einfach streichen für den täglichen Gebrauch, also wenn ich jetzt wirklich für meine learner eine learning app verwenden will.

I: Mhm. Okay. Gehen wir hypothetisch davon aus bei diesem rubric werden gewisse Dimensionen vielleicht zusammengelegt oder gestrichen damit er ein bisschen übersichtlicher und kürzer wird. Denkst du, dass dann den Lehrern oder Lehrerinnen die den rubric verwenden sollen die Möglichkeit genommen wird die dimensions auszuwählen die für sie interessant sind? Aus deiner Antwort habe ich rausgehört du persönlich würdest aus dem rubric wenn er vor dir liegt die dimensions rausnehmen die dir wichtig sind, die du anschauen möchtest und den Rest lasse ich weg. Wenn man den rubric aber in einen übersichtlicheren rubric komprimiert, dann wäre die Problematik, dass Leute die sich gewisse dimensions rauspicken wollen die Auswahl nicht mehr haben. Lebt der rubric von der Möglichkeit Dimensionen auswählen zu können?

P3: Ich finde es kommt auch darauf an was man vor sich liegen hat. Wenn ich beispielsweise nur die dimensions aufgelistet hab und dann im nächsten Schritt, ich gehe aus vom Internet, auf die dimensions klicken kann. Wenn ich da zum Beispiel lese Value of Errors und ich denke mir ja, das interessiert mich, dann schaue ich welche Frage dabeisteht. 21st century skills würden mich nicht so interessieren, da würde ich eben nicht draufklicken. Vielleicht wäre es sinnvoll das Ganze nicht in der Form eines Rasters vor einem liegen zu haben, sondern tatsächlich nur die dimensions als bullet points aufgelistet, und wenn mich

spezifische dimensions interessieren, dass ich dann draufklicken kann. In dem Fall könnte man die ganzen Kategorien dann lassen. In der Rasterform auf den zehn Zetteln verliert man sich aber ein bisschen.

I: Alright.

P3: Also hands-on practice würde das wahrscheinlich kein Lehrer in der Schule machen, dass er Apps so analysiert. Also vielleicht eher Stichwörter, mit der Möglichkeit, dass man näher darauf eingehen und nachlesen kann, fände ich schon gut.

I: Okay. Danke für deine Zeit und dein Feedback.

Think-aloud protocol 04, 10th February, 2021

I: Im Zuge dieser Think-Aloud-Session wirst du eine language learning App mithilfe des Evaluation Rubric for Language Learning Apps for Secondary Level Learners evaluieren. Der rubric deckt insgesamt 26 Dimensionen ab. Für jede Dimension sollst du der App zwischen einem Punkt und fünf Punkten geben. Kannst du eine Dimension bei einer App überhaupt nicht anwenden, dann wählst du auf der Skala die Option not applicable. Versuche bitte während des Evaluationsprozesses laut zu denken. Sage während du evaluierst bitte alles, das dir hinsichtlich des rubrics durch den Kopf geht. Besonders interessieren mich deine Gedanken hinsichtlich der Formulierungen, der Sinnhaftigkeit und Anwendbarkeit der einzelnen Domänen und Dimensionen. Während du evaluierst und dabei laut denkst, werde ich Notizen machen. Gibt es von deiner Seite aus noch offene Fragen?

P4: Nein. Doch [laughs], Entschuldigung. Also ich kann schon über die App selbst auch sprechen?

I: Ja klar, sicher.

P4: Okay.

I: Das ist in Ordnung wenn du Anmerkungen zur App hast.

P4: Mhm.

I: Wenn es keine weiteren Fragen gibt, würde ich dich nun bitten mit dem Evaluationsprozess zu beginnen.

[Participant 4 commences with the evaluation]

[A1] OK, passt. Ähm, domain A Educational Value. Level of Webb's Depth of Knowledge (A1) habe ich noch nie vorher gehört. Ich habe keine Ahnung was das ist. Ähm, aber mit den Beschreibungen von der Skala habe ich das Gefühl es geht dabei darum was die Denkprozesse oder Prozesse des Denkens, zum Beispiel nach Bloom, sind. Also ob es nur recall, memorise ist oder ob es higher thinking skills sind. So wie zum Beispiel create, das wäre das Höchste bei Bloom. Bei Elsa war das meiste im Bereich recall, match und auch compare wenn man speech samples miteinander vergleichen muss. Ich hätte da nicht gewusst ob ich zwei (Punkte) oder drei (Punkte) ankreuzen soll, weil die App irgendwie dazwischen liegt, weil das mit compare bei zwei fehlt. Ich fand das war irgendwie schwierig.

[A2] Ich finde auch bei den 21st Century Skills (A2) habe ich zuerst an die Digitale Grundbildung (referring to the subject taught in the MS) denken müssen, an den Lehrplan mit dem critical thinking, vielleicht wegen den Medien und media literacy. Ich habe da eigentlich auch keine Ahnung was die 21st century skills aussagen, und deswegen kann ich dazu eigentlich nichts sagen. Media literacy ist für mich klar, wenn man zum Beispiel eine

App verwendet muss man eine gewisse media literacy besitzen, weil sonst könnte man die App gar nicht herunterladen und verwenden. Die anderen skills, creativity, creativity steht bei Elsa auch nicht im Vordergrund. Also die ersten zwei dimensions fand ich gleich total schwierig.

[A3] Value of Errors (A3) finde ich sehr straightforward. Also was passiert wenn die Frage inkorrekt beantwortet wird. Ich finde in der App, in der regular und der advanced version, dass man dann immer gleich sieht wo der Fehler ist. Das Problem ist aber oft, dass eine Erklärung fehlt warum das ein Fehler ist, aber man kann es dann ja noch einmal anhören. Man kann dann Erklärvideos schauen, die vielleicht auch helfen.

[A4] Feedback to Teachers (A4). Ich habe sehr lange gesucht ob man das in der App machen kann, dass der Lehrer in der App die Schüler registriert und dann sozusagen eine Klasse einrichtet damit er Feedback erhält. In meinem ersten Durchlauf habe ich diese Möglichkeit nicht gefunden. Ich glaube aber schon, dass es mit dieser community Funktion geht, dass man vielleicht eine eigene community erstellt. Ich habe das aber in der App nicht straightforward gefunden. Ich glaube es ist eher so, dass die learners selbst das Feedback erhalten und der teacher nicht wirklich involviert ist.

[A5] Cooperative Learning (A5). Das ist das mit allow learners to communicate, receive messages and post message boards. Also dass sie wirklich direct messages senden können, wäre mit nicht aufgefallen, aber mit den message boards und share learning artifacts, weil man kann die aufgenommenen Antworten mit anderen Teilnehmern sharen, ähm... Aber der rubric ist eigentlich, also diese scale ist ganz straightforward formuliert glaube ich.

[A6] Accomodation of Individual Differences (A6). In uniquely different ways... ich habe dann darüber nachgedacht was diese Lerntypen sind, audio und visuell aber was genau ist nochmal kinesthetic. Ich bin zu dem Schluss gekommen, dass es doch straightforward ist, wenn man weiß was kinesthetic bedeutet. Ich musste da nochmal nachschauen. Es ist aber eigentlich dann ganz easy anzukreuzen bei dieser App. Es ist auf alle Fälle visual und auditory beinhaltet.

[B1] Naja, Design (B), Ability to Save Progress (B1). Ich musste vorher nochmal die App öffnen, weil ich nicht weiß ob es automatisch gespeichert ist wenn ich rausgehe, weil ich keinen Account erstellt habe. Also ich habe mich nicht angemeldet. Ich habe das Gefühl es kam jetzt etwas ganz Neues, deswegen ist der rubric für mich hier schwierig zu beantworten. Also spontan kann ich das gerade nicht bewerten. Mmh... ich könnte jetzt nicht sagen ob das, da müsste ich mich glaube noch ein bisschen länger mit der App beschäftigen um herauszufinden ob das wirklich wieder dort einsteigt wo ich aufgehört habe. Ich hatte das Gefühl es kam jetzt etwas ganz Neues mit neuen Wörtern die ich nachsprechen musste.

[B2] Integration (B2). How it connects to other apps, online communities, independent websites and learners' email. Da habe ich keine Ahnung. Also die App schlägt mir immer vor, dass ich mir die Pro-Version herunterladen soll. Independent websites... also ich weiß nicht worin diese Videos eingebettet sind, ob das eigene Websites sind oder ob das innerhalb der App eingebettet ist. Ich finde das sehr schwierig. Online communities, innerhalb dieser App gibt es eine online community wo man sich mit seiner Email Adresse anmelden muss. Aber

von den scales... achso ob es das integriert... naja, es integriert das schon, aber wie stark kann ich jetzt nicht sagen. Aber ja, die App integriert das auf alle Fälle.

[B3] Is the app's text, graphics, sounds (xxx) well organised? (B3) Ähm... puh... also ja, es gibt halt immer so, also as mich sehr stört sind die Pop-up Werbungen für die Pro-Version. Nach jedem Level erscheinen die Pop-Up Werbungen was ich mühsam finde. Grundsätzlich gibt es immer introductory exercises und dann gibt es auch noch so Videos die man sich anschauen kann, also Erklärungen. Und man kann sich die Wörter immer vorher anhören und dann nachsprechen, das ist schon gut organisiert, finde ich. Mich nervt es, dass in der App immer wieder Werbung für die Pro-Version der App aufscheint.

[B4] Ich finde die App sehr intuitiv. Man kann einfach anfangen und man muss einfach nur klicken und kommt dann im Endeffekt gleich zur ersten Übung. Was ich mir denke ist, das kommt dann glaube ich später noch einmal, dass man von seinem language level gleich am Anfang Daten angeben muss. Das ist ich, glaube ich, schwierig. Wenn das für lower secondary learners ausgelegt ist, braucht man guidance vom teacher, damit man dort überhaupt einmal hinkommt. Ich glaube ich habe das gerade vorher aus der Sicht des Lehrers, also eines advanced learners, gesehen. Vielleicht ist das für upper secondary learners leichter.

[B5] How easily can learners move through the app's content an options? (B5). Also das ist ja eigentlich verbunden mit dem Screen Design (B3), da habe ich jetzt ein bisschen vorgegriffen. Das ist eigentlich das, dass man wegeklicken muss, stimmt aha. Ich finde, dass man viel wegeklicken muss, aber sonst ist es sehr gut aufgebaut mit basic level und next level und man muss so durchscrollen. Wenn man zurückgeht, habe ich das Gefühl man fliegt bei einem anderen level raus. Mit den Levels habe ich dann beim durchscrollen irgendwie Probleme gehabt. Aber ja, das war nicht so tragisch. Ich habe dann einfach in das Level noch einmal hineingeschaut und bin dann nochmal zurückgegangen.

[B6] Is the app's content presented in a logical manner. Ich fand es sehr schräg mit welchen Wörtern und sounds die App beginnt. Ich fand nicht, dass das die Wichtigsten sind. Gleich am Anfang wurde /dʒ/ und /j/ geübt, das fand ich eher schräg. Ich hätte erwartet, weil es die Schwierigsten oder die Wichtigsten sind, dass man mit den Vokalen anfängt. So weit bin ich aber noch nicht einmal gekommen obwohl ich gestern einige Zeit investiert habe. Für mich war es nicht logisch nachvollziehbar, aber ich habe mit auch nicht so wirklich damit beschäftigt wie man am besten pronunciation lernt und übt.

[B7] Media Integration (B7) (xxx) mmh... ja, auf alle Fälle. Die Videos die dann die activties unterstützen, das sind ja noch ausführlichere Erklärungen. Ähm... ich würde schon sagen, dass das einfach genau dazu passt. Beim ersten scale ist ein Tippfehler drin. It does not form a cohesive programme. Beim cohesive fehlt das e am Ende. Das ist mir gerade aufgefallen. Aber irgendwie habe ich das Gefühl das die Frage schon einmal vorkam. Vielleicht habe ich aber die andere Frage auch nicht genau gelesen. Ich habe das Gefühl Media integration und Screen Design sind miteinander verwandt.

[B8] Cultural Sensitivity (B8). Ich habe nicht das Gefühl, dass diverse populations vertreten sind. Gleich am Anfang muss man auswählen, also man wählt gleich am Anfang den American Accent, wobei American sehr breit ist, General American wäre da vielleicht noch

besser. Also ich finde cultural responsive teaching kommt überhaupt nicht vor. Allerdings ist es auch nicht so, dass der typical American stereotype präsentiert wird. Ich glaube das spielt eigentlich in der App keine Rolle.

[C1] Engagement (C), Learner Control (C1). Ja, man kann auswählen beginner oder basic und dann advanced level. Ob man das age level auswählen kann, weiß ich nicht mehr, aber das difficulty level kann man sicher aussuchen. Also man kann aussuchen wo man anfangen möchte.

[C2] Interactivity (C2) (xxx) content to learners... na sicher, man muss interacten. Natürlich kann man sich das auch nur anhören, aber das ist überhaupt nicht lustig wenn man bei den activities nicht mitmacht. Ich finde es genial, ich wusste nicht, dass die speech recognition schon so weit entwickelt ist, dass sie anzeigen kann zu wie viel Prozent man native speaker ist. Man muss auf alle Fälle sehr viel interacten.

[C3] Mmh... does the app allow learners to... ähm... present the content at an appropriate rate für lower secondary learners (C3 Pace). Äh, ich glaube die zweite Frage does the app present the content at an appropriate rate for lower secondary learners kann ich nicht beantworten. Das hätte ich mir ganz am Anfang nochmal genauer anschauen müssen. Ich habe nämlich die App für mich getestet und nicht für meine Lerner. Das war am Anfang vielleicht ein bisschen falsch ausgewählt, weil ich dann higher level ausgewählt habe. If the app does not allow learners to control (xxx)... mhm. Ja, schon, natürlich kann man das kontrollieren und regulieren. Man kann ja auch manche Sachen auslassen wenn man das Gefühl hat man kann das schon. Wobei in der basic version kann man da dann nicht weitergehen, das müsste ich mir auch noch einmal genauer anschauen.

[C4] Personalise (C4). Das ist mir überhaupt nicht aufgefallen. Ich glaube nicht, dass es geht, dass man individual preferences... ich glaube da müsste ich danach suchen. Ganz straightforward hätte ich das nicht gesehen.

[C5] Interest (C5). Will the app's language learning material... ich glaube Schüler würden das schon lustig finden. Es ist schon interessant, aber ich habe das Gefühl, dass manche Sachen schon sehr auf Business abzielen. Also wenn man zum Beispiel als Erwachsener Englisch lernen will und seinen Accent verbessern möchte in business situations. Es kommen auch Wörter vor, gerade vorhin habe ich so business words gehabt wie communication und co-workers. Das ist glaube ich nicht so relevant. Aber es gibt auch so topics wie food und family und at a party, was dann natürlich schon wieder interessant ist, wobei nicht für beginners, Erste (5. Schulstufe), Zweite (6. Schulstufe), sondern eher so Dritte (7. Schulstufe), Vierte (8. Schulstufe).

[C6] Will the app's graphics (xxx) (C6 Aesthetics), ja ich finde schon. Ich finde das leicht zu beantworten, dass es motivierend ist mit den Smileys zum Beispiel. Und dann werden Luftschlagen und Konfetti eingeblendet, das finde ich schon sehr motivierend. Grundsätzlich finde ich das Design schon sehr motivierend, die Sterne die aufscheinen, wie viel Prozent man erreicht... Also das mit den Smileys ist schon ganz lustig.

[C7] Achso, does the app use game-like elements (C7 Gamification). Dass ist eigentlich das jetzt. Achso, graphics and interface (referring back to C6 Aesthetics), und das andere sind

jetzt die game-like elements. Mmh... ja, also graphics and interface war aber auch so mit den levels. Die game-like elements sind diese Sternchen und die se Smileys. Das boostet auf alle Fälle das learner engagement, das ist motivierend.

[D1] Und Language Learning (D). Ah, da kommt das mit topics und valuable and relevant language auch nochmal vor (D1 Quality of Input). Ich glaube ich müsste mir die App nochmal genau anschauen, aber grundsätzlich glaube ich, dass diese scale gut funktioniert. Ah, da kommt auch das mit people working in business vor. Die Frage ist natürlich wie misst man relevant für diese topics die dann im Themenpool vorkommen. Also die Themen family, friends oder welche phrases... ich hatte dann auch etwas zum Thema college parties and Greek societies und beer keg, und dann habe ich mir gedacht für das lower secondary level wäre das vielleicht ein bisschen problematisch. Ähm...

[D2] Does the app provide learners with the opportunity to produce written or spoken output (D2). Auf alle Fälle spoken output. Written output glaube ich gar nicht. Also entweder written oder spoken, ja OK, also written or spoken, auf alle Fälle spoken, beides nicht. The output goes beyond, sentence level. Also bisher war es glaube ich nur am sentence level oder zwei sentences. Aber so richtige texts waren, glaube ich, noch nicht nachzusprechen. Also es trifft das word level oder sentence level zu.

[D3] Appropriateness of Language Learning Material (D3)... was ist die app's language learning material appropriate, ich verstehe das nicht. Is the app's language learning material for specific grade-levels, ich verstehe die Frage nicht [participant 4 begins to laugh]. Is the app's language learnign material for specific grade-levels (xxx)..., aber die target audience ist ja specific grade-level, oder? Ich verstehe das nicht. Ich lese mal die Beschreibungen durch. Inappropriate for language learning purposes, achso, aber inappropriate ist das auch nicht es ist nur das topic inappropriate gewesen beim beer keg. Aber es ist auch inappropriate for business contexts. Ich habe keine Ahnung wonach bei dieser Kategorie gefragt wird. Die Dimension ist für mich ein bisschen schräg. OK.

[D4] How many skills does the app train? (D4 Skills) Da war ich mir zuerst nicht sicher ob die four skills listening, reading, writing und speaking gemeint sind, und ob grammar und vocabulary dabei ist, aber das ist dann meistens bei der Beschreibung extra genannt. ONLY provides vocabulary or grammar practice, none of the four skills are trained explicitly. Mhm. Also ich finde das ist very straightforward. Wobei ich würde bei dieser App sagen, dass es wirklich speaking und pronunciation ist. Vocabulary und grammar, vocabulary... stimmt denn es gibt eine Wörterbuch option wo man deutsche Wörter eingeben kann und es dann auf Englisch übersetzt wird, aber grammar kommt, glaube ich, nicht vor.

[D5] Exposure to Target Language (D5). Auf alle Fälle maximise exposure. Immer, denn es ist fast alles immer English only. Man wird schon am Anfang nach der native language gefragt, und in diesem Wörterbuch kann man nachschauen, aber das ist eher so als learnign tool in welchem ich als learner ein deutsche Wort das ich unbedingt wissen möchte nachschauen kann und das dann auf Englisch übersetze. Also ja, doch maximise exposure to the target language. Das ist auch eine straightforward scale. Jap.

[Participant 4 finishes the evaluation]

[The interviewer commences the follow-up interview]

I: OK, sehr gut. Danke schon einmal für deine Evaluation.

P4: Gerne.

I: Ich habe mir ein paar Notizen gemacht und ich würde gerne noch einmal auf zwei, drei dimensions zurückkommen.

P4: Ja.

I: Ich glaube B7 hat dich verwirrt. B7 ist die dimension Media Integration und du hast gesagt du hattest das Gefühl das vorher beim Screen Design schon einmal gelesen zu haben. Würdest du sagen die zwei dimensions sind so ähnlich, dass man aus den zwei einfach eine machen könnte.

P4: Ja, wobei ich mir nicht sicher bin ob ich das nicht vielleicht zu ungenau gelesen habe. Aber das eine ist the app's texts, graphics, videos, sound and speech ist das well organised, OK, ob es well-organised ist, da geht es mehr um das. Bei der anderen geht es um integration, aber für mich ist das sehr ähnlich. Wenn es gut integriert ist, ist es gut organisiert.

I: Mhm.

P4: also wahrscheinlich könnte man diese zwei dimensions zusammenlegen. Also für mich ist integrieren und well-organised ähnlich.

I: Die zweite dimension ist C3 Pace.

P4: Pace, mhm.

I: Ich habe das Stichwort lower secondary notiert. Das ist jetzt eine ganz allgemeine Frage. Wenn du den rubric auf eine App anwendest, würdest du sagen, dass man eine App auch im Hinblick auf die Möglichkeit sie für die upper secondary einzusetzen evaluieren kann?

P4: Auf alle Fälle.

I: In den vorherigen Sessions wurde nämlich auch erwähnt, dass es wenig Sinn hat sich immer nur auf die lower secondary zu beziehen, weil man mit dem rubric genauso eine App für die upper secondary evaluieren könntest.

P4: Ja, voll. Ich finde die ersten descriptor sind ja auch alle ohne lower secondary gewesen. Ich glaube D3 ist der erste wo das im rubric vorkommt. Bis dahin habe ich gedacht es ist grundsätzlich für alle levels ein rubric, also egal ob upper oder lower secondary. Also man kann ihn definitiv auf upper secondary auch anwenden. Ich glaube da kommt es dann eher auf die App an die man evaluiert aber nicht auf den rubric.

I: Okay. Eine andere Dimension die ich mir noch notiert habe ist D3 Appropriateness of Language Learning Material.

P4: [laughs] Das ist die dimension die ich nicht verstanden habe.

I: Du bist nicht die Einzige die es nicht verstanden hat. Ich finde es sehr interessant, weil es für mich als Entwickler des rubrics eigentlich klar ist. Aber ich sehe das Problem der Unverständlichkeit in der Formulierung durchaus. Es ist sehr ungeschickt formuliert. Ich verstehe warum das auf den ersten Blick und auch nach zweimal durchlesen unklar ist.

P4: Also das geht es um die grade-levels in der App?

I: Es geht grundsätzlich darum ob activities wenn sie von der App für eine spezifisches Level vorgesehen sind auch wirklich für die target audience geeignet sind.

P4: Ich habe nicht verstanden was die grade-levels sind. Ob es dabei um A1 or pre-intermediate learners geht, und target audience wäre dann lower secondary learners.

I: Im Kern soll es abfragen ob die activities in der App für lower und upper secondary geeignet sind. Die Formulierung lässt das aber nicht klar werden.

P4: Ich habe nicht verstanden, dass der Fokus auf dem language learning material liegt.

I: Dass damit die activities gemeint sind.

P4: Ja, genau. Wobei es ist sicher die Formulierung, aber auch wie es dort steht, dass es diesen break gibt und die Frage in die nächste Zeile springt. Die anderen Fragen, vor allem die darüber, sind kürzer, die passen gut in eine Zeile, und hier ist es eben nachteilig, dass die Frage über zwei Zeilen geht.

I: Okay, und jetzt hätte ich noch, ah genau, C6 und C7. Bei den dimensions hatte ich das Gefühl, dass du sie als sehr ähnlich wahrgenommen hast. Wenn man die erste dimension evaluiert, nimmt man vielleicht Sachen vorweg die erst in der zweiten dimension abgefragt werden?

P4: Ja, da war mein Fokus auch schon wieder auf sehr motivierend also auf motivate, und ich habe das graphics und interface ein bisschen überlesen. Da ist ja eigentlich das Design allgemein gemeint. Diese game-like elements würde ich fast mit den aesthetics zusammenlegen. Natürlich ist mein Blick da auch ein bisschen anders, vor allem weil ich mir vorher schon den rubric angeschaut habe. Und ich habe gewusst worauf man vielleicht hinauswill oder was man sich angeschaut hat, ob das vorkommt oder nicht. Wenn man den rubric vielleicht genauer Schritt für Schritt durchmacht entsteht dieses Problem nicht. Vielleicht müsstest du in Klammer dazu schreiben was genau du mit graphics and interface meinst. Also der display usw.

I: Das bringt mich gleich zu einer anderen Frage die schon etwas allgemeiner ist. Ich hätte noch ein paar allgemeinere Fragen zum rubric. Hast du das Gefühl, dass man mit dem rubric evaluieren kann ohne sich vorher mit den dimensions und domains vertraut gemacht zu haben? Braucht man ein gewisses Training damit man mit dem rubric effektiv und effizient arbeiten kann?

P4: Bei den ersten drei dimensions braucht man schon ein Training weil man wissen muss was diese Konzepte sind. Anderenfalls kann man das nicht beantworten. Sobald man mit der App ein bisschen gearbeitet hat, braucht man keine extra Erklärung oder Training für die anderen dimensions. Particularly interesting oder might be interesting ist ja trotzdem immer subjektive Interpretation des raters, also das kann nie ganz objective sein. Das heißt da braucht man eigentlich kein Training.

I: Inwiefern würdest du den rubric in deiner eigenen Praxis als Lehrerin verwenden?

P4: Gar nicht.

I: Mhm.

P4: Weil ich einfach die App anschau und dann entscheide ob ich die App cool finde, oder ich probiere die App mit eine paar Kindern in einer Lernstunde aus und frage sie was sie von der App halten. Ich würde das gar nicht so scientifically machen.

I: Spielen die Länge und die Ausführlichkeit des rubrics eine Rolle?

P4: Ja, ja.

I: Mhm.

P4: Wenn es zum Beispiel nur statements wären so wie ist es motivierend und man dann eine scale ankreuzen müsste, wäre man bei der Evaluation schneller.

I: Also du meinst man könnte die descriptors herausnehmen und die Frage der dimensions als Statement hinschreiben und dann die Skala-

P4: Ja genau, eins trifft voll zu oder umgekehrt, mhm.

I: Was hältst du von der Idee aus dem rubric, aus jeder domain drei oder vier wichtige dimensions auszuwählen und damit eine Light-Variante des rubrics zu gestalten?

P4: Das würde ich auf alle Fälle eher verwenden als die Ausführung des rubrics mit 26 Dimensionen. Aber ich weiß nicht, ich würde es wahrscheinlich nur so mit einer Likert scale machen mit trifft zu oder trifft eher nicht zu und dann die dimensions in der Form von

Statements formulieren. Zum Beispiel die App ist motivierend, ähm... ich weiß gar nicht wie man das formulieren sollte. Aber ja, ich könnte mir das eher vorstellen so etwas zu verwenden.

I: Gab es allgemein domains oder dimensions deren Sinnhaftigkeit du angezweifelt hast?

P4: Bis auf die eine dimension die ich so gar nicht verstanden habe und die paar dimensions die sich meiner Meinung nach überlappen, haben alle anderen dimension immer Sinn gemacht.

I: Okay, danke für deine Zeit und dein Feedback.

Think-aloud protocol 05, 3rd March 2021

I: Im Zuge dieser Think-Aloud-Session wirst du eine language learning app mithilfe des Evaluation Rubric for Language Learning Apps for Secondary Level Learners evaluieren. Der rubric deckt insgesamt 26 Dimensionen ab. Für jede Dimension sollst du der App zwischen einem Punkt und fünf Punkten geben. Kannst du eine Dimension bei einer App überhaupt nicht anwenden, dann wählst du auf der Skala die Option not applicable. Versuche bitte während des Evaluationsprozesses laut zu denken. Sage während du evaluierst bitte alles, das dir hinsichtlich des rubrics durch den Kopf geht. Besonders interessieren mich deine Gedanken hinsichtlich der Formulierungen, der Sinnhaftigkeit und Anwendbarkeit der einzelnen Domänen und Dimensionen. Während du evaluierst und dabei laut denkst, werde ich Notizen machen. Gibt es von deiner Seite aus noch offene Fragen?

P5: Nein, es ist alles klar.

I: Dann würde ich dich nun bitten mit dem Evaluationsprozess zu beginnen.

[Participant 5 commences with the evaluation]

[A1] Alles klar, ähm... ich habe die App Mondly angeschaut und wir starten mit domain A Educational Value. Ähm, die dimension A1 Thinking Skills, ähm... ich würde wenn ich an die App Mondly denke... äh vielleicht eine 2 hergeben, the app's content best aligns to the level 1 recall and reproduction thinking skills, recall, recognise, match, memorise. Ähm, ich finde die Formulierungen sind zwar klar aber ich finde bei der Frage fehlt ein bisschen eine Hintergrundinformation zu dem, äh, Webb's Depth of Knowledge. Äh, ich habe das halt jetzt ausgewählt wegen den Verben, also recall, recognise, match, memorise. Ähm, weil zum Beispiel create, synthesise, analyse und design würden überhaupt nicht zur App dazu passen. Die App ist eher einfach, ähm, design und äh die- (xxx) müssen also etwas matchen oder sich erinnern können und was hinschreiben aber jetzt nicht wirklich etwas kreieren.

[A2] Dimension A2 21st Century Skills... ähm, bezogen auf die App würde ich sagen, dass critical thinking oder collaboration, creativity, self-direction nicht wirklich notwendig sind oder ähm verlangt werden. Das ist fast ein bisschen Schade, aber es ist... deshalb würde ich dem Ganzen echt einen 2er oder 3er geben. The app does not require learners to engage 21st century skills. Ja. Ich finde die Formulierung ist klar, ja, ich finde es ist eindeutig was damit gemeint ist.

[A3] Dimension A3 Value of Errors, mmh, der Anwender oder die Schüler und Schülerinnen die die App benutzen, bekommen in der App soweit ich mich erinnern kann nicht die Möglichkeit sich... nein, äh, nicht die Möglichkeit die Antwort sofort auszubessern. Die Frage

wird glaube ich zurückgelegt und dann noch einmal gestellt, ähm, das heißt, äh, the app offers learners an opportunity to answer the question again but the app does not provide an explanation why their answer was incorrect the first time. Genau, es gibt keine Erklärung in der App, äh, aber trotzdem kommt die Frage später noch einmal auf. Ich finde die Frage ist klar... mmh, genau es ist eindeutig und klar formuliert, würde ich auf jeden Fall so durchgehen lassen. Und ich finde es auch wichtig, dass error correction angesprochen wird.

[A4] Dimension A4 Feedback to Teachers. How does the app allow teachers to monitor the learners' progress. Ähm, ich glaube die App ist nicht darauf ausgelegt, dass, ähm, Lehrer das überprüfen können, soweit ich mich erinnern kann. Die Kinder oder die Anwender bekommen dann immer eine, bekommen Punkte für ihre erreichten oder gemachten Aufgaben und werden dann auch gereiht. Also alle App-Nutzer werden quasi in einer Bestenliste gereiht und das wird auch angezeigt, da gibt es eine Kategorie für Weltweit und für Österreich... und da kann man sich mit dem Namen anmelden, ähm, man kann dann auch nach Personen suchen und sich Freunde hinzufügen, das heißt Lehrer könnten theoretisch ihre Schüler als Freunde hinzufügen und so tracken ob sie etwas gemacht haben, wie viele Punkte sie gesammelt haben und auch am Tag wie viel die Kinder gemacht haben. Mann kann aber keinen Lernfortschritt oder keine Bewertung sehen oder machen. Ähm, genau, ich finde die Formulierungen vom rubric in diesem Fall klar... ist alles eindeutig.

[A5] Dimension A5 Cooperative Learning. Does the app allow learners to communicate and collaborate with others. Ähm, wie vorher bereits erwähnt kann man in der App Freunde hinzufügen, richtig kommunizieren kann man nicht, es gibt keine Chatfunktion, es gibt keine Messengerfunktion oder man kann auch nichts liken oder teilen, somit nicht wirklich. Genau, man kann auch keine presentations kreieren oder größerer Projekte miteinander machen oder Aufnahmen machen oder Dokumente versenden. Genau. Die Frage ist klar formuliert, als die Dimension ist ebenfalls eindeutig. Mmh...

[A6] Dimension A6 Accommodation of Individual Differences. Does (xxx) [participant whispers]. In der App wird vor allem mit Texten gearbeitet und soweit ich das jetzt bewerten kann nach meiner kurzen Testphase nur oder kaum mit Bilden, oder gar nicht mit Bilden, ähm... es gibt schon eine Übung wo man quasi eine Bewegung machen muss und Wörter in eine bestimmte Richtung schieben muss, ähm... aber ansonsten eher eintippen oder bestimmte Wörter auswählen, falls ich die Frage richtig verstanden hab. Ähm, ja. Kinaesthetic, da weiß ich nicht ob du da was anderes damit meinst, wo man quasi mit dem Körper noch was anderes machen sollte, ja, Bewegung, mittelmäßig. Man muss halt Wörter verschieben... .Genau, also ich würde nicht sagen, dass die App darauf eingeht, dass alle Schülerinnen und Schüler unterschiedlich lernen. Es gibt in den Lektionen selbst, ähm, verschiedene Übungen aber jeder der Anwender muss eigentlich jede Übung machen um weiterzukommen. Man kann jetzt nicht auswählen zwischen unterschiedlichen Übungen, es gibt schon ein Quiz dann am Ende aber sonst ist es relativ einfach gestaltet mit den Lektionen und verschiedenen Übungen, und danach eben ein Quiz und eine Vokabellernübung und eine Lernfortschrittstest. Also ich würde sagen very little oder the app's content is presented in a way that assumes that all learners learn in the same way.

[B1] Genau, so domain B Design. Dimension B1 Ability to Save Progress. Ja die App erlaubt es einem wieder zur Übung zurückzukehren, also... genau, after exiting an app learners can reopen it and automatically return to the content/activity they were last engaging. Man

kann aber die Übung abbrechen und man kann dann eine andere Übung auswählen. Lektionen können immer wieder gemacht werden, auch die Vokabelübung, das Quiz und der Lektionstest am Enden können immer wieder wiederholt werden, ähm, man muss aber Lektionen eben abschließen um dann zum nächsten Level zu gelangen. Äh, wobei man hier vielleicht noch hinzufügen muss, dass eine Testversion eben sieben Tag lang verwendet werden kann, und danach muss man sich eine, eben die bezahlte Version kaufen. Erst durch die bezahlte Version kann man dann verschiedenen andere Übungen auch machen oder eben weitere Aktivitäten starten. Der rubric selbst... in diesem Fall ist die Frage klar, ähm... wobei der fünfte Punkt mit den fünf Punkten und der dritte Punkt sind so formuliert als würde sich das gegenseitig ausschließen, und das ist ja nicht so in diesem Fall. Man kann schon den content they were last engaging quasi öffnen, ähm, oder man kann weitermachen einerseits, andererseits kann man das auch händisch auswählen.

[B2] Dimension B2 Integration. Is the app enhanced (xxx) [participant whispers]. Ja, Mondly ist mit anderen Apps vernetzt, man kann sich auf Social Media mit Facebook verbinden, mit Google... ähm, mit der E-Mail, die E-Mail-Adresse kann man auf jeden Fall eingeben und... ähm, mit Apple kann man sich auch verbinden. Mmh, was mit online communities gemeint ist weiß ich nicht ganz genau, weil online communities sind aber auch gleichzeitig Apps. Ja, also ich würde mal sagen is enhanced by how it integrates with all four of the listed connections.

[B3] Dimension B3 Screen Design. Is the app's text, sound, graphic and speech well organised. Mmh, ja es ist auf jeden Fall gut organisiert, äh, es gibt jedoch keine oder kaum Grafiken, oder Videos gibt es eigentlich auch nicht. Es gibt Text und Geräusche und eben Sprache. Man kann sich schon etwas anhören, und die App arbeitet auch ständig damit, aber keine Bilder oder Videos. Ansonsten ist es aber sehr gut organisiert, also fünf Punkte. Die Frage ist klar formuliert, ich würde aber als step one vielleicht mal nachfragen ob es überhaupt videos, graphics, text gibt, oder wie die App generell aufgebaut ist.

[B4] Dimension B4 Ease of Use. Is the app intuitive and are learners able to (xxx) [participant whispers] with minimal guidance... Ja, also ich würde schon sagen the app is intuitive, die Schüler:innen brauchen keine Hilfe von Außenstehenden um die App bedienen zu können, es ist alles selbsterklärend. Ja, die Frage ist ganz klar formuliert und eindeutig.

[B5] Dimension B5 Navigation. How easily can learners move through the app's content and options? Als Anwender kann man auf jeden Fall relativ leicht zu allen Übungen und Lektionen kommen, ähm, es gibt zuerst Kategorien, ähm, wo verschiedene topics, verschiedene Themen aufgelistet sind, dann gibt es auch Lektionen die man eben an einem Tag machen kann, einen Tag später gibt es dann wieder neue Lektionen, es gibt eine Statistik wo die erreichten Punkte eingetragen werden. Sowohl die Minuten die man mit dem Lernen verbracht hat, Wörter, Ausdrücke und eben der Fortschritt, und eine Bestenliste sowie einen (xxx), das zählt dann zur bezahlten Version. Man kann leicht in der App navigieren, die Kategorienlandschaft, also das mit den topics ist ein bisschen... das könnte ein bisschen verwirrend sein, aber im Großen und Ganzen ist es eigentlich klar und leicht verständlich. Zur Dimension selbst... also ich würde vorher noch zur Bewertung eine vier auswählen, put forth some effort, oder eine fünf, can move through the content and options fluidly. Ich finde die dimension ist klar formuliert, die Abstufung ist auch in Ordnung.

[B6] Dimension B6 Information Presentation, ich finde nicht, dass es ganz logisch aufgebaut ist, weil es gibt diese Kategorien mit den topics und da sind die Lektionen auch drin, auch die tägliche Lektion. Da gibt es verschiedene Dinge zu erledigen, man kann sowohl, ähm, über einen Chat sprechen üben als auch den Grundwortschatz aufbauen und diese täglichen Lektionen machen. Es gibt aber unten auch noch einen eigenen Punkt, Lektionen, wo die besagten täglichen Übungen drin sind, das ist somit halt doppelt drin, das ist nicht ganz logisch aber trotzdem verständlich. Ähm, also is presented in a manner that is mostly logical. Die dimension ist klar formuliert, es gibt aber keine Erklärvideos oder Erklärungen in der App, in den Übungen selbst gibt es manchmal Wörter die unterstrichen sind und da kann man sich dann eine zusätzliche Erklärung auf Englisch und auf Deutsch anschauen, kommt aber eher selten vor. Die Dimension selbst ist klar.

[B7] Dimension B7 Media Integration. The app's text... Nachdem es eben nur Texte, sounds und speech gibt, würde ich sagen, dass die Frage nur bedingt zu beantworten, bezogen auf diese App... also da würde ich drei Punkte geben (xxx) but they still form a mostly cohesive programme. Mmh, wenn man vorher nachfragen würde ob es graphics, videos, sounds, oder anderes gibt, dann hätte man die Frage stellen können, wenn ja, are the app's texts, graphics, usw. effectively integrated. Also ich würde vielleicht zuvor klarstellen was es in der App gibt und dann im Anschluss fragen.

[B8] Dimension B8 Cultural Sensitivity, finde ich jetzt schwierig zu beantworten... weil ich nicht alle Kategorien (topics) anschauen konnte oder nur wenige Kategorien anschauen konnte. Es gibt Kategorien wie Reisen, Reisevorbereitungen, Urlaubsaktivitäten, Romantik, das sind gleich einige der ersten Kategorien, und das würde ja voraussetzen, dass die Schülerinnen und Schüler in sozial stärkeren Familie aufwachsen und auf Reisen gehen können oder Urlaubsaktivitäten machen können. Das setzt auf jeden Fall das voraus. Ich weiß aber nicht inwiefern die Fragen dann auf Kulturen eingehen, ähm... Ich finde aber das die App eher so wirkt, dass sie nicht auf sozial benachteiligte Schüler:innen eingeht wenn man gleich Reisen als ersten Punkt anbietet. Romantik ist vielleicht auch etwas fragwürdig, im Alter von 11 bis 14, wie viel Romantik die Kinder da erleben. Ansonsten finde ich die Frage gut und wichtig... schwierig zu beantworten, not applicable würde ich sagen.

[C1] Domain C Engagement, dimension C1 Learner Control... Ja, grundsätzlich kann der Anwender, die Anwenderin das Level auswählen, wobei das Level am Anfang nicht beschrieben wird. Es steht glaube ich Anfänger, dann Mittelmäßig Fortgeschritten und Fortgeschritten und es ist nicht definiert wie fortgeschritten man ist und ich glaube es ist als Schüler:in auch schwierig einzuschätzen... Also ich würde so ein mittelmäßiges... drei Punkte vergeben. The app allows learners some control related to the level at which the begin engaging its content. Die Frage ist klar formuliert, die Abstufung ist auch in Ordnung.

[C2] Dimension C2 Interactivity, es wird nicht nur Inhalt vorgelegt, es gibt nicht nur Input, sondern die Anwender können auch etwas eintippen oder einsprechen. The app frequently requires learners to engage its content through activities, da kann man fünf Punkte vergeben. Die Frage ist interessant und gut, ich finde Interaktivität in einer App sehr, sehr wichtig, dass man nicht nur liest oder etwas anhört sondern auch die productive skills gefördert werden. Ich finde die Dimension gut.

[C3] Dimension C3 Pace, nicht überall können die Kinder die Geschwindigkeit kontrollieren... eigentlich muss man in jeder Übung bestätigen, also überprüfen lassen und dann weitergehen zur nächsten Aufgabe, zur nächsten Übung. Nur bei der Wiederholung von Vokabeln werden die Vokabeln noch mal aufgesagt auf Englisch und dann ganz, ganz schnell auf Deutsch. Das ist eigentlich zu schnell, die Kinder können sich da sicher nichts mitnehmen. Ansonsten haben die Nutzer das schon unter Kontrolle. Also the app allows learners to mostly control the rate at which they move through the app's content, vier Punkte. Die dimension ist interessant, ich finde gut, dass du an das gedacht hast. Ich glaube man kann nicht zurückgehen in der App, was auch noch interessant wäre. Jeder lernt anders und anders schnell und deshalb finde ich das echt wichtig, dass es im rubric vorkommt.

[C4] Dimension C4 Personalisation, ähm, Personalisierung ist in der App nur bedingt möglich. Man kann ein Profilbild auswählen, soweit ich mich erinnern kann, man kann schon ein- oder ausschalten ob man Hintergrundgeräusche oder Musik haben möchte. Aber sonst gibt es keinen Avatar, die Farbe vom Hintergrund kann man nicht auswählen, es gibt auch keine Figuren die auf einem Spielfeld weiterrücken... Genau, das heißt ich würde auswählen... richtig personalisieren nur ein feature, das wären dann zwei Punkte. Learners can personalise one feature of the app. Die Frage ist klar formuliert, die Abstufung ist in Ordnung.

[C5] Dimension C5 Interest, will the app's language learning material appeal to lower secondary learners. Ich glaube schon, dass lower secondary learners ansprechen auf diese App, man könnten aber sicher noch mit mehr Bildern und Kurzvideos das Interesse steigern. Es gibt in einer Lektion immer sehr viele unterschiedliche Übungen aber die wiederholen sich dann immer, und da es nur Text ist, kann ich mir vorstellen, dass es für Zehn- bis Fünfzehnjährige irgendwann langweilig wird. Also drei (Punkte), some of the app's language learning material might be interesting and engaging to lower secondary learners. Die Dimension ist auch klar formuliert, ich finde es auch wichtig, dass man das interest von den Kindern erwähnt, weil um das soll es gehen, und interest ist eigentlich ein sehr, sehr großer Punkt. Wenn die Kinder kein Interesse daran haben, dann werden sie es nicht in ihrer Freizeit machen. Und das ist wahrscheinlich der Sinn dahinter, dass die Kinder sich im Bus, im Zug, in den Öffis oder zuhause hinsetzen und die App öffnen, und das regelmäßig. Oder dass man den Englischunterricht nach außen trägt und das ganze gamifiziert.

[C6] Dimension C6 Aesthetics, ja nachdem es keine Bilder gibt, würde ich sagen, dass diese Frage nicht beantwortbar ist. Wobei in der Auswahl selbst gibt es schon Bilder, in den Übungen dann aber nicht... in der Kategorie Romantik gibt es zum Beispiel ein Herz aus Blumen, das ist jetzt meiner Meinung nach nicht so ansprechend für Kinder, bei Urlaubsaktivitäten ist ein Ehepaar, also ein Mann und eine Frau, abgebildet, die älter sind, also sicher über 60. Das ist jetzt auch nicht direkt für die Zielgruppe gemacht. Auch bei den Reisevorbereitungen ist eine junge Frau am Bild die wohl so ungefähr 30 ist. Ich würde sagen das ist nicht auf learners zwischen 11 und 15 zugeschnitten. In den Übungen selbst gibt es eh kaum Bilder. Die Frage ist klar, könnte man aber auch als Unterfrage stellen.

[C7] Dimension C7 Gamification, ja, es gibt auf jeden Fall Punkte und eventuell dann auch einen Preis wenn man eine Kategorie abgeschlossen hat. In der App selbst gibt es oben drei Sterne die erfüllt werden wenn man die Aufgaben richtig beantwortet, es gibt auch eine Leiste in der angezeigt wird wie viel von der Lektion schon erledigt wurde, also einen

Fortschrittsbalken, was sicher die Motivation bei den Nutzern steigert aber wie gesagt, es gibt keine Avatare. Ein Ranking gibt es auch, also the app uses three game-like elements, vier Punkte würde ich vergeben.

[D1] Domain D Language Learning, dimension D1 Quality of Input, ich würde auf jeden Fall sagen der Inhalt der App ist verständlich und auch relevant... mmh... die Antwortmöglichkeiten, wenn man auswählen kann bei so multiple choice questions, da sind die Antwortmöglichkeiten nicht so divers. Es ist relativ klar, ohne dass man viel können muss, welche Antwort da hinein gehört, das heißt auch nicht so besonders wertvoll würde ich sagen. Und ich finde es schwierig, the app provides mostly comprehensible and valuable input. It covers language that is mostly relevant to lower secondary learners, also vier Punkte. Zur dimension selbst, ich finde es schwierig dass man comprehensible, valuable and relevant in einer dimension zusammenfasst. Ich finde in der App ist eigentlich alles comprehensible, also alles verständlich formuliert und auch klar, es gibt immer eine Deutsche Übersetzung, was aber noch lange nicht heißt, dass das ganze valuable gestaltet ist. Ich finde Deutsche Übersetzungen werden zu oft eingesetzt, oft wäre es nicht notwendig. Auch bei gesprochenen Übungen gibt es eine Deutsche Übersetzung und das ist nicht notwendig. Das würde ich dann nicht als valuable bewerten.

[D2] Dimension D2 Opportunity to Produce Output. Ja, der oder die Nutzer:in kann schon beides, written und spoken output, produzieren, ähm, wobei ich mir gewünscht hätte, dass es noch mehr written output wäre. Oft muss man Buchstaben in die richtige Reihenfolge bringen oder die richtige word order finden, da hätte man schon auch eine Schreibübung daraus machen können. Spoken output, da war ich überrascht wie viel da in die App integriert wurde. Es gibt eine Chatfunktion wo man etwas einsprechen kann, und auch in den Lektionen selbst gibt es Übungen mit Sprecherelementen. Es ist aber eigentlich alles predefined, wobei in der Chatfunktion kann man schon ein bisschen freier agieren, man chattet da übrigens mit einem Computer, also nicht mit anderen Nutzern... auf jeden Fall limited to the sentence level, also drei Punkte. Die dimension finde ich gut, mmh, vielleicht könnte man es aufteilen, eine eigene dimension für written und eine eigene für spoken output erstellen. Bei Mondly ist beides ähnlich, aber wenn in einer App nur written oder nur spoken output produziert werden kann, dann wird es schwierig mit der Beantwortung.

[D3] Dimension D3 Appropriateness of Language Learning Material, ich finde schon, dass es in Ordnung war, ähm, es werden immer wieder von der Oxford University Press Beispiele hinein genommen, ich glaube die App möchte damit die authenticity steigern, es ist aber trotzdem aus dem Kontext herausgerissen. Ich finde man könnte das Ganze noch mehr mit Texten oder mit listenings unterstreichen. Es ist eher immer auf dem sentence level, mit word order und so. Aber im Großen und Ganzen ist es schon appropriate für die target audience. Also ja, usually matched appropriately, vier Punkte. Die dimension finde ich gut formuliert, ist eindeutig und klar.

[D4] Dimension D4 Skills, mmh... äh, ehrlich gesagt schwierig zu beantworten. Ich finde, dass writing nur bedingt eingebaut wird. Man muss einmal pro Lektion oder einmal pro Übungsblock vielleicht ein Wort schreiben. Ob man das writing nennen kann, schwierig, also vier Punkte würde ich sagen. Zur dimension, eigentlich könnte man diese dimension schon früher (im rubric) einbauen, vor der dimension D2 könnte man das schon fragen.

[D5] Dimension B5 Exposure to the Target Language. Schwierig. Ich würde nicht sagen, dass die exposure to the target language maximiert wird. Es wird ständig auf Deutsch übersetzt, und auch am Ende der Übungen, also am Ende der Lektionen gibt es immer noch eine Deutsche Übersetzung obwohl das dann schon eindeutig und klar sein sollte... Also nehmen wir vielleicht zwei Punkte, (xxx) translations are very frequent... wobei, nein, eigentlich könnte man auch nur einen Punkt vergeben. Ja, ich gebe nur einen Punkt weil die App selbst eigentlich auch alles auf Deutsch hat. Also sogar Hallo steht einmal da, tägliche Lektion, Statistik, das ist alles auf Deutsch, das könnte man vielleicht umstellen aber grundsätzlich ist alles auf Deutsch. Die Dimension selbst ist eigentlich auch klar, man könnte vielleicht sogar die Antworten ein bisschen kürzen. Ansonsten ist es aber eindeutig, und ich finde die Frage auch wichtig. Genau. Ja, das war es soweit.

[Participant 5 finishes the evaluation]

[The interviewer commences the follow-up interview]

I: Danke für die Evaluierung. Ich habe mir ein paar Fragen beziehungsweise Notizen gemacht und würde noch ein paar Sachen abklären wollen. Ich würde mit der letzten domain beginnen, mit Language Learning. Dimension D2 Output, du hast gemeint du würdest diese dimension eventuell aufsplitten in zwei, einmal in written und einmal spoken. Wenn du dir die dimension durchliest, wirkt es so als müsste beides in der App vorkommen?

P5: Ähm... [reads guiding question] stimmt, du hast recht. Vielleicht habe ich das vorher falsch gelesen. Nachdem du ja geschrieben hast produce written OR spoken output-

I: Sollte man das OR vielleicht irgendwie hervorheben damit man es nicht überliest?

P5: Ich bin mir aber nicht sicher ob es das besser macht, dass da OR steht. Was ist wenn beides vorkommt. Bei dieser App wird ja auch written output verlangt, zwar nur minimal aber trotzdem. Wie beantworte ich da dann die Frage. Does the app provide written or spoken output. Yes [laughs].

I: Also könnte man die Dimension mit der Formulierung stehen lassen und nicht aufsplitten, oder findest du, dass es wichtig wäre, dass man das auf zwei individuelle Dimensionen aufteilt?

P5: ich denke es hängt davon ab was du rausfinden möchtest, ob der Fokus auf written oder spoken output liegen soll. Aber im Endeffekt kannst du es auch zusammen in der dimension lassen. Das war ein Lesefehler von mir.

I: Ok, die nächste dimension ist D3 Appropriateness of Language Learning Material. Lies dir bitte die Frage nochmal durch. Mich würde interessieren wonach für dich in dieser dimension gefragt wird. Wonach wird konkret gefragt?

P5: Naja, ob die Inhalte der Übungen in der App, ähm, auf die Zielgruppe zugeschnitten sind. Also quasi ist es in Ordnung, dass ich die... ob die Intention dahinter in Ordnung ist, ob es auf die Zielgruppe zugeschnitten ist.

I: Mhm, okay.

P5: ich finde, dass die Frage klar ist.

I: Gut, gehen wir weg von der domain D und hin zu domain B Design. Die dimensions B3 und B7, findest du, dass die zwei sehr ähnlich sind, und dass man die zwei schwer voneinander unterscheiden kann?

P5: ... [reads guiding question] Mmh, ja, also die beiden Fragen sind sehr ähnlich, und ich glaube ich habe auch bei beiden Fragen etwas ähnliches geantwortet. Und well-organised,

das stimmt schon. Wenn ein Programm cohesive ist, dann ist es auch well-organised. Das eine setzt das andere voraus. Da könntest du eigentlich eine dimension rausgeben. Ja.

I: Zu den descriptors, zu der Skala selbst. Es gibt im rubric ja Abstufungen mit Zahlen, also zum Beispiel four of the skills, three of the skills und es gibt auch Abstufungen mit adverbs, zum Beispiel frequently, mostly und so weiter. Ist es für dich in Ordnung, dass Zahlen und adverbs gemischt vorkommen, oder hast du eine Präferenz?

P5: Ich finde es mit adverbs fast besser weil es mehr Spielraum lässt, weil gerade bei Zahlen, so wie es bei der dimension B2 war, finde ich es nicht immer ganz klar ob es zwei oder drei sind. Oder auch wie viele features kann man personalisieren. Nur weil man einstellen kann ob eine Hintergrundmusik läuft oder wenn man das Ganze heller oder dunkler haben möchte, heißt es ja noch nicht, dass es eine Zahl mehr ist. Ich würde eher Adverbien verwenden, wobei es bei manchen Fragen sicher Sinn macht Zahlen einzusetzen.

I: Gut, dann hab ich noch drei allgemeinere Fragen an dich. Im rubric wird in den Formulierungen hauptsächlich auf lower secondary learners fokussiert. Wenn du jetzt mit dem rubric arbeitest und an die App allgemein denkst, könnte man die App und den rubric auch auf die upper secondary anwenden?

P5: Also die App würde ich selbst nicht für die upper secondary verwenden, denn es gibt nur drei Schwierigkeitsgrade. Ich habe den Mittleren angeschaut und der war schon sehr niedrig angesetzt. Ich glaube nicht, dass das für die Oberstufe in Englisch passt, für den rubric selbst, du meinst, dass man andere Apps die auch für die Oberstufe –

I: Genau, ich meine, dass man eventuell die Formulierung mit den lower secondary learners aus dem rubric rausnimmt, weil man mit dem rubric Apps evaluieren kann die eventuell besser für die Oberstufe geeignet sind. Besonders dimensions wie A1 und A2 fragen nach Skills die eher auf die Oberstufe abzielen.

P5: Ja, wäre wahrscheinlich sinnvoll, aber ich finde nachdem deine Masterarbeit eher darauf abzielt zwei Apps anzuschauen die für die lower secondary passend sind, würde ich es nicht zwingend rausnehmen. Wenn man den rubric für die Oberstufe verwenden möchte, kann man die Formulierung ja anpassen.

I: Wenn du an die App denkst, an Mondly, inwiefern würdest du sagen, dass man die App im Unterricht in der Schule verwenden kann.

P5: Ähm, ich bin grundsätzlich dem gegenüber aufgeschlossen, dass Technik im Unterricht verwendet wird. Ich finde es aber schwierig, dass man Kinder die Nutzung ihres eigenen Smartphones erlaubt, weil trotzdem immer wieder Nachrichten ankommen und die Kinder trotzdem wieder abgelenkt sind. Das hat jetzt aber nichts mit der App Mondly selbst zu tun. Ich finde, dass die App nicht sehr tief in die Materie eindringt und man keine bestimmten Themen auswählen kann. Man kann jetzt zum Beispiel nicht auswählen adjective or adverb oder a or an, deshalb würde ich es im Unterricht glaube ich nicht einbauen. Außerhalb des Unterrichts in der Freizeit ja, einfach um die exposure zu erhöhen. Im Unterricht selbst würde ich es nicht machen.

I: Meine abschließende Frage: inwiefern würdest du den rubric in deiner Praxis als Lehrerin verwenden, vor allem im Hinblick auf den Umfang und die Ausführlichkeit.

P5: Du meinst wenn ich eine App verwenden-

I: Genau, also wenn wir davon ausgehen, dass du eine App evaluierst, würdest du denn rubric verwenden?

P5: Mmh, ich finde einige Fragen sind schon sehr sinnvoll, weil man in der Praxis in der Schnelligkeit vergisst, dass zum Beispiel die exposure to the target language, also dass nicht zu viel Deutsch dabei ist, das man darauf vergisst, oder... das die App auf alle vier skills eingeht. Wobei das meistens wahrscheinlich schon in unseren Köpfen ist. Aber es gibt schon

einige Dinge, Gamification ist schon ein Punkt, dass man sich immer wieder in Erinnerungen rufen soll, dass Kinder gerne mit Spielen lernen. Ich finde der rubric zeigt schon einige Punkte auf die wichtig sind wenn man eine App evaluiert. Die Antwortmöglichkeiten finde ich schon ausführlich, aber eh in Ordnung, weil es sonst unklar ist. Man kann nicht einfach nur eine Abstufung von eins bis fünf machen... Manche Fragen sind schon, zum Beispiel dimension A5, das würde ich kürzen. Ich würde versuchen die Fragen so klar und kurz und knapp wie möglich zu formulieren. Aus der Frage sollte schon klar sein in welche Richtung es geht. Das ist dir bei den meisten Fragen eh sehr, sehr gut gelungen. Ich finde den rubric gut, ich finde ihn in Ordnung.