

The Effect of Automated Adaptive Corrective Feedback: L2 English questions

Dmitri Leontjev, University of Jyväskylä

The research on the amount and the types of corrective feedback beneficial for learning a second or foreign language has produced inconsistent results. Interestingly, studying corrective feedback from the perspective of a sociocultural theory of learning has the potential to resolve these differences although so far, these studies have been largely qualitative. The present study attempts to contribute to the existing research on corrective feedback from this perspective by comparing the effects of two types of automated corrective feedback on learning: adaptive feedback (i.e., feedback incrementally adapting to learners' abilities by becoming more explicit and detailed) and knowledge of response feedback. The participants were learners of English randomly assigned to two groups, receiving either adaptive feedback (experimental group) or knowledge of response feedback (control group). The aim was to establish whether adaptive corrective feedback had a positive effect on learning, the target being L2 (second or foreign language) English questions. The findings indicate a significantly higher positive effect of the adaptive corrective feedback. Furthermore, the experimental group considered the feedback to be significantly more useful for learning than the control group although there was not a clear difference between the two groups' perceived usefulness of the feedback for getting the answers right during the intervention. It is argued that adaptive corrective feedback can raise learners' awareness of their mistakes, and it is suggested that it can facilitate individualised approach to learners. Further research is suggested.

Keywords: feedback, testing/assessment, second language (L2) learning, sociocultural theory, computer-assisted language learning

1 Introduction

It has been generally assumed that corrective feedback plays an important role in learning a second or foreign language (e.g., Bitchener 2008; Carroll & Swain 1993; Ferris 1995). At the same time, there is much less consensus as to the type and the amount of corrective feedback, both on written and spoken performance, that is more beneficial for learning (e.g., Ellis 2009; Pica 1994). This is especially the case with studies comparing the effect of explicit (i.e., overt corrective feedback) with that of implicit feedback (i.e., feedback that does not overtly state that the performance is incorrect). Hence, while some studies (e.g., Ellis et al.

Corresponding author's email: dmitri.leontjev@jyu.fi

ISSN: 1457-9863

Publisher: Centre for Applied Language Studies, University of Jyväskylä

© 2014: The author

<http://apples.jyu.fi>

2006; Nassaji 2009) demonstrated the superiority of explicit corrective feedback, others (e.g., Iwashita 2003; Kang 2009) did not find any clear difference between the two kinds of feedback. Furthermore, it should not be forgotten that there are researchers who challenge the effectiveness of corrective feedback. Truscott (1996, 1999), for example, claimed that the evidence for the beneficial effect of correction had been inconsistent and suggested that corrective feedback can be detrimental for language learning, especially if it is provided regardless of learners' developmental readiness to understand their mistakes (1996: 344).

Interestingly, studies considering corrective feedback from the perspective of a sociocultural theory of learning (e.g., Aljaafreh & Lantolf 1994; Nassaji & Swain 2000) can potentially resolve these differences. These studies build on the Vygotskian concept of Zone of Proximal Development (ZPD), formulated as "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (Vygotsky 1978: 86). In other words, according to the theory, learning is a result of collaboration between the tutor and the learner within the latter's ZPD, which involves graded support, known as mediation, provided by the tutor.

In the present study, I will refer to corrective feedback provided within learners' ZPD as to *adaptive corrective feedback*, the latter defined by Vasilyeva et al. (2007: 11) as feedback dynamically adjusting to users' abilities, characteristics, and/or performance. The reason for not using the term *mediation* is because the latter can include, but is not limited to, different forms of corrective feedback (e.g., Ableeva 2010; Poehner 2008). I will use the attributive *static* to refer to feedback/assessment not considering learners' ZPD.

However, studies looking into adaptive feedback/mediation in L2 teaching and learning are not numerous and often are predominantly descriptive. The study of Aljaafreh and Lantolf (1994) serves as an excellent example of presenting the process of negotiating corrective feedback in learners' ZPD until it matches their abilities. Tracing the influence of corrective feedback on three learners' L2 development, the authors designed a Regulatory Scale consisting of thirteen feedback messages gradually becoming more explicit and detailed. Importantly, they demonstrated that any feedback can be useful if it is provided within a learner's ZPD. While the contribution of the study is undoubted, the study design was largely descriptive.

Nassaji and Swain (2000) addressed this limitation, conducting a quasi-experimental case study of two L1 Korean learners of English, the first of whom was given adaptive corrective feedback in response to her mistakes in the use of the English articles, and the other given random feedback that did not take her ZPD into account. Having collected and analysed both qualitative and quantitative data, the authors concluded that the adaptive feedback was more beneficial when compared with the feedback that disregarded the learner's ZPD. They also found that in the case where the feedback was provided in a random manner, explicit feedback was more helpful. Yet, as this was a pilot study with only two participants, their results lack generalizability.

Adapting the amount of assistance to learners' abilities also lies in the core of dynamic assessment, which is based on the concept of ZPD and combines assessment and instruction into a single process. The key difference of dynamic assessment from its static counterparts is that during the former, the learners are provided with different kinds of mediation helping them to perform beyond the

level they would be able to while working independently (Leung 2007; Poehner 2008). Dynamic testing/assessment seems to be a reasonable basis for accumulating empirical data on the effect of adaptive corrective feedback, as it allows for collecting experimental data by means of validated instruments. However, as with most of the research adopting the sociocultural paradigm, there appears to be a lack of quantitative studies in the field of dynamic testing/assessment (see section 2.1 for a discussion).

The lack of experimental evidence about the effect of adaptive corrective feedback is understandable considering the qualitative tradition in the sociocultural research, which conventionally aims at interpreting development rather than measuring it. On the other hand, studies confirming the positive effect of adaptive corrective feedback experimentally could strengthen the argument for its usefulness. Moreover, such studies have the potential to alleviate some of the criticism, especially directed towards dynamic assessment (see e.g., Poehner 2008 for a discussion). What is more, as regards classroom instruction, procedures allowing to trace the development of learners as a group could be helpful for language teachers, for example, for finding out whether a certain structure that they have been teaching is within most of their learners' ZPD.

The present study seeks to add to the body of research on corrective feedback from a sociocultural perspective by finding out whether adaptive corrective feedback provided during a computer-based dynamic test is more effective for learning than static implicit feedback (see section 3.1 for the specific research questions). On the basis of the previous (mostly qualitative) research, I could tentatively hypothesise that L2 English learners receiving adaptive feedback are more likely to develop their L2 ability than learners receiving static (implicit) corrective feedback. While recognising the value of qualitative analyses that dominate these studies, in the present study, I will place the emphasis on experimental evidence for the beneficial effect of corrective feedback provided within learners' ZPD. An obstacle for collecting such data has been the impracticality of assessing a number of learners in face-to-face sessions (which is a common way adaptive feedback / mediation is provided to learners); yet, a recent advancement in dynamic assessment addresses this issue. I will discuss this (and other research relevant to the study) in some detail in the section to follow. I will then describe the present study, introduce the data analyses, and report on the findings. I will also suggest further research to reinforce the findings of the study.

2 Background

In this section, I will present a review of the research on computerised dynamic assessment, learners' preferences and perceived usefulness of corrective feedback (which, I will argue, is important to take into account in computerised dynamic assessment), and the development of L2 English questions (which were selected as the target of the intervention).

2.1 *Dynamic Assessment*

There are two major approaches to dynamic assessment: interventionist approach and interactionist approach. The difference between them lies in the way mediation is provided during these two types of assessment. During the former, the mediation is standardised and is given in a predefined order, often in the form of corrective feedback ranging from implicit to explicit types. In the latter approach, the required mediation emerges during the interaction between the learner and the examiner (Poehner 2008).

There have also been several successful attempts at creating computerised dynamic tests where mediation is provided automatically. The drawbacks of computerised delivery include the impossibility of establishing how learners would respond if other mediation was provided (Poehner 2008: 177) and the difficulty of tracing learners' reciprocity to mediation (see Poehner (2005) for a discussion of the latter). Its advantages, however, which include the possibility of assessing a large number of learners simultaneously, (re-) assessing the learners under uniformed conditions, and generating learners' performance reports automatically, make computerised dynamic assessment an interesting research tool.

However, not many implementations of computerised dynamic assessment have been reported in the literature. The rare examples include a computerised version of Guthke and Beckman's (2000) Leipzig Learning Test, a test for diagnosing children's learning problems, and Teo's (2012) computer-based dynamic test of learners' metacognitive reading strategies. As regards L2 computerised dynamic assessment, there seems to be only one computer-based dynamic assessment system that addresses learners' problems with L2 grammar and only to the extent it is required for listening and reading comprehension (Ableeva 2010, 2012).

These tests are designed following the interventionist approach to dynamic assessment, which is close to psychometrically oriented non-dynamic tests. This approach, especially the sandwich test format, in which treatment is conducted between an unmediated pretest and a posttest (Poehner 2008) and which, consequently, favours experimental research designs, seems to be promising for the purpose of collecting evidence on the effect of adaptive corrective feedback.

However, there are not many studies on the influence of mediation in computerised dynamic assessment that are supported with quantitative data. In Teo's (2012) study mentioned earlier, the learners' abilities before and after the intervention were compared statistically, but the author did not contrast the effect of adaptive with that of static corrective feedback. Ableeva (2010) also conducted several quantitative analyses of her data, which revealed the positive effect of the mediation. Other than that, the reports have been largely descriptive.

2.2 *Learners' Perspective on Corrective Feedback*

Constructing learners' ZPD is a dialogical activity. Thus, learners' reciprocity to mediation is an integral part of the sociocultural perspective on development. In his study, Poehner (2005) designed a Learner Reciprocity Typology – a scale in which he arranged the learners' reciprocal moves from being unresponsive to mediation due to being other-regulated to incorporating it to rejecting it due to

being fully self-regulated, which, he claimed, also reflected learners' development.

Nevertheless, it seems that learners' expectations of corrective feedback can also influence their responsiveness to and, ultimately, the usefulness of the latter. It has been found that while learners generally consider corrective feedback useful, especially feedback on their lexical, structural, and grammatical errors (Amrhein & Nassaji 2010; Hyland 2001; Leki 1991), teachers' practices, including feedback, may not be effective if they do not meet learners' expectations and preferences (e.g., Schulz 2001).

Speaking of the findings regarding learners' preferences of corrective feedback, they are somewhat varied. Amrhein and Nassaji (2010) found that both high-achieving and low-achieving learners are in favour of more explicit feedback types whereas teachers generally prefer more implicit feedback. Hyland (2001), on the other hand, points out that some learners also acknowledge the usefulness of implicit feedback for developing their language skills. However, by and large, the research demonstrates that if feedback is focused on grammatical and structural errors, then learners are generally in favour of more explicit corrective feedback (Ashwell 2000; Leki 1991). Amrhein and Nassaji (2010: 116) note that by doing so learners, especially high-achieving ones, make their lives easier, placing the responsibility of correcting their mistakes on teachers.

There is, thus, a possibility that learners can attribute different meanings to feedback usefulness—usefulness for learning and usefulness for getting the correct answers effortlessly. More importantly, this suggests that learners' rejection of feedback might not always be the manifestation of their abilities but also root in their preferences of corrective feedback. The latter is especially important for computerised dynamic assessment, where it is hard to trace learners' responsiveness to mediation.

2.3 Stages of Acquisition and Corrective Feedback

Alternatively, learners' development can be seen from a different perspective—as stages of acquisition. The stages in question development identified in the context of Pienemann's Processability Theory (Pienemann 2005) can serve as an illustration of this perspective (Table 1)¹.

Table 1. Stages in question development (adapted from Pienemann 2005; Spada & Lightbown 1999)

Stage 1	Single words, phrases: <i>How are you?</i>
Stage 2	SVO: <i>The tea is hot?</i>
Stage 3	Fronting: Do: <i>*Do he work? Does he work?</i> Wh-: <i>*Where the station is?</i> Other: <i>*Is the boy is beside the bus?</i>
Stage 4	Inversion: Yes/No: <i>Has he seen you? *Have he seen it?</i> Pseudo Inversion: <i>Where is John?</i>
Stage 5	Do/Aux 2nd: <i>Why did he sell that car?</i>
Stage 6	Cancel Inversion: <i>I wonder where he has gone?</i>

According to this theory, a learner cannot, for example, move to stage 3 of question development before stage 2 questions have emerged in his/her interlanguage, and learners move through the same developmental stages regardless of their L1. Yet, one reservation should be made. This order refers to oral production. Alanen and Kalaja (2010), who studied the L2 English performance of 250 L1 Finnish grade 7-9 learners as a part of the CEFLING project (www.jyu.fi/cefling), found the same stages in writing. However, while learners tend to use more questions at higher stages as their proficiency grows (Alanen & Kalaja 2010), it seems that they do not adhere to the developmental stages as rigidly as in spoken language (e.g., Spada & Lightbown 1999).

A number of studies have also demonstrated that corrective feedback can influence the way learners use L2 English questions (e.g., McDonough 2005; White et al. 1991), especially if an opportunity for production of modified output is provided. This makes L2 English questions an interesting treatment target in studies comparing the effects of different kinds of corrective feedback.

3 Methodology

3.1 Research Questions

The present study adds to the existing research on corrective feedback by examining the adaptive corrective feedback provided automatically in a web-based assessment/tutoring system, with the goal of establishing its effect and its perceived usefulness. Specifically, the study aims at finding answers to the following questions:

- Do L2 English learners receiving adaptive corrective feedback improve their ability to form questions significantly more than learners receiving knowledge of response feedback?
- Do learners receiving adaptive corrective feedback consider it more beneficial than learners receiving knowledge of response feedback a) for getting their answers right and b) for learning?

3.2 Design

To answer the research questions, a randomised pretest/posttest control group study was conducted. L2 English questions were found suitable to serve as the content of the exercises for the following reasons:

- feedback is found to influence the rate of their acquisition;
- learners generally consider feedback on grammar useful;
- the incremental development of questions allowed for tracing changes in the participants' performance in a more exact and a meaningful way;
- the stages in the development of L2 English questions seem to be the same regardless of learners' mother tongue.

To single out the typical errors the learners made, I examined Alanen and Kalaja's (2010) data. The analysis revealed a number of typical errors the learners made when formulating stage 5 questions, i.e., wh-questions with auxiliaries (see Table 1). Thus, I was able to focus the content of the exercises to stage 5 questions only. Nevertheless, to be able to trace the learners'

development more clearly, it was decided to include several items eliciting the use of stage 4 questions (e.g., *___you also ___talking parrots?*) into the pre-/posttest exercises (see section 3.3).

The independent variable in the study was the group the learners belonged to, either the experimental group (receiving the adaptive corrective feedback) or the control group (receiving the knowledge of response feedback). The number of stage 5 (and stage 4) questions correctly formed during the pre-/posttest and the learners' self-reports regarding the perceived usefulness of the feedback were the dependent variables.

3.3 Materials

The exercises in the pre-/posttest and the intervention were based on the imaginary situation where the learners received an E-mail from a pet shop, got interested in it, and decided to buy a puppy. It was expected that doing so would make the exercises resemble a real problem-solving communicative activity, thus adding to the authenticity of the exercises (see Bachman & Palmer 1996). In addition, it allowed for contextualising the sentences with pronouns as subjects in the exercises. Two exercises were designed for the pre-/posttest (Figure 1).

Ты хочешь узнать:

- 1) местонахождение магазина
- 2) часы открытия
- 3) каких животных они продают
- 4) сколько стоят животные
- 5) где ты можешь найти фотографии животных
- 6) какую еще информацию о животных может отправить тебе магазин
- 7) как они получили твой адрес электронной почты
- 8) что означает название магазина

Напиши свои вопросы в этой рамке:

Отправить письмо

What parrots _____ your shop _____? (продавать)
 _____ you also _____ talking parrots? (иметь)
 When _____ you _____ selling parrots? (начать – предложение в прошедшем времени)
 How long _____ the parrots _____? (жить)
 When _____ they _____ to talk? (учиться)
 How fast _____ a parrot _____? (мочь летать)
 How much _____ it _____ every day? (есть/съесть)
 What words _____ they _____? (мочь сказать)
 Where _____ the shop _____ the parrots from? (покупать – предложение в прошедшем времени)

Figure 1. Pre-/posttest exercises (see Appendix 1 for a translation of the prompts)

The first exercise was writing an E-mail according to the prompts (provided in the learners' L1). It was selected as it was one of the task types used to collect the CEFLING project data (Alanen & Kalaja 2010). Six out of eight prompts elicited the production of stage 5 questions and two prompts, either stage 5 or stage 4 questions. The second exercise was a gap filling exercise in which each item had two gaps, one after the question word and the other after the subject.

The exercise contained nine items, one eliciting the use of stage 4 and eight eliciting the use of stage 5 questions.

The intervention exercises, which targeted the use of stage 5 questions only, were the following (the sample items presented in Figure 2):

- two ordering exercises to assess the learners' problems with the word order in stage 5 questions – the first with pronouns and the second with nouns as subjects (as Spada and Lightbown (1999) found that the former were easier to produce than the latter), and
- three ordered multiple-choice exercises (pronouns as subjects) aiming to discover the learners' problems with the use of auxiliaries *do*, *does*, and *did* and the use of the correct forms of lexical verbs in stage 5 questions.

Figure 2. Intervention exercises: example items

In total, there were five exercises designed for the intervention, seven items in each (Appendix 2).

The presentation of the items and the feedback to the learners was designed in the following way, similar for all the intervention exercises in both groups:

1. an item was presented to a learner;
2. following the learner's response, feedback was displayed to him/her;
3. the learner was then presented with the next item, which had the same structure as the previous item.

There was, thus, a difference between the adaptive feedback (mediation) used in the present study (see Table 3) and the way mediation is commonly provided in dynamic assessment, i.e., learners go back to the same item until they are able to self-correct or are provided with the correct answer. The reason for doing so was primarily to make the learners realise that the pattern of stage 5 questions is the same/similar with different question words, lexical verbs, and auxiliaries.

The experimental group feedback was designed to follow the implicit-to-explicit adaptation similar to Aljaafreh and Lantolf's (1994) Regulatory Scale and

looked as follows, the numbers indicating the levels of the feedback progression from implicit “think more carefully” to explicit explanation and overt correction (Table 2):

Table 2. Adaptive corrective feedback in the study

Level	Description	Example
0.	An indication that the response is correct	Your sentence: When does he come to work? Correct!
1.	An implicit hint that there might be something wrong with the answer	Your sentence: When did it appeared in your shop? Think more carefully. Try to complete the next question—it will be similar to this one.
2.	The location of the error is narrowed down	<i>Your sentence:</i> How long does it sleeps in the shop? Look at the highlighted part of your sentence. Think, is everything correct there? The following question will be similar to this one.
3.	The location of the error is further narrowed down, the nature of the error is identified, and metalinguistic clues or elicitations are provided	Your sentence: How often do you're clean the shop? You used the correct helping word do . But do we need the verb are here? The following question will be similar to this one.
4.	Examples of the correct structure are given	Your sentence: How many times must eat the puppy every day? Not quite right. Look at the following examples: How are they different from your sentence? How could you do that? What might you answer him? Where could he go ? The following question will be similar to this one.
5.	The correct response is provided with the explicit indication of what was wrong	Your sentence: When you're took the picture of the puppy? Sorry, you need did before the word you ; the verb are is not needed; and you had to use take instead of took . The correct answer is:

For the control group, the simple knowledge of response feedback was designed, i.e., the indication of whether their performance on the items was correct or not.

To administer the exercises, a web-based system called ICAnDoiT (Interactive Computer-Adaptive Diagnostic and Tutoring system) was designed. It served as a tool providing learners with instantaneous corrective feedback gradually attuning to their abilities. Additionally, it allowed for recording of the learners'

performance, including the mistakes they made, the feedback they received, etc. The ICAnDoiT system was created as a part of my on-going Ph.D. research and is currently hosted at <https://solki4.cals.jyu.fi/icandoit/htdocs/>. The usefulness of the system is that it allows language tests to be compiled using a variety of predefined task types with the possibility of adding feedback (both dynamic and static) to learners' item-by-item performance. In the following, the current state of validation process of the system will be outlined. A full account of the validation process will be given in a future report.

The exercises were piloted among 19 L1 Finnish learners of English (grade 8, average 14 years of age) in December 2010. The aim of the pilot study to establish the validity of the procedure. Additionally, the questionnaire used in the present study was piloted. As the major aim of the pilot study was to pilot the exercises, the study did not include the posttest and no control group was assembled.

The piloting resulted in a number of changes, such as modification/addition of several items in the pre-/posttest exercises. The feedback messages were also slightly modified to stress the similarity between the items. The pilot study also confirmed that the exercises elicited the production of wh-questions.

To reinforce the usability of the system, the system interface was designed according to the blueprint provided by Fulcher (2003). This was followed by a three-phase usability check, which used questionnaire replies, think-aloud protocols, and interviews as data. All in all, the usability study allowed for eliminating several usability problems, such as the difficulty to understand the mechanics of the ordering exercises.

A more comprehensive account of the piloting will be given in a future paper.

3.4 Participants and Data

The participants in this study were L1 Russian learners of English, average 14 years of age, studying at grade 8 in a school in Estonia ($n = 64$). The learners were from six different groups taught by two teachers. Each learner was randomly assigned to either the experimental ($n = 35$) or the control ($n = 29$) group.

However, the reported numbers refer to those who completed the intervention exercises. Since some learners were missing during the pretest, others during the posttest, and some cheated (as observed by either me or the teachers monitoring their performance), there were fewer learners whose performance on the exercises was analysed – 26 and 21 learners respectively. As regards cheating, it was an extraneous variable that could introduce construct-irrelevant variance. Therefore, I decided to remove the performance of the learners who cheated from the analyses.

In Estonia (and in Finland), learners' first foreign language proficiency is expected to be at level B1.2 by the end of grade 9 (the end of lower-secondary school). Judging by the descriptors (Põhikooli riiklik õppekava õigusakt: Lisa 1 2010 [Basic School National Curriculum Act: Annex 1]), by the end of grade nine, learners are expected to ask wh-questions (e.g., when asking for directions). This reinforced the possibility that wh-questions should be within some of the participants' ZPD. Moreover, before the intervention, I asked the teachers whether by the time of the study, the learners had been taught to form questions in English (including wh-questions with auxiliaries), which they confirmed.

Judging by the teachers' reports and the state curriculum, I assumed that these questions were at least in some of the learners' ZPD.

The data come from the learners' performance on the exercises they took in the ICAnDoiT system. Additionally, the learners completed an online questionnaire (Appendix 3) which aimed at discovering their experiences with the feedback during the intervention. The questionnaire was conducted in the learners' mother tongue.

3.5 Procedure and Scoring

Before the pretest, it was explained to the learners that they were to complete several exercises so that they could see how well they were able to form questions in English. The learners were also advised to consult the help menu or ask for help from the persons monitoring their performance if they did not know any of the words in the exercises. They were given help only on vocabulary, not grammar. To save time, in the first exercise of the pretest, the learners were instructed to write the questions only. All the learners were working on the same exercises. The only difference was in the feedback the two groups received during the intervention.

The following sample from a learner's performance log demonstrates how the feedback incrementally adapted to the experimental group learners' abilities, gradually becoming more explicit and detailed (Table 3). The feedback this learner and the others received was originally in their L1 and was translated into English for the present article.

The log illustrates that the learner had a certain consistent error in *wh*-questions with the auxiliary *does*. After being provided with metalinguistic clues regarding the nature of his mistake, he was able to select the correct option and use it consistently until the end of the exercise.

After completing the intervention exercises, the learners were asked to fill in the questionnaire. Two academic periods were allocated for this part of the study. The procedure was not speeded, as all the learners managed in less time than that.

The posttest was conducted a week after the intervention and was the same as the pretest. One academic period was allocated for the posttest, but the learners completed the exercises in less time than that. Those who were missing during the pretest and the intervention were asked to complete the intervention exercises and the questionnaire while the others were working on the posttest. Two scoring schemes were used to score the learners' responses to the pre-/posttest items. In the first scoring scheme, I gave each correctly formed stage 5 question one point. In the second, I awarded each correct stage 5 question two points and each correct stage 4 question, one point. It was decided to use a composite score (i.e., the sum of the learners' scores on the two pretest/posttest exercises) in the statistical analyses for the following reasons:

- both exercises assessed the same construct;
- this allowed for estimating the learners' abilities more precisely.

Table 3. A learner’s performance on the *does*-exercise

Selected option	Feedback
Where do it plays in the shop?	Your sentence: Where do it plays in the shop? Think more carefully! Try to complete the next question—it will be similar to this one.
Why do it looks sad in the photo?	Your sentence: Why do it looks sad in the photo? Look at the highlighted part of your question. Think, is everything correct there? The following question will be similar to this one.
When do it goes to sleep?	Your sentence: When do it goes to sleep? You used the correct helping verb. But think about the word it . What do you have to add to the helping verb do ? What should happen to the verb goes ? The following question will be similar to this one.
When does it close on holidays?	Correct!
What does it like to eat?	Correct!
When does he come to work?	Correct!
How long does it sleep at night?	Correct!

4 Results

The exercises were designed based on the performance of the Finnish learners of English and were also piloted among them. Thus, ensuring the comparability of the pilot study group with the present study participants was necessary for reinforcing the construct validity of the exercises for the present study group.

For comparing the present study and the pilot study participants’ performance, an independent samples t-test was conducted on the square-root transformed variable (percent correct on the two pretest exercises). It demonstrated that the performance of the present study participants ($M = 4.01$, $SD = 2.85$, $n = 47$) was not statistically different from the pilot study participants’ performance ($M = 3.77$, $SD = 2.34$, $n = 19$), $t(64) = 0.32$, $p = .748$. Moreover, the present study learners made similar mistakes as the Finnish learners had made in the exercises, so the designed exercises (including the distractors in the multiple-choice exercises) and the feedback addressed their problems equally well.

This was followed by a modern item analysis of the present study participants’ pretest performance conducted using *Winsteps* Rasch analysis software. It showed that there were no outfitting items in both the scoring that only took into account stage 5 questions ($0.59 \leq \text{infit MNSQ} \leq 1.4$) and the partial credit scoring ($0.55 \leq \text{infit MNSQ} \leq 1.36$). The person separation statistics of the two variables were 1.4 (Cronbach’s alpha .84) and 1.49 (Cronbach’s alpha .86)

respectively, which is satisfactory (e.g., Fisher 2007). In other words, taken together, the pretest exercises could distinguish between high (or rather middle) and low performers.

Most of the following statistical analyses were conducted using IBM SPSS software. The results are presented in two sections, the first comparing the performance of the two groups and the second, the experiences of the two groups with the feedback in the study. Exact statistics will be provided whenever possible.

4.1 The Effect of the Adaptive Feedback as Contrasted with the Knowledge of Response Feedback

To establish whether the adaptive feedback had any effect on the learners' ability to produce stage 5 questions, the differences in the learners' scores on the pretest and the posttest were compared. The descriptive statistics for the pretest and the posttest scores are reported in Table 4 and illustrated in Figure 3 below.

Table 4. Learners' pre-/post-test performance: descriptive statistics

Groups	Pre-test			Post-test		
	Mean	SD	Median	Mean	SD	Median
Experimental ($n=26$)	4	4.14	2.5	5.35	4.17	4.5
Experimental, partial credit ($n=26$)	8.81	8.93	5.5	11.92	8.87	10
Control ($n=21$)	4.19	3.26	4	3.9	2.81	3
Control, partial credit ($n=21$)	9.38	7	9	9	8	6

However, before studying the changes in the learners' performance after the treatment, I decided to reinforce the condition that the two groups' ability to form L2 English questions did not differ significantly before the treatment. Due to the verisimilitude of the figures obtained on the two scoring schemes, only the results on the partial credit scores are reported.

As the dependent variable was not normally distributed, a Mann-Whitney U test was used. It demonstrated that the experimental group learners ($Mdn = 5.5$) did not perform significantly differently from the control group learners ($Mdn = 9$), $Z = -.59$, $p = .561$. To corroborate the finding, a differential item functioning analysis was conducted. It confirmed that the learners in both groups performed similarly on all of the items, the highest *Welch's t* value being for item 11 (the second exercise), $t(33) = 1.49$, $p = .15$.

To establish whether the difference in performance between the two groups was statistically significant, I conducted an independent-samples t-test on the gain scores variables (the difference between the posttest and the pretest scores), which were normally distributed (e.g., for the partial credit scoring, $W(26) = .973$, $p = .696$ for the experimental group and $W(21) = .953$, $p = .386$ for the control group).

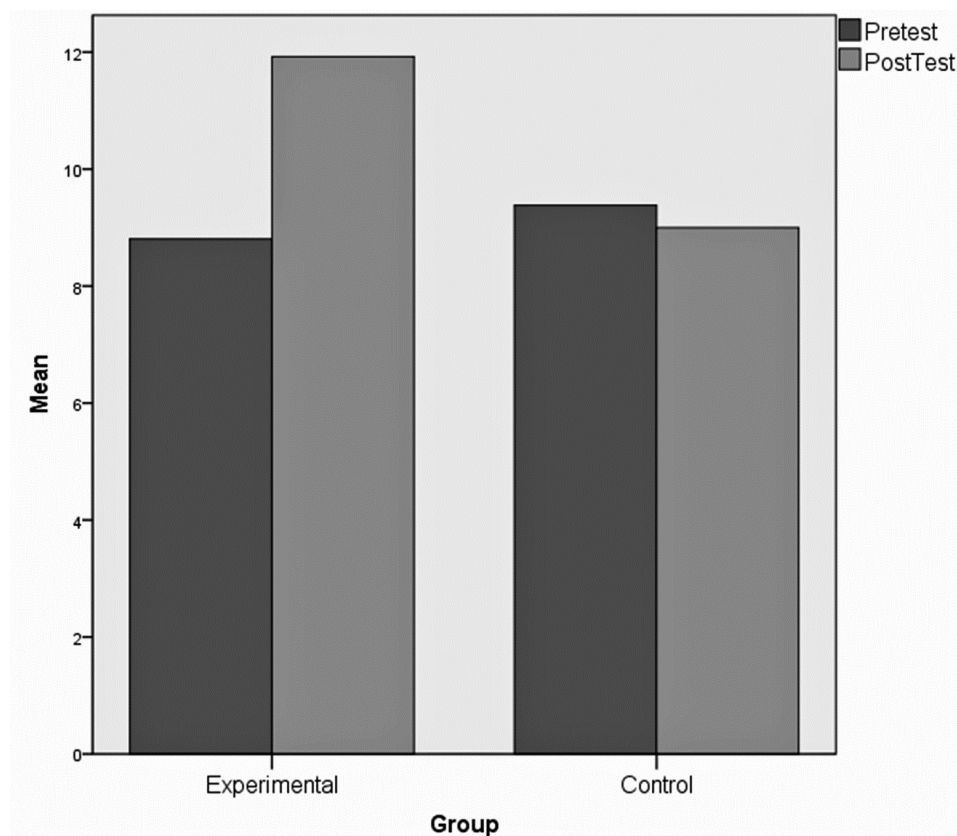


Figure 3. Learners' performance on the pretest and the posttest (partial credit scoring)

The t-test demonstrated that the experimental group ($M = 1.35$, $SD = 2.64$) increased the number of correctly formed wh-questions with auxiliaries significantly more than the control group ($M = -0.28$, $SD = 2.47$), $t(45) = 2.17$, $p = .035$), where there was, in fact, a small decrease. The mean difference of 1.63 scale units, 95% CI [0.12, 3.15] indicated a moderate effect size, $d = 0.64$. The difference between the experimental group ($M = 3.12$, $SD = 5.48$) and the control group ($M = -.38$, $SD = 5.95$) was slightly less pronounced for the partial credit scoring scheme, $t(45) = 2.09$, $p = .042$. The mean difference of 3.50 scale units, 95% CI [.13, 6.86], indicated a moderate effect as well, $d = 0.61^2$.

However, the significant difference between the two groups does not imply that the knowledge of response feedback was useless for all the control group learners. The following Table 5 illustrates the opposite.

It can be deduced from the learner's performance log that he was not sure at first about the word order in the wh-questions with modal auxiliaries. By the end of the exercise though, he was able to use the correct word order more consistently. Moreover, during the posttest, he formed all questions with the model auxiliary correctly as compared to half of the questions of the same kind formed correctly during the pretest.

The experimental group learner's performance log presented in Table 3 reveals a somewhat different picture in that he appeared to be more other-regulated than the control group learner whose performance is illustrated in Table 5. Interestingly, during the pretest, the experimental group learner was not able to form any stage 5 questions with *does*. During the post-test, however,

he formulated two wh-questions with *does* correctly. I will elaborate on these two examples in the **discussion** section.

Table 5. A control group learner's performance on the first ordering exercise

Selected option	Feedback
WHERE / MUST / IT / SLEEP / IN THE FLAT?	Correct!
CAN / I / TEACH / MY DOG? / WHAT TRICKS	Your question is formed incorrectly.
WHY / MUST / YOU / CLOSE / THE SHOP EARLY?	Correct!
CAN / I / BUY / WHAT TOYS / FOR MY PET?	Your question is formed incorrectly.
WHEN / CAN / I / GET / MORE PHOTOS?	Correct!
WHERE / CAN / I / LEARN / MORE ABOUT DOGS?	Correct!
HOW LONG / CAN / IT / STAY / ALONE?	Correct!

What is interesting to note about the learners' performance is that the intervention seems to have also resulted in a somewhat higher number of level 4 questions (both correct and incorrect) produced by the learners. A closer look at the learners' performance reveals a rather interesting trend. Three out of four experimental group learners who failed to produce any questions higher than stage 3 (e.g., **What animals shop sells?*) during the pretest produced at least one stage 4 question (e.g., *Where's the shop located?*) during the posttest. One of those three also managed to produce three stage 5 questions. The fourth learner produced four correct stage 5 questions but no stage 4 questions during the posttest. It is hard to say to what extent knowledge of response feedback can facilitate the same development, as there was only one control group learner who produced one question at stage 4 and one at stage 5 during the posttest while having failed to produce any questions at these stages during the pretest. Not much can be said about the same trend in formulating stage 5 questions, as three experimental group learners out of six who failed to form any stage 5 questions during the pretest formed at least one (either correct or incorrect or both) during the posttest and two out of three control group learners were able to do the same.

4.2 Learners' Self-Reports

Twenty-eight experimental group and twenty-three control group learners completed the questionnaire. To compare the two groups' self-reports, their responses to one Likert-scale and two dichotomously scored items were analysed (see Appendix 3). The Likert-scale item asked the learners to rate the extent to which the feedback helped them to find the correct answers during the intervention. The first dichotomous item asked them whether they had learned anything having completed the intervention exercises. The second dichotomous item asked them whether the feedback had helped them to learn it.

A Mann-Whitney U test demonstrated that the experimental group ($Mdn = 3.5$) did not rate the usefulness of the feedback for completing the intervention exercises differently from the control group ($Mdn = 3$), $Z = -0.59$, $p = .963$. Moreover, although a higher proportion of the experimental group learners (64%) thought that they had learned something compared with the control group (48%), the difference was not statistically significant either, as demonstrated by a Chi-square test, $X^2(1, n = 51) = 1.40$, $p = .238$.

On the other hand, 14 learners from the experimental group (50%) answered positively when asked whether it was the feedback that had helped them to learn something, whereas only five learners from the control group (about 21%) were of the same opinion. A Chi-square test indicated that the difference was statistically significant, $X^2(1, n = 51) = 4.31$, $p = .038$, $\phi = -.29$.

To interpret these results, I also looked at the learners' responses to the open-ended questions in the questionnaire. The qualitative analysis of the responses revealed some recurring patterns exemplified in Table 6.

Table 6. Learners' reported reasons for the feedback usefulness

Experimental	Control
<ul style="list-style-type: none"> • They showed me when I can use <i>can</i>. • They helped me by giving examples. 	<ul style="list-style-type: none"> • It showed that the answer was incorrect. • I realised I was doing right and continued.
<ul style="list-style-type: none"> • They hinted that the word was in the wrong place. • I didn't remember the rule, and the feedback helped me to. • Everything was explained: why the sentence was incorrect and how to correct it. • Because I understood my mistake. 	<ul style="list-style-type: none"> • I don't know. • I DON'T KNOW. • They were of no use. I often didn't even look at them.

5 Discussion

One of the aims of the present study was to determine whether the adaptive corrective feedback was more likely to facilitate learning than the try-again feedback (provided irrespective of the learners' ZPD). The findings demonstrate that the feedback adapting to the learners' abilities resulted in a significant increase in the number of correctly formed wh-questions with auxiliaries in the experimental group as compared with the control group who received the static implicit feedback group (where, in fact, there was a small decrease). There was at least a short term moderate positive effect of the adaptive feedback. The findings, therefore, confirm the hypothesis that adaptive feedback provided automatically can facilitate learning. This adds to the findings of the earlier studies regarding the influence of corrective feedback negotiated within learners' ZPD.

The analysis of the performance of those learners who failed to produce any stage 4 and stage 5 questions during the pretest revealed that after the intervention, stage 4 questions emerged in their performance. Certainly, from the point of view of Processability Theory, the emergence of stage 4, and not stage 5, questions indicated that some learners were simply not ready to

advance to the latter higher level of question development. The intervention, however, was not designed to facilitate the development of stage 4 questions. Therefore, this issue deserves further examination, the more so as this part of the analysis looked at a very limited number of cases.

There is also some indication that in the control group, the learners' improvement in many cases might have to do with the increase in accuracy rather than the emergence of the correct structure(s) in their unassisted performance. Stage 4 questions in the learners' performance, which I discussed in the previous paragraph, can serve as an example of that. Another example can be the qualitative difference between the pre-/posttest performances of the two learners whose treatment performance is described in Tables 3 and 5. The evidence for that, however, is rather inconsistent. A future study can explore this possibility.

Importantly, the figures also reveal that implicit feedback, favoured by some teachers according to Amrhein and Nassaji (2010), is not always facilitative for learning. The implicit feedback not being helpful could be explained by the finding of Nassaji and Swain (2000), who discovered that the learner given the feedback irrespective of her ZPD was more likely to benefit from more explicit feedback. Thus, it would be interesting to conduct a similar study where the control group received explicit feedback (e.g., explicit correction and/or explicit explanation of the error) to compare the effect of adaptive feedback with that of explicit corrective feedback.

Alternatively, it could have been the learners' preferences for different feedback types that resulted in a higher acceptance of the adaptive feedback. This could have added to the facilitative effect of the adaptive feedback in the experimental group and hindered the usefulness of the feedback in the control group. It is also worth noting that the control group learners, even when considering the feedback helpful, were often unsure of the reason(s) for that. Thus, it seems that feedback adapted to learners' abilities might be accepted more readily than static implicit feedback.

On the other hand, the results demonstrated that the experimental group learners did not consider the feedback any more useful for getting their answers right during the treatment than the control group (probably because it did not give away the correct answers in most of the cases). Thus, it seems that learners do indeed attach different meanings to the word *usefulness*. More importantly, there is a possibility that the learners' perceived usefulness of the feedback could have negatively influenced the utility of certain feedback types which otherwise matched their abilities. That is to say, some learners skipped the feedback messages they *considered* useless and not because those feedback messages did not match their abilities. However, the data in the present study do not allow for drawing any conclusions in this regard. This would also be an interesting question to address in a further study.

The above interpretation does not mean that teachers should avoid giving implicit feedback to their learners—doing so would deprive learners of an important step on their way of becoming self-regulated in the use of a second/foreign language. On the contrary, the performance of some learners (including some of the control group learners) demonstrated that they did not need explicit and detailed feedback to self-correct during the treatment and increase their scores on the posttest exercises. Amrhein and Nassaji (2010) rightfully note that learner autonomy is one goal of pedagogy, and by preferring

explicit correction, learners may unnecessarily place the responsibility of correcting their mistakes onto teachers, which contradicts this goal. Rather, from the perspective of a sociocultural theory of learning, the findings should be interpreted so that adapting the feedback to the learners' ZPD was beneficial to a larger number of learners than providing the static feedback that disregarded the learners' ZPD.

6 Conclusion

The present study aimed at finding out whether adaptive corrective feedback had a facilitative effect on learning (in this case, L2 English questions), and whether this effect was significantly different from that of the knowledge of response feedback. Additionally, it compared the self-reports of the two groups of learners on the perceived usefulness of the feedback.

The study demonstrated that the learners who had received adaptive corrective feedback during the intervention produced significantly more correctly formed L2 English wh-questions with auxiliaries than the control group. The learners also tended to accept the adaptive feedback as useful for learning more readily than the knowledge of response feedback. The latter, however, might have also derived from the learners' preference for more explicit feedback types as the previous research suggests.

The findings of the study have several implications. Adaptive corrective feedback provided to learners while they practice on a second/foreign language should allow them to self-diagnose their problems as well as to learn something. The finding that the adaptive feedback helped the learners to become aware of their mistakes and produce more correct responses during the posttest suggests that a similar procedure has implications for teaching. Learner profiles, similar to the one presented in the study (Table 3), would allow teachers to see the typical mistakes their learners make but also help them with the difficult task of finding out how much help their learners currently need with certain mistakes. Additionally, as I have suggested at the beginning of the paper, teachers would be able to see whether the required structure is within (most of) their learners' ZPD or more teaching is required.

Amrhein and Nassaji (2010) suggest that teachers should change their learners' feedback preferences if these preferences are not beneficial for their learners. One way the assessment/tutoring system used in this study, or a similar one, could help teachers achieve this goal is that they could discuss the performance profiles with their learners, so that the latter would see how implicit feedback had helped them. What is more, the experience of automated adaptive feedback might influence learners' beliefs about the efficacy of different feedback types without teachers having to follow it up with discussions, which would save teachers time and effort. Whether this experience alone or followed with discussions could change learners' preferences of corrective feedback seems to be an interesting topic to explore.

There are, however, several limitations to the study that might affect the generalizability of its results. Despite the decent overall number of participants, the fact that not everyone completed the pretest, the posttest, and the questionnaire resulted in a smaller number of cases in the analyses and might have affected the findings. Moreover, the pretest and the posttest contained only

two exercises (17 items in total). Finally, due to the school schedule, a delayed posttest could not be conducted. Therefore, it is impossible to tell whether the adaptive feedback led to a long-lasting learning effect. At the same time, the posttest was conducted a week after the intervention, so the learning effect lasted for at least a week.

A similar study with a larger number of participants, more exercises/items in the pre-/posttest as well as with a delayed posttest could reinforce the findings of the present study. Additionally, further studies comparing adaptive corrective feedback with other types of corrective feedback, such as explicit correction or random feedback, should allow for creating a more comprehensive picture demonstrating whether corrective feedback negotiated within learners' ZPD is indeed superior to static corrective feedback. The no-feedback condition for the control group might also be used to address Truscott's (1996) claim about the negative effect of corrective feedback.

Nevertheless, despite the limitations of the study, it is hoped that it has provided useful insights into the applications of adaptive corrective feedback (that is to say, mediation) its effect on learning, and its usefulness as perceived by learners. I also hope that the study stimulates research on the effect of corrective feedback as seen from a sociocultural perspective. Collecting more experimental data would enable meta-analyses of the effectiveness of adaptive corrective feedback, thus strengthening the argument for its usefulness.

Endnotes

1. There are apparent epistemological differences between the paradigms underlying the concept of universal developmental stages and the sociocultural perspective on development. Specifically, while the former presupposes a uniform order of acquisition and, consequently, that instruction can only be effective when learners are developmentally ready to advance, according to the latter it is instruction that directs the development to follow, and there are, in effect, no prescribed developmental stages (e.g., Leung 2007). Resolving these differences, however, is beyond the scope of the present paper.
2. The shape of the distribution in the control group was slightly not symmetric. Thus, I supplemented the analysis with a Mann-Whitney U test, which showed that the difference between the gain scores on the stage 5 questions only scoring was statistically significant, $Z = 2.04$, $p = .040$, $r = .30$. That is to say, it confirmed the result obtained on the t-test as far as the stage 5 questions only (which were the target of the intervention) were considered. The difference in the gain scores obtained on the partial credit scoring was not significant, $Z = -1.85$, $p = .06$. What is more, Wilcoxon signed-rank tests demonstrated that the improvement after the treatment was significant in the experimental group, e.g., for the partial credit scoring, $Z = -2.65$, $p = .007$, $r = .37$, but not in the control group, $Z = -.26$, $p = .805$.

References

- Ableeva, R. 2010. *Dynamic assessment of listening comprehension in second language learning*. Unpublished doctoral dissertation, Pennsylvania State University, University Park, PA. [Retrieved June 12, 2013]. Available at <http://etda.libraries.psu.edu/paper/11063/6495>.
- Ableeva, R. 2012. *Transfer Tasks: Diagnosing Second Language Development*. Paper presented at AAAL 2012 Annual Conference, Boston, MA, USA.
- Alanen, R., & P. Kalaja 2010. *The emergence of L2 English questions across CEFR proficiency levels*. Paper presented at AAAL 2010, Atlanta, Georgia, USA.
- Aljaafreh, A. & J. P. Lantolf 1994. Negative feedback as regulation and second language learning in the Zone of Proximal Development. *The Modern Language Journal*, 78(4), 465–483.
- Amrhein, H. R. & H. Nassaji 2010. Written corrective feedback: What do students and teachers think is right and why? *Canadian Journal of Applied Linguistics*, 13(2), 95–127.
- Ashwell, T. 2000. Patterns of teacher response to student writing in a multiple-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing*, 9(3), 227–257.
- Bachman, L. F., & A. S. Palmer 1996. *Language testing in practice*. Oxford: Oxford University Press.
- Bitchener, J. 2008. Evidence in support of written corrective feedback. *Journal of Second Language Writing*, 17(2), 102–118.
- Carroll, S. & M. Swain 1993. Explicit and implicit negative feedback. *Studies in Second Language Acquisition*, 15(3), 357–386.
- Ellis, R. 2009. A typology of written corrective feedback types. *ELT Journal*, 63(2), 97–107.
- Ellis, R., S. Loewen, & R. Erlam 2006. Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, 28(2), 339–368.
- Ferris, D. R. 1995. Student Reactions to Teacher Response in Multiple-Draft Composition Classrooms. *TESOL Quarterly*, 29(1), 33–53.
- Fisher, W. P., Jr. (2007) Rating Scale Instrument Quality Criteria. *Rasch Measurement Transactions*, 21(1), 1095.
- Fulcher, G. 2003. Interface design in computer-based language testing. *Language Testing*, 20(4), 384–408.
- Guthke, J. & J. F. Beckmann 2000. The learning test concept and its application in practice. In C. S. Lidz & J. G. Elliott (Eds.), *Dynamic assessment: Prevailing models and applications* (pp. 17–69). Amsterdam: JAI/Elsevier.
- Hyland, F. 2001. Providing Effective Support: Investigating Feedback to Distance Language Learners. *Open Learning: The Journal of Open, Distance and E-Learning*, 16(3), 233–247.
- Iwashita, N. 2003. Negative Feedback and Positive Evidence in Task-Based Interaction: Differential Effects on L2 Development. *Studies in Second Language Acquisition*, 25(1), 1–36.
- Kang, H-S. 2009. The relative efficacy of explicit and implicit feedback in the learning of a less-commonly-taught foreign language. *IRAL - International Review of Applied Linguistics in Language Teaching*, 47(3-4), 303–324.
- Leki, I. 1991. The preferences of ESL students for error correction in college-level writing classes. *Foreign Language Annals*, 24(3), 203–218.
- Leung, C. 2007. Dynamic assessment: Assessment for and as teaching? *Language Assessment Quarterly*, 4(3), 257–278.
- McDonough, K. (2005). Identifying the impact of negative feedback and learners' responses on ESL question development. *Studies in Second Language Acquisition*, 27(1), 79–103.
- Nassaji, H. 2009. Effects of recasts and elicitations in dyadic interaction and the role of feedback explicitness. *Language Learning*, 59(2), 411–452.

- Nassaji, H. & M. Swain 2000. A Vygotskian perspective on corrective feedback in L2: The effect of random versus negotiated help on the learning of English articles. *Language Awareness*, 9(1), 34-51.
- Pica, T. 1994. Questions from the language classroom: Research perspectives. *TESOL Quarterly*, 28(1), 49-79.
- Pienemann, M. 2005. An introduction to Processability Theory. In M. Pienemann (Ed.), *Cross-linguistic aspects of Processability Theory*. Amsterdam: John Benjamins Publishing Company. 1-60.
- Poehner, M. E. 2005. *Dynamic assessment of oral proficiency among advanced L2 learners of French*. Unpublished doctoral dissertation, Pennsylvania State University, University Park, PA.
- Poehner, M. E. 2008. *Dynamic assessment: A Vygotskian approach to understanding and promoting L2 development*. Berlin: Springer.
- Põhikooli riiklik õppekava õigusakt: Lisa 1. 2010. Pub. L. No. RT I 2010, 6, 22. [Retrieved May 27, 2014]. Available at <https://www.riigiteataja.ee/aktalisa/1281/2201/0017/13275423.pdf>.
- Schulz, R. A. 2001. Cultural differences in student and teacher perceptions concerning the role of grammar instruction and corrective feedback: USA-Colombia. *The Modern Language Journal*, 85(2), 244-258.
- Spada, N. & P. M. Lightbown 1999. Instruction, first language influence, and developmental readiness in second language acquisition. *The Modern Language Journal*, 83(1), 1-22.
- Teo, A. 2012. Promoting EFL students' inferential reading skills through computerized dynamic assessment. *Language Learning & Technology*, 16(3), 10-20. [Retrieved March 10, 2013]. Available at <http://www.llt.msu.edu/issues/october2012/action.pdf>.
- Truscott, J. 1996. The case against grammar correction in L2 writing classes. *Language Learning*, 46(2), 327-369.
- Truscott, J. 1999. The case for "The case against grammar correction in L2 writing classes": A response to Ferris. *Journal of Second Language Writing*, 8(2), 111-122.
- Vasilyeva, E., S. Puuronen, M. Pechenizkiy, & P. Rasanen 2007. Feedback adaptation in Web-based learning systems. *International Journal of Continuing Engineering Education and Life-Long Learning*, 17(4-5), 337-357.
- Vygotsky, L. S. 1978. *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- White, L., N. Spada, P. M. Lightbown, & L. Ranta 1991. Input enhancement and L2 question formation. *Applied Linguistics*, 12(4), 416-432.

Appendix 1. The pretest/posttest exercises (the prompts translated into English)

Exercise 1

You are interested in:

- 1) location of the shop
- 2) opening hours
- 3) what pets they sell
- 4) how much the pets cost
- 5) where you can find the pets' photos
- 6) what other information about the pets the shop can send you
- 7) how they got your E-mail address
- 8) what the name of the shop means

Exercise 2*

- 1) What parrots _____ the shop _____? (to sell)
- 2) _____ you also _____ talking parrots? (to have)
- 3) When _____ you _____ selling parrots? (to begin / to start – the sentence is in the past tense)
- 4) How long _____ the parrots _____ ? (to live)
- 5) When _____ they _____ to talk? (to learn)
- 6) How fast _____ a parrot _____ ? (can fly)
- 7) How much _____ it _____ every day? (to eat)
- 8) What words _____ they _____ ? (can say)
- 9) Where _____ the shop _____ the parrots from? (to buy – the sentence is in the past tense)

* The task was preceded by the instructions where the learner was asked to imagine that his/her grandfather wanted to buy a parrot and asked the learner to forward his questions to the pet shop.

Appendix 2. The intervention exercises

Task 1*

1. WHEN/CAN/I/GET/MORE PHOTOS?
2. HOW LONG/CAN/IT/STAY/ALONE?
3. WHERE/MUST/IT/SLEEP/IN THE FLAT?
4. WHAT TOYS/CAN/I/BUY/FOR MY PET?
5. WHERE/CAN/I/LEARN/MORE ABOUT DOGS?
6. WHY/MUST/YOU/CLOSE/THE SHOP EARLY?
7. WHAT TRICKS/CAN/I/TEACH/MY DOG?

Task 2*

1. WHAT ELSE/MUST/MY FAMILY/KNOW/ABOUT DOGS?
2. WHERE/CAN/MY FATHER/PARK/NEAR THE SHOP?
3. WHAT/CAN/THE PUPPY/DO/IN MY FLAT?
4. HOW/CAN/MY GRANDPA/TEACH/PARROTS TO TALK?
5. HOW MANY TIMES/MUST/THE PUPPY/EAT/EVERY DAY?
6. WHEN/CAN/PUPPIES/GO/OUTSIDE?
7. WHY/MUST/PARROTS/LIVE/IN A CAGE?

Task 3**

1. How often [do you clean] the shop?
 - a. do you clean
 - b. do you're clean
 - c. you are clean
 - d. are you clean
 - e. you clean
2. What else [do you sell] in your shop?
3. How [do I choose] the dog food?
4. What [do you feed] the puppies?
5. When [do I take] the puppy to the doctor?
6. Why [do you leave] the pets alone at night?
7. How often [do I wash] my puppy?

Task 4**

1. When [does it close] on holidays?
 - a. does it close
 - b. do it closes
 - c. does it closes
 - d. do it close
 - e. it closes
2. How long [does it sleep] at night?
3. What [does it like] to eat?
4. When [does it go] to sleep?
5. When [does he come] to work?
6. Where [does it play] in the shop?
7. Why [does it look] sad in the photo?

Task 5**

1. Why [did I get] only one E-mail?
 - a. did I get
 - b. I was get
 - c. did I'm get
 - d. I'm got
 - e. did I got
2. How [did you find] my E-mail address?
3. When [did you take] the picture of the puppy?
4. How many puppies [did you sell] last month?
5. Why [did he open] a pet shop?
6. When [did it appear] in your shop?
7. Where [did it live] before the pet shop?

*The order in which sentence parts, as separated with a “/”, were displayed to the learners was randomised every time each item was retrieved from the item bank; the parts were never displayed in the correct order.

**The options, as presented for item 1, had the same structure in every item; the order of the options was randomised every time each item was retrieved from the item bank. The correct option is provided in the square brackets.

Appendix 3. Questionnaire items discussed in the study (English translation)

Please tell us how useful the hints were for you (how well they helped you to do the exercises). Choose only one option:

1. very useful (they helped me a lot)
2. quite useful (they helped me quite a lot)
3. not really useful but not useless either (they helped me a little)
4. quite useless (they did not help me much)
5. useless (they were of no help to me)

Did you learn anything after completing the exercises?

yes no

Please tell us what you learned:

Do you think the hints you received helped you to learn?

yes no

Please tell us how exactly the hints helped you to learn:

Received October 17, 2013
Revision received June 10, 2014
Accepted June 13, 2014