Computer Assisted Language Learning

ISSN: 0958-8221 (Print) 1744-3210 (Online) Journal homepage: https://www.tandfonline.com/loi/ncal20

Toward a neuroscience-informed evaluation of language technology

M. Ali Bolgün & Tatiana McCaw

To cite this article: M. Ali Bolgün & Tatiana McCaw (2019) Toward a neuroscience-informed evaluation of language technology, Computer Assisted Language Learning, 32:3, 294-321, DOI: 10.1080/09588221.2018.1516675

To link to this article: https://doi.org/10.1080/09588221.2018.1516675

Published online: 04 Nov 2018.

COMPUTER ASSISTED LANGUAGE LEARNING
2019, VOL. 32, NO. 3, 294–321
https://doi.org/10.1080/09588221.2018.1516675

Toward a neuroscience-informed evaluation of language technology

M. Ali Bolgün and Tatiana McCaw

Defense Language Institute Foreign Language Center, Monterey, CA, USA

ABSTRACT

With the ever-increasing number of available language technology products, there is also a need to evaluate them objectively. Unsubstantiated beliefs about what language technology can and cannot do inside or outside the language classroom often influence decisions about the choice of language technology to be used. The declarative/procedural model, which makes a clear distinction between the declarative and procedural memory systems, can help to provide an objective, neuroscience-informed evaluation of language technology. The central argument in this paper is that language technology caters predominantly to the declarative memory system. This system is very effective in forming explicit metalinguistic knowledge but does not lead to automatic production or procedural ability. For technology to promote procedural ability, it should instead cater to the procedural memory, which involves the implicit neurofunctional computational system. This paper provides a language technology evaluation flowchart to help professionals evaluate the language technology products they will use and gauge their expectations of those products more realistically. It also provides a list of factors to be taken into account in maximizing the benefits of technology.

KEYWORDS
Declarative; procedural; implicit; explicit; evaluation of technology; language education; apps

Introduction

Given the speed with which technology in general and educational technology in specific is evolving and the multiplicity of web tools and applications (apps) available (DuBravac, 2013), a common belief is that the power of technology is consequentially transferred to improved language teaching and learning. In fact, educational technology expenditure is one of the largest in the technology market and is expected to reach over $250 billion dollars by 2020 (PR Newswire, 2016). While the assessment of technology's effects on language learning has been notably difficult

CONTACT M. Ali Bolgün alibolgun@gmail.com

(Burston, 2006), the intuitive expectation may be that, with the significant increase in the use of technology, language learning should prove easier or student proficiency levels be higher. This belief, which is not substantiated by empirical research (Golonka, Bowles, Frank, Richardson, & Freynik, 2014; Macaro, Handley, & Walter, 2012; Stevenson & Liu, 2010; cf. Wang & Vasquez, 2012), assumes that language learners are now fortunate in that, with this surfeit of technology, it is easier for them to attain the proficiency levels that older generations attained via time-consuming and arduous labor.

Conversely, there are language professionals who assert that technological tools are simply tools that, in and of themselves, do not lead to faster or better comprehension of a foreign language (DuBravac, 2013; Fuchs & Snyder, 2013; Kern, 2006; Lettvin, South, & Stevens, 2016; López-Burton, 2014, *inter alia*). This assertion, however, does not explain why this is the case.

A neuroscientific perspective – specifically, the declarative/procedural (DP) model – may explain why technology[1] has not lived up to the expectation that it would revolutionize language learning (Anwaruddin, 2017; Macaro et al., 2012). In fact, the central argument we make is that language technology has been effective in building metalinguistic knowledge but not as effective in forming procedural ability because language technology caters predominantly to the declarative memory system. This memory system involves explicit metalinguistic knowledge, does not lead to automatic production and is limited in capacity. For language technology to help learners 'acquire' a language at higher levels, it should also cater to the procedural memory system, which is used almost exclusively when acquiring one's native language and involves implicit neurofunctional computation (Chomsky, 2016; Paradis, 2009; Ullman & Lovelett, 2018). Accordingly, this article proffers a flowchart to help language professionals make objective and neuroscience-informed evaluations of language technology products, and gauge their expectations of such products. We hope this will develop a neuroscience-informed mindset among language professionals so that they may devise techniques that benefit from such tools and apps.

A brief history of the use and evaluation of technology in language education

Excitement about the use of technology in language teaching is not new. From the phonograph to the radio, the telephone to the spectrograph, the dormiphonics technique to language labs (Keating, 1963), and television to computers, researchers have demonstrated a great deal of interest, discussing their pedagogical advantages and motivational benefits and citing the stimuli they bring to the classroom and the type of learning (e.g. independent learning) they promote (Salaberry, 2005).

With the growth of personal computer use, computer assisted instruction (CAI) generated excitement, and researchers warned that if educators did not take advantage of this breakthrough, unfavorable outcomes would ensue (Dunkel, 1987; Lindenau, 1984). However, assessment of CAI indicated that most of the software programs did essentially the same things that books had been doing, just electronically (Garrett 1991; Kleinman, 1987). This led to the development of computer-assisted language learning (CALL), the purpose of which was the empirical evaluation of emerging technologies. Starting in the 1980s, communicative CALL emerged, shifting focus from accuracy to fluency. The feedback[2] feature emerged as technological advancement allowed for more interactive options, expanding the usefulness of technology for a wider range of proficiency levels. The 1990s saw increased levels of interactivity and feedback, allowing for greater student autonomy. Additional interactivity allowed for computer-mediated communication, essentially expanding communication capabilities, real-world relevance, and learners' ability to take charge of the learning process (Blake, 2013).

In the first few years of the twenty-first century, Web 2.0 emerged – with the term itself first used in 2004 – and created excitement because this new approach to web page design allowed users to both collaborate and create content as opposed to simply consuming the content created by earlier – Web 1.0 – designers (Han, 2011). Web 2.0 apps available to consumers are now measured in the millions, with educational apps being the third most popular category as measured by the number of downloads (Statista, 2018). This, along with ubiquitous mobile devices, makes it possible to take language education beyond the confines of the brick-and-mortar classroom (Kukulska-Hulme & Shield, 2008).

These developments prompted the need to assess the effectiveness of the new technology. However, the level of impact of current technology on language education in terms of its ability to positively contribute to higher levels of language proficiency has still not been established conclusively (Macaro et al., 2012); 'for most technologies, actual increases in learning or proficiency have yet to be demonstrated' (Golonka et al., 2014, p. 92). Two discrete meta-analyses of studies of the effectiveness of technologies used in language learning found that technology-supported language instruction/learning is as effective overall as such instruction/learning conducted without technology (Grgurović, Chapelle, & Shelley, 2013; Zhao, 2003). Grgurović et al. (2013) further found that 'in studies using rigorous research conditions' – such as, 'employing random placement of subjects into conditions' (p. 191) – CALL groups performed better than non-CALL groups. Similarly, a comprehensive meta-analysis of research studies on the effectiveness of online learning[3] concluded that 'students in online conditions performed *modestly better* [emphasis added], on average, than those learning the same material through face-to-face instruction' (Means, Toyama, Murphy, Bakia, & Jones, 2010, p. xiv). The authors of these three studies caution that the findings should be interpreted carefully due to reasons that include, among others, small effect sizes, the research setting mostly being in higher education (as opposed to being equally represented across primary, secondary, college, and private language schools), language proficiency of participants mostly being at the beginner level (as opposed to including participants at different levels of language proficiency), and random sampling being rarely employed. In addition, Means et al. (2010) state that studies often fail 'to report retention rates for students in the conditions being contrasted and, in many cases,' there is 'potential bias stemming from the authors' dual roles as experimenters and instructors' (p. xviii).

The gap between high expectations of language technology and lesser impact often lead to frustration without knowing the main cause. For example, frustration about web apps not leading to the advertised goals of the products or users' perceived language proficiency goals is often communicated without stating the fundamental reason as to why this is so (see, for example, Andersen (2017) and Groves, Hopkins, and Reid (2015)).

Burston (2006) argues that the reason newer technologies have failed to bring about higher levels of proficiency in language education – that is, higher than those achieved in classrooms using older technology –$9# may be because they are used in ways that resemble those older-technology classrooms. Prensky (2005) expresses a similar sentiment, arguing that technology adoption in the schools is a four-step process: (1) 'dabbling' which simply means the haphazard involvement of individuals with access to few computers; (2) 'doing old things in old ways' – essentially means doing the same things exactly as before, but with technology; for example, putting courses, books, and curricula online instead of using printed resources; (3) 'doing old things in new ways' involves using, for example, tutorial videos for explaining grammatical concepts in a flipped classroom.[4] Prensky argues that despite the improvement, simulations, whether 'in sand, on paper, and in their heads,' have been used for millennia and, as such, fall under the third step; and (4) 'Doing new things in new ways,' for which Prensky does not offer specific uses of technology; rather, he argues that we need 'new curricula, new organization, new architecture, new teaching, new student assessments, new parental connections, new administration procedures, and many more elements.'

The DP model may be adapted to evaluate language technology objectively as it delineates how, under what conditions, and to what extent the two distinct brain memory systems learn and acquire different elements of human language. As such, it offers a set of determinants against which a given language technology can be evaluated. The next section provides a brief account of the model as it relates to this paper.

The neuroscientific view of 'learning' and 'acquisition'

In the DP model[5] proposed separately by Paradis and Ullman (Paradis 2009; Ullman, 2004, 2015, 2016; Ullman & Morgan-Short, 2012), two distinct brain memory systems (namely *declarative* and *procedural*) work in different ways. Accordingly, both declarative and procedural memory systems are long-term memory systems. However, declarative memory handles explicit, metalinguistic knowledge while procedural memory is responsible for implicit, procedural ability. Speakers acquire native languages through the procedural memory system, except for the vocabulary – $9#the form-meaning *or* sound-meaning relationship, as explained below – $9#whereas adult learners typically have to rely on the declarative memory system.

One crucial distinction that the DP model makes is between 'vocabulary' and 'lexicon' (in fact, between 'vocabulary' and all other implicit features of language). Lexicon 'refers to the implicit grammatical properties of lexical items,' and '*not* the form-meaning relations that represent what is called the vocabulary' (Paradis, 2004, p. 12). Vocabulary is subserved by declarative memory in both L1 and L2. Note that the grammatical properties of words are usually not explicit; 'they are not learned as part of the word but are acquired as elements of the implicit grammar of each language subsystem' (Paradis, 2009, p. 14). These properties are independent of the words' semantic meaning constraints and phonological forms; for example, whether a verb is intransitive or reflexive, requires that it take a direct or an indirect object; or, whether a noun is a count or a mass noun and hence whether it can take the plural form. These features are language-dependent; a noun may be a mass noun in one language but a count noun in another. For example, *information* in English is a mass noun and does not normally take a plural suffix (as in **informations*). *Bilgi*, 'information' in Turkish, on the other hand, is a count noun and can take the plural suffix *-ler* (as in *bilgiler*). One language may require a feminine article before a given noun (e.g. *die Sonne* 'sun' in German), while in another language the noun that refers to the

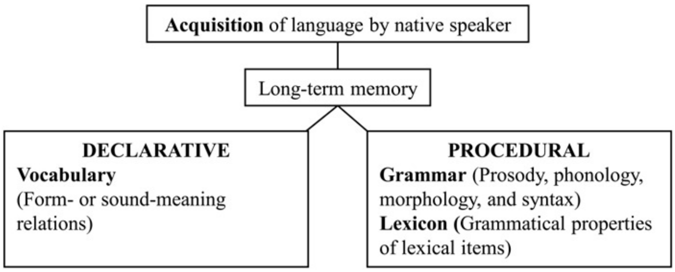

Figure 1. Involvement of long-term memory types in language acquisition (based on the DP model).

same entity may require that a masculine article precede it (e.g. *le soleil* 'sun' in French). Vocabulary is found in dictionaries; it is a list of words and their meanings. They are seen and heard, and speakers learn their form-meaning associations. Conversely, lexicon is acquired through use by encountering words in varying sentential contexts (Paradis, 2009).

As Figure 1 schematically shows, language acquisition (as is the case when one acquires a native language) is handled through the procedural memory system, except for the vocabulary. Contrast this with Figure 2, which schematically shows that the declarative memory is used almost exclusively for language learning. When adults learn a language, it is this declarative memory system that will be used. Procedural memory is involved only to the extent that language elements are practiced and repeatedly used.

Most language learners have a number of sound- or form-meaning associations at their disposal, but they lack the competence related to the words' morphosyntactic properties, which may vary from those of the learners' native language (Paradis, 2009). Stated another way, when processing *vocabulary*, both native and non-native speakers use the declarative memory system. This explains why memorizing lists of vocabulary – including being repeatedly exposed to vocabulary through, for instance, vocabulary learning apps – does not necessarily lead to language acquisition, but instead leads to metalinguistic knowledge of vocabulary. However, when it comes to *lexicon*, or any other automatized processes, such as sentence construction, L1 speakers use the procedural memory system, whereas L2 speakers continue to use the declarative memory system unless they have proceduralized at least portions of L2. As they repeatedly use the lexical items in context, and as these uses are tallied by the procedural memory, learners also begin to rely more on the procedural memory system.

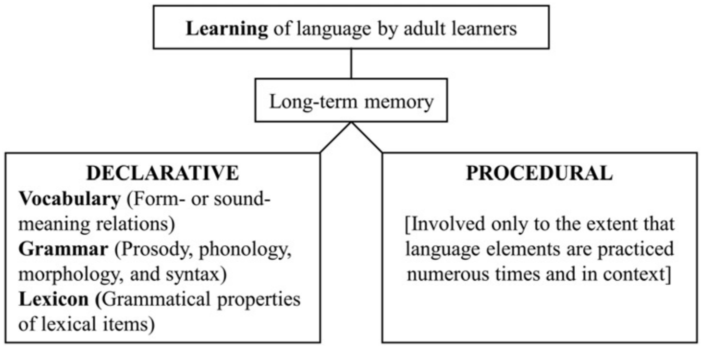

Figure 2. Involvement of long-term memory types in language learning (based on the DP model).

When native speakers generate a sentence, they usually do not consciously select words; this is taken care of by the 'implicit neurofunctional computational system' just as it automatically selects other phenomena, such as tense, agreement, and word order (Paradis, 2009, p. 19). Second language speakers do not produce sentences in the same fashion. Their construction of sentences, as well as their selection of words, are declarative, controlled tasks. Thus, sentence processing is not only more complex than single-word processing, but also supported by a different type of memory system and consequently involves different cerebral structures that occupy different anatomical locations (Paradis, 2009).

As adult second or foreign language learners are exposed to and practice language elements, they acquire those elements through, and rely upon, the procedural memory system. For elements that have not been acquired, they rely upon their knowledge in the declarative memory system. Novice adult learners of a foreign language initially rely almost exclusively upon the declarative memory system; their language production is laborious, slow, and likely contains mistakes. Proficient learners (e.g. those that used the target language in the target culture and/or were exposed to it for many years) likely rely more upon the procedural memory system (Paradis, 2009; Ullman & Lovelett, 2018).

One research finding important for language learning is that metalinguistic knowledge does not gradually become implicit linguistic competence (Ellis, 2005; Ellis & Larsen-Freeman, 2006; Paradis, 2009; Ullman & Lovelett, 2018). '[T]here is no continuum between metalinguistic knowledge and implicit linguistic competence' (cf. Anderson, 1982)[6], 'but only between the degree of use of one system and of the other' (Paradis, 2009, p. 28). This is because metalinguistic knowledge and implicit linguistic competence, by their different natures, are 'intrinsically incapable of affecting each other's content and structure. Only the output of implicit competence can interact with metalinguistic knowledge because it is observable and conscious' (Paradis, 2009, p. 44). The use of procedural memory may replace the use of the declarative, as proficiency increases, or it may replace it in rapid succession, when speakers resort to their use of metalinguistic knowledge in declarative memory to compensate for a gap in their incomplete implicit competence in procedural memory. These memory systems do not directly feed, impact, transmit information to, or influence each other in any way. "Individuals may know the rule and be able to verbalize it and apply it in writing, in consciously controlled speech, or in a grammaticality judgment task; […] nevertheless, [they] may continue to systematically produce an incorrect form represented in their (inaccurate) implicit linguistic competence. (Paradis, 2009, p. 29)".

Over time and to the extent that the implicit competence improves, there should be less need for the learner to depend on the declarative memory, as the procedural memory handles language production automatically.

It is unfortunate for adult language learners that the availability of procedural memory for acquiring language as a whole decreases with age (Paradis, 2009; Ullman, 2016). In addition, the optimal periods for acquiring the various parts of language differ. Accordingly, the earlier an individual is exposed to the prosody of the language, the better. Later comes phonology, followed by morphology, and then syntax. Since each of these parts has a different optimal period, different learners will internalize different sets of implicit rules in each linguistic module. This often results in acquiring one portion of the grammar completely while other portions are only partially acquired (Paradis, 2009).

The dilemma that learners face is as follows: they must pay attention to and notice what is to be learned; otherwise they cannot learn it. However, what is *acquired* is not observable, and, therefore, is not noticeable (Paradis, 2009; cf. Schmidt, 1990). This means that learners cannot pay attention to elements that are not observable because they cannot even be perceived. What they may do is pay attention to input, which is the surface form of that which is perceived. However, they cannot pay attention to intake, which is what serves as material for the computational underlying connections on the basis of frequency of use. This is shown in experiments where participants who were trained on an artificial grammar lacked conscious awareness of the competence they had acquired (Dienes, Altmann, Kwan, & Goode, 1995). In other words, they could produce grammatical sequences but were unable to explain how they did so.

Another crucial observation to make within the framework of the DP model is that if an element is not used, its underlying structure will not be internalized. The element that is used is not the one that has an effect; the implicit tallying of its underlying structure is what leads to internalization. 'What is tallied is the number of times the underlying structure is implicitly abstracted' (Paradis, 2009, p. 54). For example, native speakers from a very young age automatically and accurately inflect non-words[7] that they have never before heard (Berko-Gleason, 1958). This shows that the frequency of 'specific words' used in certain combinations is not what is important; rather, it is the procedural memory's ability to process the underlying properties of these words.

Implications of the neuroscientific findings of the DP model on technology use in language education

The memory type that technology caters to

Given the neuroscientific account of acquiring versus learning languages outlined above, it is evident that the most effective way to appropriate a language is through acquisition (see also Krashen, 1982); that is, through a great deal of communicative interaction and practice in the target language with native speakers in which the procedural memory system is activated and in charge. For technology to have a greater impact on language education, it should enable such interaction. However, we should also note that unlike children, adults have largely lost the ability to learn a language without thinking about its structure and, as such, they rely on alternative mechanisms to learn it (DeKeyser, 2000). Learners must be able to 'notice' language elements to learn them. Technology can help learners with *noticing* language elements, via animations, blinking fonts, colors, and sounds, among others. While it is true that what is noticed results in metalinguistic knowledge, not proceduralized knowledge, this metalinguistic knowledge could enable the learner to construct correct sentences, the use of which then serves as a means to implicit intake. Technology should then allow for the repeated conscious use of correct forms. This, in turn, would provide input from which intake is implicitly abstracted and tallied, leading to acquisition.

The importance of creating novel sentences and receiving corrective feedback

Language technology cannot on its own lead to acquisition of language elements needed for one to operate at higher levels of proficiency if it simply presents a set of words or sentences stored in its memory, as is the case with vocabulary learning software. This is because it is practically impossible to predict, store in computer memory, and teach all possible sentences a user will need to proficiently use at higher levels the language being learned. This means that, with the exception of a limited number of formulaic everyday expressions (such as, 'nice to meet you,' 'good morning,' and 'have a nice day') that they memorize at the lower levels of proficiency and with the exception of also a limited number of sentences they can emulate by simply changing a word or two within model sentences (such as, 'Can I have … ?' or 'I like … ing.'), language learners will have to create their own novel sentences and even combine them into paragraphs in order to eventually reach higher levels of language proficiency. For example, to qualify as a Level 2 speaker on the Interagency Language Roundtable (ILR) language proficiency scale, one would, among other things, have to demonstrate that they can hold 'casual conversations about current events' (ILR, 2011) – which can vary from a news item on a traffic accident to another news item on a high-level visit by foreign dignitaries. As the level goes up even higher (e.g. Level 3), the speaker would need to be 'able to speak the language with sufficient structural accuracy and vocabulary to participate effectively in most formal and informal conversations in practical, social and professional topics' (ILR, 2011). For instance, they would need to produce several-paragraph-long arguments on topics such as gun control. Accomplishing such tasks requires language skills far beyond simply memorizing sentences stored in computer memory.

Learners who aim to reach such high levels of language proficiency will have to form novel sentences as they learn a language. In fact, this should be encouraged and they should be provided with ample opportunities to do so, and in doing so, they will likely make mistakes on which they should receive feedback. Therefore, technology – especially, in its *tutor* role (Kern, 2006) or *tutorial* category (Garrett, 2009) – must allow for (a) the formation of well-constructed novel sentences within a discourse and (b) when the output is flawed (i.e. when the learner produces ungrammatical sentences), it should be able to provide feedback that permits explicit analysis. The feedback should not only illustrate whether a sentence is accurate but also why it is inaccurate and how to modify it for greater accuracy. This is because languages are rife with subtleties that deviate from grammar rules; therefore, simply providing a correct alternative to an ungrammatical sentence without an adequate explanation does not necessarily guide the learner in composing well-formed sentences. This task becomes more complicated as language proficiency levels increase because not only do learners then need sentence, paragraph, and discourse-level feedback but they also need such feedback to show the interrelatedness of sentences, involving deixis, and may include sociocultural explanations of the importance of following sociocultural norms, among others.

Well-formulated and timely feedback is crucial in ensuring successful adult language learning and is perceived as 'very useful' by learners (Liu et al., 2013). In fact, we argue that the lack of such feedback is what, to a large extent, limits the currently available language technology. Even on social network sites, where learners can receive feedback from the other participants, feedback is not always forthcoming. When learners do receive feedback, the quality of that feedback is debatable and can vary greatly from one network to another (Liu et al., 2013). This is because not all native speakers know, for example, the distinction between two seemingly synonymous expressions, and even when they do, these speakers may not be able to formulate their intuition in a way that would make sense to the learner (Haegeman & Guéron, 1999). Consequently, while users of social media can theoretically provide feedback (Brick, 2011), the feedback provided may not necessarily be as useful for language learners.[8] When using social media sites (or other technology), teachers must ensure that learners are receiving timely and accurate feedback on as much of their language output as possible (cf. Gruba & Clark, 2013), as error correction can be useful in improving, and even in accelerating, adult foreign language learning and acquisition when it is done correctly (Lyster, Lightbown, & Spada, 1999; Schulz, 2001). Note that while feedback and error correction do not always lead to immediate learning, they can be seen as a catalyst for later learning (MacKey & Philp, 1998). Acquisition is not an instantaneous process (Doughty & Williams, 1998), and achieving positive effects through error corrective feedback is a long-term process. This means that language learners should receive corrective feedback continuously until the language elements are proceduralized; only then can learners automatically produce those elements. Therefore, providing feedback is not simply important, but crucial in language learning and acquisition (Brandl, 2008).

Why, one may ask, is current technology problematic in terms of error correction? All technical difficulties aside, the reasons include the following: error correction involves multiple strategies, such as recasts, clarification requests, asking students to repeat, asking questions, and metalinguistic feedback. For example, guided feedback gives the students a chance to self-repair. Some guided feedback strategies are metalinguistic feedback, a teacher's request to repeat, asking questions, pinpointing, and pausing. For technology to be more effective in error correction, it must do more than just figure out what to correct. For example, to provide precise feedback about a learner's error, not only do experienced

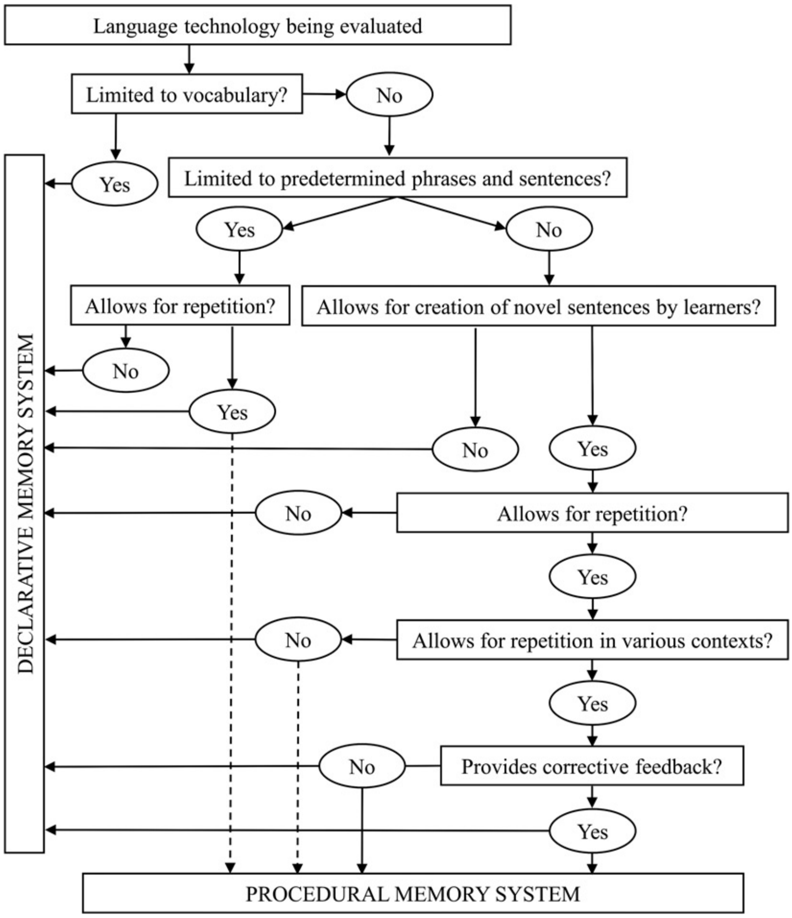

Figure 3. Flowchart for determining the memory type a technology-as-tutor caters to.

teachers determine the error, but they also understand it and, more importantly, identify the source of the error. In some cases, this is a daunting task, especially when one considers how complex grammar can be.

Technology should also be able to employ more than just one feedback strategy of providing what the correct answer or language form is. For instance, elicitation, which enables learners to draw on their metalinguistic knowledge, has been shown to be very effective as a feedback strategy (Loewen & Philp, 2006; Lyster & Ranta, 1997). Recasts, on the other hand, are effective with phonological errors (Mackey & Philp, 1998). Learners have shown the ability to note sound discrepancies between their own pronunciation and a teacher's pronunciation, a skill that makes such errors more salient than morphological errors. Such pronunciation corrections should be done simultaneously and within the context of one-on-one speaking sessions with a native speaker. This means that while pronunciation correction software or apps may be useful at the level of pronouncing words in isolation or within isolated sentences, higher levels of proficiency require that the learner pronounce words comfortably in connected speech at the paragraph level. This is most efficacious with teachers who are ideally native speakers of the target language, partly because students may have follow-up questions regarding the teacher's feedback or correction, and the teacher may have follow-up questions on students' understanding. Technology has not reached a level of sophistication sufficient to deal with such an exchange.

Corrective feedback, therefore, plays a crucial role in a learner's process of learning and eventually acquiring a language (Brandl, 2008; Russell & Spada, 2006). However, computers cannot easily, accurately, and consistently offer variegated error correction. Also, unlike experienced teachers, computers are unable to take into account students' affective states. Experienced teachers continuously evaluate class dynamics and decide whether to provide feedback at a given moment during instruction. When they decide to do so, they also consider how to correct and which (out of a possible few) errors require correction and how extensive this feedback should be. These decisions are based on the teacher's knowledge of learners' attributes (e.g. personalities, cognitive styles, motivation levels, learning preferences, etc.) within the context of classroom dynamics. For example, teachers sometimes decide to involve the learners in the correction process, rather than unilaterally correcting errors. These are the major limitations of currently available technology.[9]

A neuroscience-informed evaluation of technology

In light of the discussion above, it should be clear that evaluation of language technology in terms of its effectiveness in language education presents challenges. We offer the following flowchart (Figure 3) that language professionals may find useful in evaluating language technology, especially those that are designed to perform 'tutor' roles.[10]

As indicated in the flowchart, if the language technology is limited to vocabulary – even when allowing for repetition of that vocabulary, it will by default involve only the declarative memory system (Recall the crucial distinction the DP model makes between 'vocabulary' and 'lexicon' as discussed earlier.). If the language technology includes a set of formulaic expressions and discrete sentences and no capability for allowing the user to create novel sentences, it will typically involve the declarative memory system. Such technology may also potentially involve the procedural memory system to a limited extent and only if it at the same

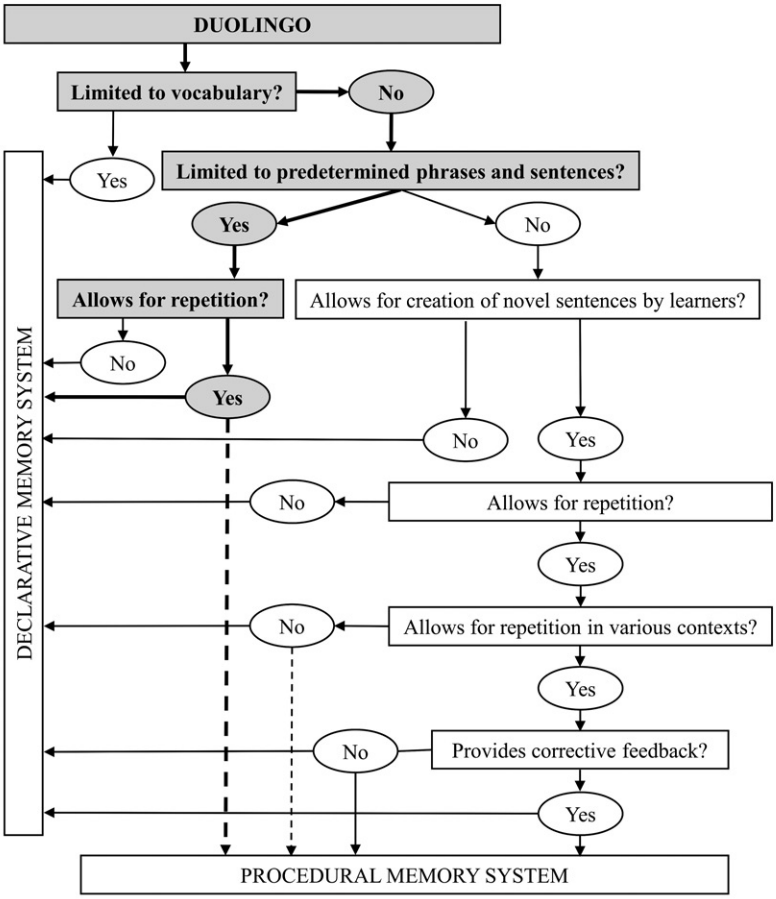

Figure 4. Flowchart as used to determine the memory type *Duolingo* caters to.

time allows for repetition since 'procedural memory learns[11] […] with repeated exposure' (Ullman, 2016, p. 960).[12] Limited involvement of the procedural memory system is indicated by dashed lines in Figure 3. If the language technology allows for the creation of novel sentences but does not allow for repetition of these sentences in context, it will involve the declarative memory system because adult language learners' creation of novel sentences is a laborious, slow, controlled, and conscious process and such processes are handled by the declarative memory. If the language technology does indeed allow for both creation and repeated use of novel sentences but does not have the corrective feedback capability, the procedural memory system will be involved but it will most likely lead to the acquisition and eventual fossilization of incorrect forms, along with the correct ones, as novice adult language learners are bound to make errors when forming sentences. An ideal language technology that involves the procedural memory system would be one that – in addition to knowledge and practice of vocabulary – allows for creation of novel sentences and their repeated use in various contexts while providing corrective feedback all along. Note that the declarative memory system will be involved under all conditions for two reasons: (1) it is nearly impossible for most adults to learn a language without consciously processing at least portions of it and (2) vocabulary, which is an essential part of language learning, is handled by the declarative memory system (as described by the DP model above).

Evaluating language technology with the flowchart

Below is a sample evaluation of *Duolingo*, using the flowchart in Figure 3 to further illustrate how it can be used in language technology evaluation. *Babbel*, *Kahoot!*, *Pimsleur*, *Quizlet*, and *Rosetta Stone*[13] were also evaluated. However, only *Duolingo* is discussed in this paper due to page limitation; the evaluation results of the other products are listed – without detailed discussion – in Figure 5 according to the memory types they cater to. These products were chosen because they fall in the technology-as-a-tutor category (see Kern (2006) and Garrett (2009) for categorization of technology). *Duolingo*-Turkish[14] evaluation was done with the assumption that the users were attempting to learn Turkish only through *Duolingo*, and not in conjunction with any other source. Also, the evaluation is valid as of 10 February 2018 and does not pertain to versions that may have been available after that date.

Duolingo

Duolingo is built around teaching vocabulary and a limited number of phrases and simple sentences in any one of over 30 languages. Users of the program are introduced or exposed to words or sentences, and then they are made to practice them in the following ways:

- choosing the right word from among a few picture-prompts
- typing the Turkish word or sentence in English or the English ones in Turkish
- saying the word or sentence aloud, following the prompt.

Occasionally, users are quizzed on words or expressions that have not yet been introduced. The users have the option of revealing the English translation to find out what it means by hovering the cursor over the prompt. This word or sentence is then later asked again, aiming at helping the users to memorize it. Also, some grammar descriptions that users can read are provided, if they wish to do so.

Language technology product	Declarative	Procedural for a limited and fixed number of sentences	Procedural
Babbel	✓	✓	
Duolingo	✓	✓	
Kahoot!	✓		
Pimsleur	✓	✓	
Quizlet	✓		
Rosetta Stone	✓	✓	

Figure 5. Some language technology products and memory type they cater to.

If the learner responds to the prompt correctly, s/he is given another one, and if and when s/he responds to all the prompts correctly and completes the lesson, s/he is given a digital prize and motivating feedback (such as, 'Great! You levelled up!'), and s/he progresses to the next lesson. If the user responds to all the prompts correctly without making any mistakes, s/he moves to the next lesson after about 10 prompts. If the learner fails to supply the right response to one or more than one prompt, then the number of prompts within each lesson increases since the incorrectly answered prompts appear again for the user to try again, exposing the user to the word or sentence multiple times.

Duolingo does not allow user-generated, novel sentences; it only permits the uttering of sentences[15] (by repeating after the voice prompt), or translating the sentences that are prompted from English to Turkish or vice versa. That is not to say that *Duolingo* prohibits users from forming novel sentences; it is just that users do not have the option of forming novel sentences that they may type or speak into the system and receive automated instantaneous feedback on them.

Given the above, the path in the flowchart is highlighted in Figure 4.

In accordance with the DP model summarized earlier in the paper, and the evaluation shown in Figure 4, we predict the following: the vocabulary-learning component will activate the declarative memory system by default. However, since *Duolingo* allows for repetition of the phrases and sentences, it would also activate the procedural memory system but only to an extent that is strictly limited to the number of those sentences (see the explanation in endnote #12), provided that the sentences share the same underlying structure and the user revisits previously completed lessons and practices them often. *Duolingo* sends out reminders via email to registered users, reminding and encouraging them to practice the language for which they signed up. Users of *Duolingo* (and, in fact, of all similar technology, some of which are listed in Figure 5) can expect to learn some or all of the vocabulary presented and practiced therein, and utter a limited number of sentences (The number of words users can learn or the ease with which they can learn and retain them depends, among other factors, on the capacity of their declarative memory.). They can also expect to utter memorized sentences or ask memorized questions. However, they cannot expect to maintain simple face-to-face conversations even on familiar topics (unless they use the language technology product in conjunction with a language course – such as a college language course – that encourages and enables creation of novel sentences and provides timely corrective feedback). On the ILR language proficiency scale,[16] users can expect to potentially show some of the characteristics of ILR Level 0+ and perhaps, but to a significantly lesser extent, Level 1 (ILR, 2011).

Our prediction about the level *Duolingo* users may reach is corroborated by the findings of another study that evaluated *Duolingo*-Spanish. In that study, participants who received the lowest scores on the WebCAPE placement test taken prior to starting *Duolingo*-Spanish (i.e. 'the beginner/novice group' who would qualify for a first semester Spanish 101 course) 'had the biggest improvement' as indicated by their post-test scores, while those who already had higher levels of proficiency in Spanish (i.e. those whose placement test scores would qualify them for third semester Spanish 201) 'showed more modest improvement' (Vesselinov & Grego, 2012, p. 17). Stated another way, just as we predict, *Duolingo* has the potential to raise the users' language proficiency level from 0 to a point where the user shows some of the characteristics of 0 + (and, to a substantially lesser extent, Level 1) on the ILR scale. However, if a user is already at that level, it cannot raise the user's language proficiency to higher levels – assuming that the user is not using *Duolingo* in conjunction with any other language program or course (see Krashen (2014) for his review of Vesselinov & Grego (2012) as it relates to *Duolingo*).

Figure 5 is a matrix of language technology products and the memory type they cater to. As indicated in Figure 5, *Kahoot!* and *Quizlet* cater to the declarative memory because in their typical use of quizzing, users will be engaged in conscious, metalinguistic processing of words and sentences; not in repeated use of them. However, if *Kahoot!* and *Quizlet* are used to allow some sentences to be practiced repeatedly, they may also activate the procedural memory to a limited extent, as do the other products listed in Figure 5.

Discussion, recommendations, conclusion, and further research

In its tutor role, technology has a deterministic function (cf. Anwaruddin, 2017), which can be evaluated. Therefore, the argument that ‘technology is just a tool’ is neither sufficient nor helpful. Instead, this article offers a neuroscience-informed explanation to the puzzlement as to why the spectacular advances in technology have not delivered equally spectacular high levels of language proficiency. In a nutshell, the reason is that despite the level of sophistication technology has reached, it simply cannot override the way the brain acquires language. As explained earlier, the brain has two distinct memory systems – declarative and procedural – and currently available technological products that function in the tutor role cater predominantly to the former, mostly aiming at teaching vocabulary. Learning vocabulary, while an important element in language education, is not sufficient in helping the language learners reach higher levels of proficiency. This is because vocabulary (as opposed to lexicon) is handled by the declarative memory. Lexicon and all other implicit features of a language are handled by the procedural system. For the procedural system to be activated, users of these products must have the option of creating novel sentences in context, about which they should receive corrective feedback. This is missing from these products.

Creating novel sentences is crucial for language learners because this gives them the opportunity to use in context the words and grammar structures that they may have learned earlier, and to receive feedback about the proper use of those words and newly constructed sentences. The feedback may include elements, such as pronunciation of words within a sentence (which may be different from the way the words are pronounced in isolation), intonation of the sentence, appropriateness of words as used in the sentence (perhaps, some words should be substituted for other ones), choice of grammar structure (perhaps, ‘used to’ needs to be used instead of ‘would’), proper use of cohesive devices (maybe ‘normally’ is more appropriate than ‘usually’) within a sentence and across sentences, as well as cultural appropriateness of the sentence (perhaps, the sentence has culturally offensive elements or the idea that the learner is trying to convey is stated differently in the language s/he is attempting to learn), and so forth. This practice of creating sentences, receiving feedback on them, and reformulating sentences and uttering (or writing) them again is what would eventually and subconsciously involve the procedural memory.

We hope that the flowchart, along with the arguments we put forth, will be helpful in at least two ways: (1) for language technology evaluation and (2) for approaching language education issues with a neuroscience-informed mindset.

Regarding the language technology evaluation, the flowchart is helpful in that it enables language professionals to objectively determine the strengths and limitations of a given technology product and to predict the extent to which it can help learners reach the language proficiency levels they aim to reach, thereby potentially eliminating frustration. Its objectivity comes from the fact that the evaluator has to follow the flow-chart steps that are developed according to findings in neuroscience. Doing so eliminates bias, or teacher and learner beliefs as to what technology can or cannot do, or attraction to a technology product due to its novelty, among others.

Following the 'mindset' mentioned in the previous paragraph, language professionals should always utilize technology in ways that enable learners to *produce* the TL. For example, even an activity as simple as matching a number of words with pictures or other words on an interactive whiteboard need not be done silently, with students simply dragging items around a screen. Instead, the teacher can encourage students to narrate what they are doing during the activity, providing corrective feedback throughout the activity.

Similarly, when choosing a particular technology, a teacher can determine, among other factors, if it provides multiple sources of input, such as aural and visual support. For example, in the case of flashcards apps, when determining which application to use, all else being equal, the choice should be the application that has captioning or pronunciation help that accompany the words. Students are thereby more likely to improve recognition memory, as bimodal input has a positive effect on learners' decoding ability, potentially leading to improved future intake of larger amounts of comprehensible input. Additionally, bimodal input increases language processing capacity (Bird & Williams, 2002; Hulstijn, 2003) probably because 'phonological information derived from text and sound both contribute to improvements in processing of spoken words' (Bird & Williams, 2002, p.527).

Language technology companies should realize the importance of corrective feedback followed by extensive practice in various contexts when working with even the simplest-looking grammar structures and how lack of such feedback can demotivate the language learner. While user feedback and collaborative learning is valuable, with no authority to make the final call as to which of the multiple explanations is correct, users can be left without reliable explanations. Here is an example from *Duolingo*: '*Ben güzelim*' is given as a prompt, and users are asked to translate it into English. The intuitive expectation is that this simple sentence should not cause complications; users should be able to respond by typing the answer 'I am beautiful.' Nevertheless, even simple (or simple-looking) sentences necessitate thorough, on-the-spot explanations, followed by practice with similar sentences, in order to acquire the underlying grammar structure. In the sentence *'Ben güzelim,'* some *Duolingo* users posted a number of legitimate questions for the online discussion group members to answer, including those related to the suffix *'–im'* (at the end of the word *'güzelim'*), and why the suffix should be used in the form of *'-im'* (IPA: [im]) and not in any of the other possible options (i.e. allomorphs) of *'-ım,'* *'-um,'* or *'-üm,'* (IPA: [ɯm], [um], and [ym]) or why this suffix should be used at all (i.e. why can we not simply say '*Ben güzel*'?), and whether or not the subject pronoun *'Ben'* can be left out and only *'güzelim'* used (because Turkish is pro-drop and the suffix *'-im'* already indicates the intended pronoun).

Those involved in the discussion groups attempt to sincerely answer these questions; in fact, there is noticeable sense of mutual respect and collaboration. However, group members do not always answer all the questions that are posted and when they do, their explanations are not always accurate or well-formulated (this was also observed by Haegeman & Guéron (1999), and Liu et al. (2013), cited earlier in the paper; see also Ware & O'Dowd (2008) on peer-review issues). Incidentally, accurate or not, grammar explanations lead to metalinguistic knowledge, not to procedural ability (see discussion on the DP model earlier in the paper). Still, accurate metalinguistic knowledge enables users to form more accurate sentences, which in turn could serve as valuable input for the procedural memory system. Lack of accurate metalinguistic knowledge leads to errors that may fossilize.

As a possible solution to the above, explanations could be closely monitored and a database of accurate explanations to frequently asked questions could be complied and maintained. When needed, users can be directed to relevant explanations. Multiple supplementary sentences and exercises can be added for each of the explanations.

Regarding the pronunciation of words and sentences, technology needs to be improved significantly in at least two ways: (1) better speech recognition that will accurately recognize user pronunciation and (2) feedback on the user pronunciation of words or sentences that goes beyond just telling the user his/her pronunciation is simply correct or incorrect. Currently, technology lags behind both.

Pertaining to the first point are some of *Duolingo*'s prompts that ask users to click the microphone icon and say the word or the sentence that is displayed on the screen. During our evaluation, in some cases, even when our pronunciation was intentionally not accurate, the system would still accept it as correct. For instance, the second vowel (the undotted 'i') in the question *'Nasıl?'* (which means 'How?' in Turkish) is the high back unrounded vowel [ɯ]. We pronounced it with [i], which is the high front unrounded vowel, and the system still accepted it as correct. Such limitations on automatic speech recognition would most likely lead users to assume that their pronunciation is correct and, therefore, not seek help on it, leading to the fossilization of incorrect pronunciation.

Pertaining to the second point, in situations in which user pronunciation is not accepted, the user has no way of knowing why his/her pronunciation was not accepted. Was it because s/he mispronounced all or only some of the syllables in all the words of a given sentence? Was it because of bad intonation? Was it because of a software glitch? What aspect of the user pronunciation was inaccurate, and what should the user do to correct his/her pronunciation?

Pronunciation and feedback become increasingly more important and more complicated as the proficiency level rises. At higher levels of proficiency, with paragraph-long exchanges and faster speech, pronunciation of words begins to differ from the pronunciation of words in isolation. This is an important issue that language technology has to resolve.

Further research can provide additional insight regarding the issues and arguments presented in this paper. This may involve neuroimaging tools to identify the memory system activated when the research subject is producing language. Longitudinal studies that involve implementation of a given language technology could be conducted and participants monitored periodically using neuroimaging to determine which memory system is currently used, and at what point the procedural memory takes over (e.g. after how many hours of study, or after how many repetitions of a given language element). Such findings could inform language professionals in their efforts to improve language technology that addresses the procedural memory system.

Notes

1. Note that, though important, this paper is not concerned with the technical or economic aspects of technology, such as temporary or permanent unavailability, or level of user-friendliness or complexities of websites and software, or subscription fees, among many others. The discussion applies to situations in which users do not encounter such issues and they have full access to technology.
2. Feedback and its importance are discussed later in the paper.
3. The online learning that they evaluated was not necessarily for language teaching; it was online learning in general.
4. 'Flipped classroom' refers to an instruction model in which students work outside the classroom on content that would traditionally be delivered in a lecture style by a teacher in the classroom. In the flipped classroom model, the bulk of class time is devoted to meaningful application activities with the focus on target language communication (Bergmann & Sams, 2012).
5. Also referred to as Paradis/Ullman perspective (Libben, 2006), Ullman/Paradis account (Clahsen & Felser, 2006), declarative/procedural model of Paradis (Falk, Lindqvist, & Bardel, 2015), among others, we interpret the two to be compatible.

Any differences that may exist between Paradis' and Ullman's accounts are negligible for the purposes of this paper.

6. Anderson (1982), unlike Lewicki, Czyzewska, and Hoffman (1987) and Reber, Allen, and Regan (1985), theorizes that procedural knowledge develops from declarative knowledge. For Bialystok and Ryan (1985) and Bialystok and Smith (1985) these are independent. For them, declarative knowledge develops along a continuum from unanalyzed to analyzed, whereas procedural knowledge develops along a continuum from controlled to automatic. In the DP model, they are independent memory systems.
7. The most renowned, of course, is the 'WUG' test, devised by Berko-Gleason (1958). As part of this test, children are asked to complete sentences, such as this one: *This is a WUG. Now there is another one. There are two of them. There are two*_________.
8. In fact, such limitations, coupled with problematic peer evaluation (Ware & O'Dowd, 2008), lead some researchers to question the value of even pursuing research on social networks with respect to language learning (Lamy & Mangenot, 2013).
9. This is not to say that technology will never reach a level of sophistication that incorporates all the current limitations stated above. In fact, applications that employ natural language processing are now capable of providing more detailed feedback when compared to those that are more traditional (see, e.g. Nagata, 2002). Also, see Tegmark (2017) for a brilliant account of the incredible progress made so far in computer technology and artificial intelligence, and for a prediction about the levels of sophistication they may reach in the future.
10. We left out technology in its 'tool' and 'medium' roles since, as the names of the roles indicate, they either provide access to resources or an ability to connect participants. Such technologies do not contribute to or interfere with participants' language learning in the sense that they do not 'teach' content (see Garrett (2009) and Kern (2006) for their categorizations of technology). This is not to say that technology should not be used, for example, for distance education, or for conducting language classes over the Internet; in fact, there may be many reasons for doing so. See, for example, Schwienhorst (2011) for a thorough discussion about the role multi-user domains, object oriented may play in language learning, and in helping establish or support learner autonomy. However, when technology is used as a medium, we should keep in mind that communicating over the Internet is not intrinsically superior to face-to-face communication, as it will not necessarily involve procedural memory to any greater degree. For example, in an action research involving three separate groups, one engaging in technology-mediated task-based instructional design, one engaging in the same task-based – but no technology – design, and a third that engaged in equivalent textbook activities, Solares (2014) found that the three groups achieved similar linguistic gains.
11. Here, the verb 'learn' is not used contrastively with 'acquire,' which would have been a better choice. 'Learn,' here, should be interpreted to mean as forming neuronal connections.
12. The involvement of the procedural memory system is possible here because at least some of the sentences that are repeated within the same or different modules of a given technology product, such as *Duolingo*, are likely to share the same underlying syntactic structure. For example, the two sentences, *La bicicleta no es pequeña* 'The bicycle is not small' and *El tren no es rojo* 'The train is not red,'

share the same underlying structure even though their surface forms differ. Being exposed to sentences with different surface forms but the same underlying syntactic structure repeatedly could potentially involve the procedural memory, which tallies 'the number of times the underlying structure is implicitly abstracted' (Paradis, 2009, p. 54). In fact, the 'procedural memory system […] is well suited for learning implicit knowledge about rules, sequences, and categories' (Ullman & Lovelett, 2018, p. 43) and the system 'proceeds gradually through repeated exposure' (Ullman & Lovelett, 2018, p. 41; also, Knowlton & Moody, 2008). Weber, Christiansen, Petersson, Indefrey and Hagoort (2016) demonstrate that brain 'regions known to be involved in syntactic processing' (i.e. the procedural memory system) show fMRI syntactic repetition effects, one of which is repetition enhancement (an increase in neural activity when stimuli are repeated) to infrequent unfamiliar structures (p. 6877). However, a crucial point to keep in mind about procedural memory system involvement in this case (where sentences are repeated but no novel sentences are created) is that it is likely to be very limited to perhaps acquiring a few syntactic features, if at all. Therefore, it should *not* be construed as being sufficient enough to enable the users of the products mentioned in this paper to acquire a foreign language at higher levels of proficiency.

13. *Duolingo* is a trademark of Duolingo, Inc. *Babbel* is a trademark operated by Lesson Nine GmbH. *Kahoot!* is a trademark of Kahoot! *Pimsleur* is a trademark of Beverly Pimsleur, used by Simon and Schuster under exclusive license. *Quizlet* is a trademark of Quizlet, Inc. *Rosetta Stone* is a trademark of Rosetta Stone, Ltd. We, the authors, have no commercial or proprietary interest in any of these companies or the products mentioned in this article. Also, we are not affiliated with any organizations or companies having a direct financial interest in the materials or products discussed in this article.
14. Turkish was chosen as the language of study in *Duolingo* because one of the authors of this paper is a native speaker of Turkish and this gave us a native-speaker intuition advantage when it came to acceptability judgement of sentences being presented or asked as part of an activity. Also, the same author is a subject matter expert in Turkish and has the academic background to evaluate the grammar explanations provided therein or the explanations given by other users through *Duolingo*'s online discussion platform.
15. The exception to that is the option to have artificial-intelligence powered simple conversations with *Duolingo Bots*. However, these conversations are never user-generated, and the context is limited.
16. For ILR language proficiency level descriptions, please see ILR (2011).

Acknowledgements

The authors would like to thank Dr. Anita Bowles and the anonymous reviewers for their comments and suggestions, Mr. Michael McGuire for English-editing the manuscript, and Mr. Jeff Hansman for English-editing an earlier draft.

Disclosure statement

The views expressed in this article are those of the authors and are not necessarily the official policies or views of, or endorsed by, the Defense Language Institute Foreign

Language Center (DLIFLC), or any other institutions or companies that are mentioned in this paper.

Notes on contributors

M. Ali Bolgün holds a PhD in Applied Linguistics and currently works as a language curriculum development (CD) advisor. His research interests include neurolinguistics, corpus linguistics, discourse analysis, and language technology and its impact on language education.

Tatiana McCaw holds an EdD in Instructional Leadership and currently oversees a faculty development program. Her research interests include neurolinguistics, learning strategies, critical thinking, and technology integration in foreign language teaching.

ORCID

M. Ali Bolgün http://orcid.org/0000-0002-7589-6660
Tatiana McCaw http://orcid.org/0000-0001-6924-8272

References

Andersen, T. (2017, October 2). Has technology revolutionized language learning? *Voxy*. Retrieved from https://voxy.com/blog/2017/10/language-learning-revolutionized-technology/

Anderson, J. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*, 369–406.

Anwaruddin, S. M. (2017). Beyond determinism and instrumentalism: Re-conceptualizing technology for CALL. In J. Perren, K. Kelch, J. S. Byun, S. Cervantes, & S. Safavi (Eds.), *Applications of CALL theory in ESL and EFL environments* (pp. 22–35). IGI Global.

Bergmann, J., & Sams, A. (2012). *Flip your classroom: Reach every student in every class every day*. Washington, DC: ISTE.

Berko-Gleason, J. (1958). The child's learning of English morphology. *Word*, *14*, 150–177.

Bialystok, E., & Ryan, E. B. (1985). A metacognitive framework for the development of first and second language skills. In D.E. Forest-Pressley, G. MacKinnon, & T. Waller (Eds.), *Metacognition, cognition, and human performance* (pp. 207–252). New York: NY: Academic Press.

Bialystok, E., & Smith, M.S. (1985). Interlanguage is not a state of mind: An evaluation of the construct for second language acquisition. *Applied Linguistics*, *6*, 101–117.

Bird, C., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics*, *23*, 509–533.

Blake, R. J. (2013). *Brave new digital classroom: Technology and foreign language learning* (2nd ed.). Washington, DC: Georgetown University Press.

Brandl, K. (2008). *Communicative language teaching in action: Putting principles to work*. Upper Saddle River, NJ: Pearson Prentice Hall.

Brick, B. (2011). Social networking sites and language learning. *International Journal of Virtual and Personal Learning Environments*, *2*(3), 18–31.

Burston, J. (2006). Working towards effective assessment of CALL. In R. P. Donaldson & M. A. Haggstrom (Eds.), *Changing language education through CALL* (pp. 249–270). London and New York: Routledge.

Chomsky, N. (2016). *What kind of creatures are we? (Columbia themes in philosophy)*. New York, NY: Columbia University Press.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*, 3–42.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*, 499–533.

Dienes, Z., Altmann, G. T. M., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 1322–1338.

Doughty, C. J., & Williams, J. (Eds.). (1998). *Focus on form in classroom second language acquisition*. New York, NY: Cambridge University Press.

DuBravac, S. (2013). *Technology in the L2 curriculum*. New York, NY: Pearson.

Dunkel, P. (1987). Computer-assisted instruction (CAI) and computer-assisted language learning (CALL): Past dilemmas and future prospects for audible CALL. *The Modern Language Journal*, *71*, 250–260.

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, *27*, 305–352.

Ellis, N. C., & Larsen-Freeman, D. (2006). Language emergence: Implications for applied linguistics—Introduction to the special issue. *Applied Linguistics*, *27*, 558–589.

Falk, Y., Lindqvist, C., & Bardel, C. (2015). The role of L1 explicit metalinguistic knowledge in L3 oral production at the initial state. *Bilingualism: Language and Cognition*, *18*, 227–235.

Fuchs, C., & Snyder, B. (2013). It's not just the tool: Pedagogy for promoting collaboration and community in social networking in CMC. In M. N. Lamy & K. Zourou (Eds.), *Social networking for language education* (pp. 117–134). Basingstoke: Macmillan

Garrett, N. (1991). Technology in the service of language learning: Trends and issues. *The Modern Language Journal*, *75*, 74–101.

Garrett, N. (2009). Computer-assisted language learning trends and issues revisited: Integrating innovation. *The Modern Language Journal*, *93*, 719–740.

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning*, *27*(1), 70–105.

Grgurović, M., Chapelle, C., & Shelley, M. C. (2013). A meta-analysis of effectiveness studies on computer technology-supported language learning. *ReCALL*, *25*(2), 165–198.

Groves, M., Hopkins, D., & Reid, T. (2015, June 12). Just how effective are language learning apps? *The Conversation*. Retrieved from https://theconversation.com/ just-how-effective-are-language-learning-apps-42913

Gruba, P., & Clark, C. (2013). Formative assessment within social network sites for language learning. In M. N. Lamy & K. Zourou (Eds.), *Social networking for language education* (pp. 177–193). Basingstoke: Macmillan

Haegeman, L. M. V., & Guéron, J. (1999). *English grammar: A generative perspective*. Malden, MA: Blackwell Publishing.

Han, S. (2011). *Web 2.0*. New York, NY: Routledge.

Hulstijn, J. (2003). Connectionist models of language processing and the training of listening skills with the aid of multimedia software. *Computer Assisted Language Learning*, *16*(5), 413–425.

ILR. (2011). *Interagency Language Roundtable*. Retrieved from http://www.govtilr.org/

Keating, R. F. (1963). *A study of the effectiveness of language laboratories: A preliminary evaluation in twenty-one school systems of the Metropolitan School Study Council*. New York, NY: Institute of Administrative Research, Teacher College, Columbia University.

Kern, R. (2006). Perspectives on technology in learning and teaching languages. *TESOL Quarterly*, *40*(1), 183–210.

Kleinman, H. (1987). The effect of computer-assisted instruction on ESL reading achievement. *The Modern Language Journal*, *71*, 267–273.

Knowlton, B. J., & Moody, T. D. (2008). Procedural learning in humans. In J. H. Bryne (Ed.), *Learning and memory: A comprehensive reference (vol. 3): Memory systems* (pp. 321–340). Academic Press of Elsevier.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.

Krashen, S. D. (2014). Does Duolingo "trump" university-level language learning? *International Journal of Foreign Language Teaching*, *9*(1), 13–15. Retrieved from http://www.sdkrashen.com/content/articles/krashen-does-duolingo-trump.pdf

Kukulska-Hulme, A., & Shield, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. *ReCALL*, *20*(3), 271–289.

Lamy, M. N., & Mangenot, F. (2013). Social media-based language learning: Insights from research and practice. In M. N. Lamy & K. Zourou (Eds.), *Social networking for language education* (pp. 197–213). Basingstoke: Palgrave Macmillan.

Lettvin, E., South, J., & Stevens, K. (2016). Idea to retire: Technology alone can improve student learning. *The Brown Center Chalkboard*. The Brookings Institution. Retrieved from http://www.brookings.edu/blogs/brown-center-chalkboard/posts/2016/03/25-technology-improving-student-learning-south-stevens-lettvin

Lewicki, P., Czyzewska, M., & Hoffman, M. (1987). Unconscious acquisition of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 523–530.

Libben, G. (2006). How language learners comprehend and produce language in real time. *Applied Psycholinguistics*, *27*, 72–74.

Lindenau, S. (1984). The teacher and technology in the humanities and arts. *The Modern Language Journal*, *68*, 119–124.

Liu, M., Evans, M. K., Horwitz, E., Lee, S., McCrory, M., Park, J.-B., & Parrish, C. M. (2013). A study of the use of social network sites for language learning by university ESL students. In M. N. Lamy & K. Zourou (Eds.), *Social Networking for Language Education* (pp. 137–157). Basingstoke: Palgrave Macmillan.

Loewen, S., & Philp, J. (2006). Recasts in the adult English L2 classroom: Characteristics, explicitness, and effectiveness. *Modern Language Journal*, *90*(4), 536–556.

López-Burton, N. (2014). The use of technology. In N. López-Burton & D. Minor (Eds.), *On being a language teacher: A personal and practical guide to success* (pp. 236–256). New Haven and London: Yale University Press.

Lyster, R., Lightbown, P. M., & Spada, N. (1999). A response to Truscott's "What's wrong with oral grammar corrections?" *Canadian Modern Language Review*, *55*(4), 457–467.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, *19*(1), 37–66.

Macaro, E., Handley, Z., & Walter, C. (2012). A systematic review of CALL in English as a second language: Focus on primary and secondary education. *Language Teaching*, *45*(1), 1–43.

MacKey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *The Modern Language Journal*, *82*, 338–356.

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. U.S. Department of Education, Office of Planning, Evaluation, and Policy Development. SRI International. Retrieved from https://www2.ed.gov/rschstat/eval/tech/evidence-based-practices/finalreport.pdf

Nagata, N. (2002). BANZAI: An application of natural language processing to web-based language learning. *CALICO Journal*, *19*(3), 583–599.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. Philadelphia, PA: John Benjamins Publishing Company.

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Philadelphia, PA: John Benjamins Publishing Company.

PR Newswire. (2016, May 25). Global report predicts EdTech spend to reach $252bn by 2020. *Cision PR Newswire*. Retrieved from https://www.prnewswire.com/ news-releases/global-report-predicts-edtech-spend-to-reach-252bn-by-2020-580765301.html

Prensky, M. (2005). Shaping tech for the classroom: 21st-century schools need 21st-century technology. *Edutopia: What works in education*. George Lucas Foundation. Retrieved from http://www.edutopia.org/adopt-and-adapt-shaping-tech-for-classroom

Reber, A., Allen, R., & Regan, S. (1985). Syntactic learning and judgements: Still unconscious and still abstract. *Journal of Experimental Psychology: General*, *114*, 17–24.

Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 133–164). Philadelphia, PA: John Benjamins Publishing Company.

Salaberry, M. R. (2005). The use of technology for second language learning and teaching: A retrospective. In Y. Zhao (Ed.), *Research in technology and second language learning: Developments and directions* (pp. 39–56). Greenwich, CT: Information Age Publishing.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129–158.

Schulz, R. (2001). Cultural differences in student and teacher perceptions concerning the role of grammar instruction and corrective feedback. *The Modern Language Journal*, 244–258.

Schwienhorst, K. (2011). *Learner autonomy and CALL environments*. New York, NY: Routledge.

Solares, M. E. (2014). Textbooks, tasks, and technology: An action research study in a textbook-bound EFL context. In M. Gozalez-Lloret & L. Ortega (Eds.). *Technology-mediated TBLT: Researching technology and tasks* (pp. 79–113). Philadelphia, PA: John Benjamins Publishing Company.

Statista. (2018). Number of available apps in the Apple App Store from July 2008 to January 2017. *Statista: The statistics portal*. Retrieved from https://www.statista. com/statistics/263795/number-of-available-apps-in-the-apple-app-store/

Stevenson, M. P., & Liu, M. (2010). Learning a language with Web 2.0: Exploring the use of social networking features of foreign language learning websites. *CALICO Journal*, *27*(2), 233–259.

Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence*. New York, NY: Knopf.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*, 231–270.

Ullman, M. T. (2015). The declarative/procedural model: A neurobiologically motivated theory of first and second language. In B. VanPatten & J. Williams, (Eds.), *Theories in second language acquisition: An introduction*. (2nd ed.) (pp. 135–158). New York, NY: Routledge.

Ullman, M. T. (2016). The declarative/procedural model: A neurobiological model of language learning, knowledge, and use. In G. Hickok & S. L. Small (Eds.), *Neurobiology of language* (pp. 954–968). San Diego, CA: Elsevier.

Ullman, M. T., & Lovelett, J. T. (2018). Implications of the declarative/procedural model for improving second language learning: The role of memory enhancement techniques. *Second Language Research*, *34*(1), 39–65.

Ullman, M. T., & Morgan-Short, K. (2012). The neurocognition of second language. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 282–300). New York, NY: Routledge.

Vesselinov, R., & Grego, J. (2012). *Duolingo effectiveness study: Final report*. Retrieved from https://s3.amazonaws.com/duolingo-papers/other/vesselinov-grego.duolingo12.pdf

Wang, S., & Vasquez, C. (2012). Web 2.0 and second language learning: What does the research tell us? *CALICO Journal*, *29*(3), 412–429.

Ware, P. D., & O'Dowd, R. (2008). Peer feedback on language form in telecollaboration. *Language Learning and Technology*, *12*(1), 43–63. Retrieved from http://llt.msu.edu/vol12num1/wareodowd/default.html

Weber, K., Christiansen, M. H., Petersson, K. M., Indefrey, P., & Hagoort, P. (2016). fMRI syntactic and lexical repetition effects reveal the initial stages of learning a new language. *Journal of Neuroscience*, *36*(26), 6872–6880. DOI: https://doi.org/10.1523/JNEUROSCI.3180-15.2016

Zhao, Y. (2003). Recent developments in technology and language learning: A literature review and meta-analysis. *CALICO Journal*, *21*(1), 7–27.