Learning new grammatical structures in Task-Based Language Learning: The effects of recasts and prompts

van de Guchte, M.; Braaksma, M.; Rijlaarsdam, G.; Bimmel, P.

DOI
10.1111/modl.12211

Publication date
2015

Document Version
Final published version

Published in
The Modern Language Journal

The Modern Language Journal

Learning New Grammatical Structures in Task-Based Language Learning: The Effects of Recasts and Prompts

MARRIT VAN DE GUCHTE
University of Amsterdam
Research Institute of Child Development and Education
Amsterdam Roeterseilandcampus, Building D
Nieuwe Achtergracht 127
Amsterdam 1018 WS
The Netherlands
Email: vandeguchte@uva.nl

MARTINE BRAAKSMA
University of Amsterdam
Research Institute of Child Development and Education
Amsterdam Roeterseilandcampus, Building D
Nieuwe Achtergracht 127
Amsterdam 1018 WS
The Netherlands
Email: braaksma@uva.nl

GERT RIJLAARSDAM
University of Amsterdam
Research Institute of Child Development and Education
Amsterdam Roeterseilandcampus, Building D
Nieuwe Achtergracht 127
Amsterdam 1018 WS
The Netherlands
Email: G.C.W.Rijlaarsdam@uva.nl

PETER BIMMEL
Couperusweg 6
1217 TA Hilversum
The Netherlands
Email: pbimmel@xs4all.nl

In the present study, we examine the effects of prompts and recasts on the acquisition of two new and different grammar structures in a task-based learning environment. Sixty-four 14-year-old 9th grade students (low intermediate) learning German as a foreign language were randomly assigned to three conditions: two experimental groups (one receiving prompts, the other recasts) and a control group. The study involves two subsequent interventions: The first targeted a complex structure, dative case after a preposition; the second a simple structure, comparatives. Pretests, immediate posttests, and delayed posttests included written and oral accuracy as well as oral fluency. Statistical comparisons on both written and oral posttests showed that prompts and recasts were effective, when compared to the control group, with prompts being superior to recasts. Furthermore, the effect of recasts depended on the structure: Recasts were more effective for the comparative than for the dative on written accuracy.

Keywords: task-based language learning; complex versus simple structure; recasts; prompts; trade-off

OVER THE LAST THREE DECADES, RESEARchers as well as teachers have shown much interest in task-based language learning. There is steadily growing empirical evidence that task-based language learning promotes language acquisition through different kinds of processes such as negotiation of meaning, uptake of corrective feedback, noticing the gap between incorrect performance and correct use of the target structure, metalinguistic reflection, and automatization (see Robinson's [2011] review article on task-based language learning). In an increasing number of classrooms, the classic PPP model

The Modern Language Journal, 99, 2, (2015)
DOI: 10.1111/modl.12211
0026-7902/15/246–262 $1.50/0

(presentation–practice–production) makes room for learning languages by means of tasks. Such tasks primarily focus on meaning. In this process, rich and authentic input (Krashen, 1981) asks students to produce output (Swain, 2000, 2005) and promotes interaction (Long, 1996; Long & Robinson, 1998), thus stimulating language acquisition.

A central issue in task-based language learning, however, is whether input, output, and interaction are sufficient to acquire language (Skehan et al., 2012). Several researchers have expressed serious doubts (Doughty & Varela, 1998; Long, 1991; Skehan, 1996). They agree that tasks should initially focus on meaning, but suggest that in addition there should also be attention to form.

One way to focus on form is to provide the student with corrective feedback (CF). In recent years, the effects of implicit as well as explicit CF have been studied extensively (for a review, see Lyster, Saito, & Sato, 2013; Schoorman & Schlak, 2012, and meta-analyses by Li, 2010; Lyster & Saito, 2010; Mackey & Goo, 2007; Russell & Spada, 2006). Although these meta-analyses show that, in general, CF is beneficial to the acquisition of the target language, the issue remains: Which form of CF is most effective under which conditions and for which linguistic structures?

Most studies comparing the effects of recasts and prompts (e.g., Ammar & Spada, 2006; Ellis, Loewen, & Erlam, 2006; Loewen & Nabei, 2007; Lyster & Izquierdo, 2009; Yang & Lyster, 2010) have reported on the acquisition of linguistic features with which the students were already (partially) familiar. To our knowledge, no classroom studies on new linguistic features, which also differ in complexity and relatedness to the L1, exist. It is for this reason that we designed the present experiment. We investigated the effects of recasts and prompts on two new and different grammar structures in a task-based language learning environment.

CORRECTIVE FEEDBACK

One way to stimulate attention to formal structures during meaningful communication is the provision of corrective feedback (CF), defined by Lightbown and Spada (1999) as "any indication to the student that their use of the target structure is incorrect" (p. 171). How CF contributes to language acquisition can be explained by both skill acquisition theories and theories on implicit learning. In addition, we will discuss Skehan's (1996, 1998) dual mode system on information processing. Anderson (1993, 2000) considers the acquisition of a language as learning any other skill, with declarative knowledge turning gradually into procedural knowledge. In the case of language acquisition, this could mean that students memorize, for example, a grammar rule, and through practice (proceduralization of knowledge) they are able to use the rule without thinking about it. Both Lyster (2004) and DeKeyser (2010) support practice of declarative knowledge through CF because of its ability to promote restructuring of the interlanguage. CF as a way to practice, however, can only be effective if embedded in communicative interaction and when students' individual differences and the teaching context are taken into account (DeKeyser, 2010).

Ellis (1993), a supporter of theories of implicit learning, sees a more limited role for explicit knowledge. He points out that explicit knowledge as provided by CF may facilitate implicit learning in only two ways. First, CF may contribute to the process of noticing: "[I]f students are armed with explicit knowledge of a linguistic feature, they are more likely to notice its occurrence in the communicative input they receive and thus to learn it implicitly" (p. 149). Second, CF may assist noticing the gap because a student who possesses explicit knowledge of certain language structures will likely be able to notice the gap between their own incorrect and the correct target structure given by a teacher or peer.

An additional view to language acquisition is Skehan's (1996) dual mode system, which describes two ways of processing information. The *exemplar-based system* includes discrete lexical items as well as ready-made formulaic chunks of language; the *rule-based system* consists of abstract representations of the underlying patterns of the language. These systems cooperate and "combine in a synergistic manner to yield results, and degrees of learning, that are more than simply the sum of the parts" (p. 43). Drawing on this dual model system, Skehan argues that exemplar-based CF "may not be so effective" because "there is not the (. . .) connection with a rule which can produce general change" (p. 89). The rule-based system, however, "is more likely to be more sensitive to feedback since the precision and system which accounts for rule-organization will make the feedback more informative" (p. 88). Nonetheless, achieving greater accuracy is not without consequences for other dimensions of the task performance. According to Skehan's (1996, 1998) trade-off hypothesis, learners possess limited attentional resources and a focus on, for instance, accuracy, may have a negative impact on fluency and/or complexity. Robinson (2001, 2005), on the other hand, disagreeing with

this hypothesis, claims that people have multiple attentional resources, which are not in competition with each other but can be used at the same time.

Having reviewed three different theories on language processing, several key factors on how CF may promote language learning emerge: The feedback should be noticeable, create opportunities for practice (output), and trigger access to the rule-based system.

Recasts and Prompts

A key concern is the extent to which the two CF types—recasts and prompts—promote noticing (the gap), create opportunities for practice, and trigger access to the rule-based knowledge system. As defined by Lyster and Ranta (1997), recasting is "the teacher's reformulation of all or part of a student's utterance minus the error" (p. 46). Recasts are considered to be an implicit form of CF and are by far the most frequent form of negative feedback in classrooms of all kinds (Long, 2007). The second feedback type is prompting, which, according to Lyster and Ranta (1997), comes in various shapes and types: clarification requests, repetitions, metalinguistic feedback, and elicitations.

Based on L1 acquisition research that shows that children notice and use linguistic information (e.g., repeating the utterance), Doughty and Varela (2008) regard recasts as an ideal way to focus on form in the communicative classroom. In addition, it is also the ability of correcting a learner's mistake without breaking down the communicative flow (Long, 2007) that make recasts very suitable for meaning-focused task-based language learning and teaching (TBLT). The effectiveness of recasts, however, is disputed. Nicholas, Lightbown, and Spada (2001), for instance, comment that the student may not always notice the corrective element of a recast. Similarly, Lyster (1998) found that students perceived recasts as confirmation of meaning rather than feedback on form. Another point of criticism is that students may not be aware of the exact location of the error and as a consequence may not notice it. Finally, de Bot (1996) comments that no trace in memory is left by recasts, because the student is not actively involved but only listens to the input. He reasons that it is this low level of attention that generates only weak or no connections in memory, hence constraining restructuring of the interlanguage. In regard to these negative claims about recasts, Goo and Mackey (2013) recently pointed out some methodological and interpretive problems in the small number of studies on which the claims are based, including issues like form-focused instruction. They state that in some recast-versus-prompts studies, form-focused instruction has been included as a part of the experimental treatment. They doubt whether the differential effects of prompts over recasts can be fully attributed to the feedback type because of the "moderating role of the form-focused instruction" (p. 152).

The ambiguous findings on the effectiveness of recasts have led to a comparison of other forms of CF, like prompts. In the majority of classroom studies following this line of investigation, prompts appeared to be more effective than recasts (Ammar, 2008; Ammar & Spada, 2006; Ellis et al., 2006; Loewen & Philp, 2006; Yang & Lyster, 2010). Factors such as noticeability (Schmidt, 1990), the ability to generate modified output (Swain, 2000, 2005)—also considered as a way of practicing—and the metalinguistic feedback being rule-driven, may explain the superiority of prompts over recasts.

Previous Research on Corrective Feedback

An ever-growing number of classroom and laboratory studies have compared the effects of recasts and prompts on L2 acquisition. In this review, we focus on studies that compared prompts and recasts by measuring student outcomes in terms of acquisition of linguistic structures. Thus, we exclude studies that measured uptake and repair and studies that focused on the acquisition of vocabulary (Dilans, 2010).

Before comparing recasts with prompts, we briefly summarize the findings of studies that investigated the effectiveness of recasts. Several studies have reviewed their effects (Ellis & Sheen, 2006; Long, 2007; Lyster & Saito, 2010; Miller & Pan, 2012; Nicholas et al., 2001) and conclude that recasts have proven their effectiveness in laboratory settings (Carroll & Swain, 1993; Egi, 2007, 2010; Han, 2002; Leeman, 2003; Long, Inagaki, & Ortega, 1998; Mackey & Philp, 1998). However, recasts were not effective, or only to a limited extent, in classroom settings (Doughty & Varela, 1998; Ellis et al., 2006; Lyster, 2004).

In the majority of laboratory studies that compared recasts with prompts, no significant differences were found. Carroll and Swain (1993) examined the effects on the acquisition of English dative alternation of explicit correction, recasts, and two types of prompts, and compared them with data obtained in a control group with no feedback. All of the treatment groups

performed better than the control group; the group receiving explicit correction outperformed the other groups, but no significant differences were found between prompts and recasts. The same results are reported by McDonough (2007) and Lyster and Izquierdo (2009). They investigated the effects of recasts and prompts on the acquisition of, respectively, the simple past and progressive activity verbs in a Thai EFL context (Lyster & Izquierdo, 2009) and grammatical gender by adult second language students of French (McDonough, 2007). Lyster and Izquierdo concluded that "students receiving recasts benefited from the repeated exposure to positive exemplars as well as from the opportunities to infer negative evidence, whereas students receiving prompts benefited from the repeated exposure to negative evidence as well as from opportunities to produce modified output" (pp. 453–454). Nassaji (2009) compared recasts with elicitations in dyadic interactions between native speaker English teachers and adult ESL students. His study favored recasts by demonstrating that these led to a higher percentage of immediate postinteraction correction than the elicitations did. In addition, explicitness seemed to be a key factor: For both recasts and elicitations, the more explicit forms led to more immediate and delayed postinteraction correction than the implicit forms.

These classroom studies can be categorized into studies with either young or adult students. Lyster (2004) investigated the effects of form-focused instruction (FFI) in combination with corrective feedback (recasts versus prompts) on 10–11-year-old immersion students' ability to accurately assign grammatical gender in French. Results of the written tasks in particular, and to a lesser degree the oral tasks, revealed that FFI is more effective when combined with prompts than with recasts or no feedback. Ammar and Spada (2006) also focused on younger students with their study, including 12-year-old francophones as participants in an intensive ESL course. They examined the impact of recasts in comparison to prompts and no corrective feedback on students' acquisition of English third person determiners. On the oral task, both the recast and the prompt group outperformed the control group on the posttests. The delayed posttest results also revealed that the prompt group was even better than the recast group. On the written task, the prompt group outperformed the recast group on the immediate and delayed posttest. However, the investigators do not conclude that prompts are the ideal CF technique because analyses by proficiency level showed that low-performance students benefited more from prompts than recasts, whereas high-proficiency students benefited equally from both prompts and recasts. This was affirmed by Ammar's (2008) study, which was a secondary analysis of Ammar and Spada's (2006) earlier data. Comparisons of prompts and recasts showed that prompts may be more effective than recasts in leading to L2 morphosyntactic development, especially in the case of low-proficiency students. Results from the computerized task showed no differences between the groups in terms of accuracy. In addition, Mackey and Philp (1998) found that developmentally more advanced students benefit more from recasts than developmentally unready students in facilitating an increase in production of targeted higher level morphosyntactic forms.

The participants in all other classroom studies were adult students of a second or foreign language. Ellis et al. (2006) reported on the effects on the acquisition of the past tense. The participants, 77% of whom were of East Asian origin, were low intermediate students of L2 English. Implicit knowledge was measured by an oral imitation test; explicit knowledge was measured by an untimed grammaticality judgment and a metalinguistic knowledge test. Although no significant differences were found on the immediate posttests, prompts outperformed recasts on the delayed imitation and grammaticality judgment posttests. In a similar study, Ellis (2007) not only measured acquisition of the English past tense but also the acquisition of comparatives. It appeared that prompts were overall more effective than recasts, but more so for comparatives than for past tense. As a result, Ellis concluded that the effectiveness of the feedback depends on the grammatical structure. We shall address this point later in the article.

In studies by Loewen and colleagues (Loewen & Nabei, 2007; Loewen & Philp, 2006), no significant differences were found between prompts and recasts on accuracy. Nevertheless, Loewen and Philp found on the immediate posttests that prompts led to an accuracy rate of 75%, whereas recasts led to an accuracy rate of only 53%. Although prompts led to a higher success rate, Loewen and Philp found the achievements of the recast group encouraging for classroom teachers by suggesting that recasts are likely to be productive for students and have pedagogical benefits: They are considered "less threatening to students' confidence and less intrusive to the flow of communication" (p. 551). The most recent study to date, conducted by Yang and Lyster (2010), concerns a quasi-experimental

investigation held in China in EFL classrooms at university level. The study compared the effectiveness of recasts, prompts, and no feedback on the use of regular and irregular English past tense. The effects of prompts were larger than those of recasts in terms of increasing accuracy in the use of regular past tense, whereas prompts and recasts had similar effects on improving accuracy in the use of the irregular past tense.

The differences in findings among studies comparing recast and prompts in laboratory and classroom settings might lie in the fact that recasts are more salient in laboratory interactions than in the classroom. According to Nicholas et al. (2001), the laboratory setting makes the student more aware of the feedback being corrective; the fact that the feedback was restricted to only one or two features made it easier for the students to recognize what was intended by the feedback.

CF and Linguistic Structure

CF research shows that the effectiveness of CF depends heavily on the grammatical structure being investigated (Lyster et al., 2013). What works for one linguistic structure may not be effective for another (Sheen, 2011). Although there is no agreement on what is meant by *type of structure*, the differences in target structures are often explained by factors as simple/complex, easy/difficult to learn, rule-based/exemplar-based, and L1–L2 relatedness.

The current body of research into what type of instruction is beneficial to which target structures shows no agreement. Hulstijn and de Graaff (1994), for example, claim that complex rules can be taught best in an explicit way and simple rules in an implicit way. They argue that learners are not able to notice complex structures in the input and that therefore explicit learning of the rule is required. In comparison, simple rules may be noticed more easily by the learner, which may lead to acquisition. Krashen (1981) takes the opposite point of view by claiming that only simple rules can be taught and that difficult rules are best learned in an implicit way. Regarding L1–L2 transfer, Andringa, de Glopper, and Hacquebord (2011) found that explicit instruction was more effective for simple structures if these structures were more or less similarly realized as in a student's L1.

To date, only two classroom studies have been designed to compare the effects of oral prompts and recasts on different target structures. Ellis (2007) compared the effects of recasts and prompts on the acquisition of the English past tense and comparatives. He considered the comparative to be more difficult than the past tense, because the rule for the comparative requires both morphologic and syntactic analysis, while the past tense only asks for a morphologic analysis. Ellis found that prompts were overall more effective than recasts, but more so for the comparative (relatively difficult) than for the past tense (more simple). Yang and Lyster (2010) also investigated the effects of recasts and prompts on two different structures: the regular and irregular English past tense. The prompt group performed significantly better than the control group on irregular past tense forms at the delayed oral posttest. In addition, the effects of prompts were larger than those of recasts on the oral production test of regular past tense forms. However, prompts and recasts had similar effects on the oral production test of irregular past tense forms. Drawing on Skehan's (1998) dual knowledge system, they argued that prompts had more effect on regular past tense because this target structure is considered a rule-based feature (add *-ed* to the base form of a regular verb). They reasoned that prompts, more than recasts, trigger access to the rule-based system. Since there are no clear rules for irregular past tense forms, these are considered to be exemplar-based features.

So far, German language structures have received very little attention in empirical investigations of CF (Lochtmann, 2002). To enlarge the scope of linguistic structures, in relation to the effectiveness of CF, we selected two different German linguistic structures for the current study: dative case after a preposition of place and comparatives.

THE PRESENT STUDY

The current study focuses on whether previous findings on the effects of prompts and recasts on (partially) known target structures also applies to two new German grammar structures that differ in difficulty and relatedness to the L1. By investigating Skehan's (1996, 1998) trade-off hypothesis, the study also contributes to the literature on attentional resources in task-based language learning. Five research questions guided the study:

RQ1. Do recasts have a positive effect on the accuracy of new grammar structures?

RQ2. Do prompts, operationalized as metalinguistic feedback followed by elicitation, have a positive effect on the accuracy of new grammar structures?

RQ3. Which type of CF is more effective?

RQ4. Does the effectiveness depend on the targeted structure?
RQ5. Does a student's focus on accuracy have a negative effect on oral fluency?

METHOD

Participants

Sixty-four 9th-grade students learning German as a foreign language participated in this study. They were recruited from three school groups at a Dutch secondary school. The majority of the participants (96%) were native speakers of Dutch. Apart from German, participants also learn English and French as foreign languages at school. The mean age was 14.3 years. The students had been engaged in learning German for 19 months at the A2 level of the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001), for 2 hours per week, 40 weeks per year. In accordance with the protocol of the University of Amsterdam's Faculty of Humanities' Ethics Committee, all parents were informed about the studies and the possibility of nonparticipation.

Target Structures

In line with our purpose to examine whether corrective feedback facilitates the learning of a completely new structure, we chose two German target structures with which the students at A2 level of the CEFR were not already familiar. Moreover, we wanted to examine the interaction of prompts and recasts with a complex and a simple German structure that also differ in their relatedness to the L1.

Dative Case After a Preposition of Place (Dative). The target structure in Task 1 and Testing Task 1 concerns the dative case of an article after a two-way preposition in German (*in*, *an*, *auf*, *hinter*, *neben*, *unter*, *über*, *vor*, *zwischen*[1]). A two-way preposition may be used with a dative to indicate the current location; combined with an accusative it signals direction toward something or someone. In the present study, only the locative dative was used. In this example, *auf dem Bett liegt ein Kissen*,[2] the preposition *auf* demands the dative case and consequently the neuter definite article *das* changes into *dem*. We define this structure as syntactic, rule-based, and difficult for two reasons: (a) Learners have to undertake several analytical steps to arrive at the correct form (see Hulstijn & de Graaf, 1994, for complexity and difficulty of target structures); specifically, learners are required to apply a syntactic rule: The article *der* of a masculine word changes into *dem*, the article *die* of a feminine word changes into *der*, the article *das* of a neutral word changes into *dem*, and the article *die* of a plural word changes into *den*. (b) Learners have no existing knowledge regarding the dative structure because it has no equivalent in Dutch. Transfer between L1 and L2 is therefore unlikely.

Comparative. For Task 2 and Testing Task 2, comparatives were selected as the target structure. We consider this structure to be simple and basically morphological, including both rule- and exemplar-based elements, because (a) most of the comparatives are formed by simply adding *-er* to the adjective or adverb. In addition, several German irregular comparative forms get an umlaut mark (*größer* 'bigger,' *höher* 'higher') or have suppletive forms (*besser* 'better,' *mehr* 'more'); (b) the forming of comparatives in the L1 (Dutch) is very similar to the forming of German comparatives. Learners may apply the same rule, adding *-er* to the adjective or adverb, and the L1 also contains suppletive forms that show similarities to German. Adding an umlaut mark to a comparative form is not found in Dutch.

Design

The study comprises two interventions, where each consisted of a task with a focus on a particular language structure. The design entailed three testing sessions for each task. One week prior to the start of Task 1, the students performed two pretests: an oral test followed by a written test. One week after the intervention period (which lasted 3 weeks), students took posttests; 3 weeks later they took delayed posttests. Task 2 (which lasted 3 weeks) started 2 weeks after Task 1 and followed the same pattern of measurements. Two intact classes were randomly assigned to the recast ($n = 20$) and prompt ($n = 21$) conditions. The control group ($n = 23$), an intact class, followed the form-focused regular curriculum, which included written exercises on both target structures during the intervention period. These students did not receive experimental treatment. During the intervention periods, students of the experimental groups worked 1 hour a week on the treatment tasks, and during the remaining hour they read a book and practiced listening skills without any grammar instruction or feedback on grammar structures. Because the teachers were required to provide the students with approximately one or two feedback moments every week, they needed the whole hour

to do so for 20 or 21 students. Sometimes, students did not need a full hour to work on the task and continued with reading activities. Then, the task materials remained on the table, which enabled the teacher to look at the material, ask questions, and provide feedback. The control group worked 1 hour in their textbooks, including written exercises on the two target structures. During the other 60 minutes, they performed reading and listening activities.

Prompt Group. As shown in Example 1, the prompt was given by the teacher in two phases: (a) provision of metalinguistic feedback on the student's false utterance and (b) elicitation of the correct answer.

EXAMPLE 1

Prompt on Dative Case After Preposition

S: In das Mitte steht ein Stuhl. [Error: Grammatical] ['In the middle is a chair']
T: Almost, it's not in das Mitte. 'After the preposition 'in' follows the dative case. [FB: metalinguistic]
T: Okay, try again. In d——— Mitte? [FB: Elicitation]
S: In der Mitte

Note: S = student; T = teacher; FB = feedback

In order to ensure a valid comparison, none of the learner groups received instruction and practice on the new target structures (see Goo & Mackey, 2013, on instruction and prompts) prior to task performance. The students of the prompt group received information on the grammar structures during the first part of the prompts (metalinguistic feedback), enabling them to respond to the teacher's request to moderate their output (elicitation).

Recast Group. The teacher of the recast group was asked to reformulate the student's false utterance, minus the error (see Example 2).

EXAMPLE 2

Recast on Dative Case After Two-Way Preposition

S: In das Mitte steht ein Stuhl. [Error: grammatical] ['In the middle is a chair']
T: Ach so, in der Mitte steht ein Stuhl. [FB: recast]

Like the students in the prompt group, the students in the recast group were not instructed in the new structures but received information on the correctness of the target structure through the positive example of the recast.

Control Group. Because of school regulations, the control group could not perform the two oral intervention tasks, but were required to follow the standard curriculum of the course book, which was strongly focused on forms (for the distinction between focus-on-form and focus on formS, see Long, 1991). The two target structures, dative and comparative, were part of their curriculum during the period of the interventions. They did not receive personal oral feedback on these grammar structures. According to the regular practice in those language classes, students were instructed on the target structures, performed written exercises in their workbooks, and reviewed them with help from the correction model. At the end of the lesson, during a joint class moment, the teacher asked the students whether they had any questions about the exercises.

Instruction Materials and Procedures

Task 1. The students worked in pairs and designed the room of their dreams with a virtual budget of €10,000. First, the students performed pretask activities that equipped them with vocabulary useful for the upcoming task performance. The goal of the task phase for each pupil was to describe the room in triads. The two other students were not able to see the room and were asked to draw the room on paper according to the description. Afterward, they compared their drawings and posed questions on any furniture that was placed incorrectly. After finishing the task, the students presented their dream bedroom individually to a research assistant who recorded their performance on video camera. This oral presentation was subsequently assessed by the teacher using a scale ranging from 1 to 10 (low to high, with 6 judged to be sufficient).

Task 2. The students were asked to participate as researchers in a German television program for consumers entitled *Deutschland testet* ('Germany tests'). They were given the assignment to compare two products. From the input in the pretask, the students selected vocabulary that could be useful for task performance. The task phase was oriented toward an oral presentation of the comparison test that had been conducted in triads. Each student presented the comparison of his two products, and the two other group members took notes and explained which product they wanted to buy. After finishing the task, the students presented their product comparisons individually to a research assistant who recorded their performance on video camera. This oral presentation was assessed by the teacher using the same scale as in Task 1.

Procedures

The two experimental groups were taught by two different teachers because their classes were taught at the same time. The groups were randomly assigned to the teachers, who both hold master's degrees in teaching German language and literature and were trained in the feedback type they were to provide. (The teacher of the prompt group is the first author of this article.) During the training session, both teachers observed examples of possible student failures and practiced with providing either recasts or prompts.

During the pre- and main task phase, the teachers of the intervention groups walked around the classroom to provide each student with one or two feedback moments per lesson (1 hour). They looked at the students' design of their dream room (Task 1) or product comparison (Task 2), followed by questions, such as *Wo steht das Bett in deinem Zimmer*? ('Where is the bed in your room?') and *Wo hängt das Poster in deinem Zimmer*? ('Where is the poster in your room?'), regarding Task 1, and *Warum möchtest du ein iPhone kaufen? Gib mir drei Gründe* ('Why do you want to buy an iPhone? Give me three reasons'), regarding Task 2.

The control group was taught by a third teacher who also holds a master's degree in teaching German language and literature. The researcher received a copy of the lesson plans of the control group and visited the classes on a regular basis. As for the intervention groups, a colleague in the department regularly observed the lessons with the help of a feedback logbook in order to ascertain whether both teachers adhered to expectations regarding the provision of recasts and prompts and to take note of the number of feedback moments per student. The observations revealed no differences in treatment except in terms of feedback type. In each lesson, both teachers wrote down the number of feedback moments for each student. Students of the recast group received a mean of 6.3 (*SD* 4.37) recasts on the dative structure and 4.5 (*SD* 1.79) on the comparative structure, spread out over 3 weeks, that is, approximately 2.1 recasts on the dative and 1.5 recasts on the comparative per hour. Students of the prompt group received a mean of 5.33 (*SD* 2.11) prompts on the dative structure and 5.30 (*SD* .85) on the comparative structure, in other words, 1.78 prompts on the dative and 1.77 on the comparative, per hour. Because students had fewer errors in the comparative than in the dative structure, they received more feedback on the dative. No significant differences in the number of feedback moments were found between conditions on either the dative ($F[1,39] = .83$, $p = .37$) or the comparative structure ($F[1,39] = 3.71$, $p = .06$).

Testing Materials and Procedures

For the written tests, a fill-in-the-gap test was chosen in order to measure monitored language of the German grammar structures; by comparison, the communicative oral testing tasks were intended to measure the students' implicit knowledge (Ellis, 2005) and oral fluency. In order to avoid the retrieval of explicit knowledge on the structures, students performed the oral task first, followed by the written task. For the oral test on comparatives, no delayed posttests were administered in the control group because the students were not motivated to perform them. We chose not to force them because of possible negative influence on the reliability of the results.

Written Accuracy Tests. Following DeKeyser (1993) a fill-in-the-gap test was chosen to assess accuracy. Not only was this format most familiar to the students; it also forced them to provide a correct alternative rather than simply indicating the occurrence of an incorrect structure, a limitation of many grammaticality judgment tests. The written tests were administered by the teacher during the lessons. In both settings, students were given sufficient time, which turned out to be 15 minutes maximum. The test was administered in three versions: For the pretest, each participant received version A, for the posttest, version B, and for the delayed posttest, version C, thus all conditions received the same test at each measurement occasion. All written tests for the dative and comparative structures were piloted, prior to the intervention, in two 10th-grade classes ($n = 48$) that did not take part in the intervention study. Cronbach's alpha scores varied between .67 and .79, suggesting reasonable reliability. Finally, the different versions of the tests were randomly assigned to a measurement occasion.

The written test on the dative included 14 fill-in-the-gap sentences where students were to supply the correct case of an article after a two-way preposition that was used with its meaning of static location. In order to disentangle application of the dative rule from knowledge of noun gender, grammatical gender was noted in brackets after the noun (*Tisch [der]*) ['table (the)']. Reliability (Cronbach's alpha) was .85 for the pretest, .95 for the immediate posttest, and .93 for the delayed posttest.

The written test on the comparative included eight fill-in-the-gap sentences where students were asked to provide the correct comparative form. Reliability (Cronbach's alpha) was .65 for the pretest, .81 for the posttest, and .70 for the delayed posttest.

Oral Accuracy Tests. Oral accuracy in both tasks was assessed during a postintervention classroom period by means of an oral communication test held with one of the three interlocutors: two native speakers of German and the first author. The tests consisted of communicative tasks similar to those used in the treatment sessions. Once again, three parallel test versions were designed. The interlocutors, located in different rooms, had received instruction regarding test administration and followed a fixed protocol to ensure identical test administration. Each student was tested individually and audiorecorded, and, at each measurement occasion, student was tested by another interlocutor.

The recordings were rated for accuracy by a native speaker of German and the first author. The first author evaluated all recordings; the second rater evaluated 70%. For both tasks, interrater reliability was high: 95% for the dative and 90% for the comparative.

Knowledge of the dative was tested with a picture description test (Appendix A). The interlocutor used a 3-dimensional picture of a bedroom; the German noun plus its gender of the furniture and accessories in the picture were provided. Not only would this approach focus on students' grammatical over their lexical knowledge; it would also facilitate the occurrence in their production of the target structures. The interlocutors were asked to elicit a minimum of 15 target structures. Accuracy scores were expressed as proportions correct (correct use of dative case after a two-way preposition/total number of articles used after a two-way preposition). Use of prepositions other than the target structures was disregarded. Assessment focused on the repairs, not on the utterance before the repair.

For the comparison test, the interlocutor used a picture of two mobile phones that included a table indicating price, dimensions, weight, memory, internet speed, touchscreen resolution, and battery size for each mobile phone (Appendix B). Once again, the German noun forms were provided. Students were asked to compare nine features of the mobile phones. Accuracy scores were expressed as proportions of correct forms (correct use of comparatives/total number of comparatives used). Incorrect use was indicated when either the comparative form itself was incorrect or when it was used incorrectly in context. When a student repaired, the repair was assessed and not the utterance before the repair.

Oral Fluency Test. Fluency in the oral tests was evaluated by two raters, a native speaker of German and the first author. Raters were instructed to pay attention to students' speech rate and pausing. Speech rate and pausing were chosen as measures of fluency; on these measures, strong associations have been found between objective measures and subjective ratings on fluency (see DeJong et al., 2013). In order to achieve rating consistency, the raters listened to 15 preselected performances for each task at different levels in the scale. The fluency rating scale ranged from 1 (very low fluency) to 5 (very high fluency). Discrepancies were discussed until agreement was reached. The order of recordings was randomized and each sample was rated independently by each rater. The second rater evaluated 70% of the recordings. Interrater reliability was high in both studies, 91% in the first experiment and 90% in the second.

Analyses

The effects of the two variables (i.e., target structure and feedback type) were investigated through an immediate and a delayed posttest. Because of the correlational nature of the data (i.e., two types of feedback, two target structures, two posttests) a general linear mixed model (GLMM) statistical analysis was used that included fixed and random factors for each dependent measure. Test time (pretest = M0, posttest 1 = M1, posttest 2 = M2), type of feedback, and target structure were entered in the model as fixed factors; the random factor was subjects. In addition, the model included three 2-way interactions and one 3-way interaction. An alpha level of .05 was used for all statistical tests.

RESULTS

Written Accuracy Test

Table 1 presents the means and standard deviations resulting from the written accuracy tests. The results for written accuracy showed a main effect, for (a) feedback type $F(2,61) = 61.77$, $p < .001$, (b) time $F(2,304) = 64.81$, $p < .001$, and (c) structure $F(1,304) = 353.16$, $p < .001$. Pairwise comparisons showed that the recast group and prompt group outperformed the control group

TABLE 1
Group Means (Proportions Correct) and Standard Deviations for Written Accuracy Tests

		M0		M1		M2	
	Group	*M*	*SD*	*M*	*SD*	*M*	*SD*
Dative	Recasts ($n = 20$)	.029	.071	.086	.120	.068	.112
	Prompts ($n = 21$)	.065	.218	.724	.318	.503	.346
	Control ($n = 23$)	.093	.090	.093	.102	.090	.104
Comparative	Recasts ($n = 20$)	.412	.177	.519	.136	.473	.136
	Prompts ($n = 21$)	.423	.156	.851	.140	.732	.037
	Control ($n = 23$)	.310	.154	.301	.148	.293	.187

(recasts vs. control, $MD = .068$, 95% confidence interval [CI] = 0.14, $p = .05$; prompts vs. control, $MD = .35$, 95% CI = .29, .42, $p < .001$). The prompt group outperformed the recast group ($MD = .29$, 95% CI = .22, .35, $p < .001$). The main effect for feedback type is moderated by target structure and time. The analyses revealed a significant feedback type/structure interaction $F(2,304) = 16.21$, $p < .001$, feedback type/time interaction $F(4,304) = 44.83$, $p < .001$, and feedback type/structure/time interaction $F(4,304) = 2.61$, $p = .036$. The last interaction indicates that the effect of recasts and prompts differed according to structure over time (see Figure 1).

With regard to the effect of prompts on the dative, we see a very large increase between M0 and M1 ($d = 2.42$) and a decrease after M1 ($d = -.67$). For the effect of prompts on the comparative, we also see a very large increase between M0 and M1 ($d = 2.89$) and a decrease after M1 ($d = -1.16$).

With regard to the effects of recasts on the dative, there is an increase between M0 to M1 ($d = .58$) and a small decrease after M1 ($d = -.16$). For the comparative, there is also an increase between M0 and M1 ($d = .68$) and a decrease after M1 ($d = -.34$). The interaction effect shows that, compared to the prompts, recasts had a significantly larger effect on the comparative than on the dative $t(304) = -3{,}126$, $p = .002$.

Oral Accuracy Test

Table 2 displays the means and standard deviations of the oral accuracy tests. The results for oral accuracy showed main effects for (a) feedback type $F(2,66) = 27.51$, $p < .001$, (b) time $F(2,284) = 31.05$, $p < .001$, and (c) structure $F(1,284) = 210.99$, $p < .001$. Pairwise comparisons showed that the recast group and prompt group outperformed the control group (recasts vs. control, $MD = .097$, 95% confidence interval [CI] = .023, .17, $p = .011$; prompts vs. control, $MD = .302$, 95% CI = .23, .38, $p < .001$). The prompt group also outperformed the recast group ($MD = .205$, 95% CI = .13, .28, $p < .001$). The main effect for feedback type is moderated by time. A significant feedback type/time interaction ($F[4,284] = 15.4$, $p < .001$) indicates that the effect of prompts and recasts differed over time. Reviewing the effect of prompts

FIGURE 1
Interaction Effect Target Structure/Feedback Type/Time for Written Accuracy

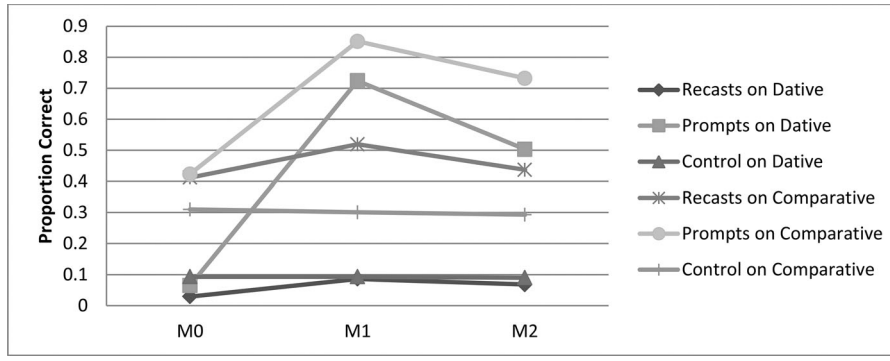


TABLE 2
Group Means (Proportions Correct) and Standard Deviations for Oral Accuracy Tests

		M0		M1		M2	
	Group	*M*	*SD*	*M*	*SD*	*M*	*SD*
Dative	Recasts ($n = 20$)	.067	.160	.208	.189	.155	.169
	Prompts ($n = 21$)	.063	.129	.523	.344	.486	.339
	Control ($n = 23$)	.030	.053	.056	.105	.035	.063
Comparative	Recasts ($n = 20$)	.383	.209	.440	.304	.497	.205
	Prompts ($n = 21$)	.404	.296	.812	.181	.691	.244
	Control ($n = 23$)	.428	.215	.425	.194	—	—*

Note. *For the oral test on comparatives, no delayed posttests were administered.[3]

on the dative, we see an increase between M0 and M1 ($d = 1.77$) and a decrease after M1 ($d = -.11$). For the effect of prompts on the comparative we see the same pattern: an increase between M0 and M1 ($d = 1.66$) and a decrease after M1 ($d = -.56$).

Observing the effects of recasts on the dative, we see an increase between M0 to M1 ($d = .81$) and a decrease after M1 ($d = -.30$). The same results hold for the comparative: an increase between M0 and M1 ($d = .22$) and an increase after M1 ($d = -.22$).

The intervention effect for feedback type was not moderated by structure. In other words, no structure/feedback type interaction $F(2,284) = 1.57$, $p = .209$ was found.

Oral Fluency Test

The means and standard deviations for oral fluency appear in Table 3. The results for oral fluency showed a main effect for (a) feedback type $F(2,62) = 5.3$, $p = .007$, (b) time $F(2,259) = 54.86$, $p < .001$, and (c) structure $F(1,259) = 4.46$, $p = .036$. The analyses also revealed a significant structure/feedback type interaction, $F(2,259) = 3.09$, $p = .047$. This suggests that the effect of recasts and prompts on fluency differed according to structure: The recast group performed the dative task more fluently than the prompt group did, while for the comparative task no differences were observed. The analyses revealed no significant differences for either feedback type/time interaction, $F(3,259) = .216$, $p = .885$, or structure/feedback type/time interaction, $F(3,259) = .540$, $p = .655$.

Accuracy Versus Fluency

To examine Skehan's (1996, 1998) trade-off hypothesis, correlation coefficients (Pearson's *r*) were calculated between oral accuracy and oral fluency. For the dative task, a significant negative correlation between accuracy and fluency could be ascertained only for the prompt group in the delayed posttest ($r = -.518$, $p = .016$), not for the recast group ($r = -.436$, $p = .055$). For the comparative task, no significant correlations were found between oral accuracy and oral fluency.

DISCUSSION

The present study investigated the effect of (a) recasts and prompts on secondary school stu-

TABLE 3
Group Means (1–5) and Standard Deviations for Oral Fluency Tests

		M0		M1		M2	
	Group	*M*	*SD*	*M*	*SD*	*M*	*SD*
Dative	Recasts ($n = 20$)	2.70	.865	3.30	.657	3,70	.733
	Prompts ($n = 21$)	2.24	.890	2.76	.700	3,14	.727
	Control ($n = 23$)	2.13	.757	2.74	.810	2.75	.715
Comparative	Recasts ($n = 20$)	2.35	.875	3.05	.759	3.20	.696
	Prompts ($n = 21$)	2.10	.700	2.91	.768	3.33	.913
	Control ($n = 23$)	1.96	.825	2.57	.590	—	—*

Note. *For the oral test on comparatives, no delayed posttests were administered.[3]

dents' accuracy of two new and different grammar structures, (b) prompts and recasts with two kinds of target structures that differed in difficulty and their relatedness to the L1, and (c) a focus on oral accuracy on oral fluency.

The results demonstrate that recasts can affect the acquisition of new grammar structures. Results of both the oral and written tests show that the recast group performed significantly better than the control group. Lyster and Izquierdo (2009) concluded in their study that the recast group either benefited from repeated exposure to positive exemplars of the grammar structure and/or was able to compare the recast and their own incorrect utterance, in which case the recasts served as negative evidence. Since the recast group was not provided with instruction on the two target structures, it would appear that the students mostly benefited from positive exemplars. The combination of hearing positive exemplars of the grammar structures via recasts and the students' use of the structures during the oral presentations in the classroom may have resulted in oral proficiency gains. This might also explain the difference in effect sizes for the oral and written tests: From pretest to posttest 2, the mean effect size for the written tests for both structures was small ($d = .41$) and for the oral tests medium ($d = .54$).

Besides the exposure to positive and negative exemplars, the effectiveness of the recasts might also derive from the fact that the recasts provided were short and consisted of only one corrective change (Goo & Mackey, 2013). This may have added to their saliency while making few demands on cognitive capacity. Even without providing overt language-focused instruction, then, recasts proved to be effective in enabling learners to acquire new grammatical structures.

RQ2 addressed whether prompts had an effect on the accuracy of newly acquired grammatical structures. This question, too, can be answered in the affirmative. On both oral and written tests, the prompt group outperformed the control and the recast group.

Disaggregating that effect was the focus of RQ3. Specifically, prompts were more effective than recasts in promoting the acquisition of new grammar structures. This finding is in line with earlier studies (Ammar, 2008; Ammar & Spada, 2006; Ellis et al., 2006; Loewen & Philp, 2006; Lyster, 2004; Yang & Lyster, 2010). Prompts may be superior because they are more salient to the students and are therefore noticed more readily. Another possible explanation for their superiority is that their first part, the metalinguistic feedback, triggers access to the rule-based knowledge system. The student makes a connection with a rule and is therefore able to apply the rule in a different setting. The second part of the prompt, the elicitation, might also trigger learning because students are asked to engage in self-repair (Swain, 1995). Lyster (2004) characterizes this as an opportunity for students to practice, which could lead to restructuring of the interlanguage. In line with de Bot (1996), we observed that the elicitation led to students' active involvement in the process, with a high level of attention. This active role can generate strong connections in memory and therefore may lead to subsequent learning.

Learner age and performance level may also be a contributing factor. The 14-year-old participants in the study performed on a low intermediate level and might not have been able to benefit from recasts as much as developmentally more advanced students (Mackey & Philp, 1998). At the same time, students were eager and willing to respond to prompts, reporting that it was motivating to get the teacher's attention and that they "learnt a lot" during this moment of attention. In line with Skehan (1998), we conclude that conscious awareness of the feedback is crucial, because appreciation for receiving corrections and the required transformation into new output "predispose the student towards a rule-based perspective which is more likely to lead to longer-term change" (p. 57).

The development of a rule-based perspective might also explain the differences in mean effect sizes between the oral and written tests from pretest to posttest 2. The mean effect size for both structures of the written tests ($d = 2.13$) was more than one and a half times larger than the mean effect size of the oral tests ($d = 1.35$). In contrast to recasts, prompts might have stimulated the acquisition of explicit knowledge, which the written test, as compared to the oral tests, tapped into particularly well.

RQ4 investigated whether the effectiveness of recasts and prompts depends on the target structure. An affirmative answer would require significant interaction between feedback type and structure. Indeed, we found such an interaction effect on written accuracy, which indicates that the difference in effect on the dative and comparative structure was larger for the recast than for the prompt group. Compared to prompts, recasts had a larger effect on the comparative structure than on the dative structure, a difference that may be attributable to the nature of the target structures. The dative structure is considered to be a complex, syntactic, and rule-based feature,

with no similarities to the students' L1. The recast group was not informed on the rules of the dative structure during the feedback moment and, as a consequence, lacked the kind of explicit knowledge that could have facilitated performance on the written test. It would have been difficult for learners to deduce rules from the input because complex syntactic structures appear to be difficult to notice (Hulstijn & de Graaff, 1994).

That recasts were more successful in promoting the comparatives is plausible because the students already possessed explicit knowledge on the forming of the comparative from their L1. This facilitated applying the rule as well as recognizing the correct form in the recast and comparing it with their own incorrect utterance. These results support Long's (2007) view that recasts serve the acquisition of new linguistic structures that bear meaning and are noticeable. It appears that the structure's relatedness to the L1 also enhances its salience and its connection to meaning.

Another interaction between group and structure was found for oral fluency. The difference in fluency performance between the dative and the comparative was larger for the recast than for the prompt group. The recast group performed the dative task more fluently than the comparative task, whereas the prompt group performed the comparative task slightly more fluently than the dative task. Presumably, the students of the recast group did not think about a rule while carrying out the dative task and because of this lack of focus on accuracy, they freed up attentional resources that could be devoted to fluency. The prompt group, on the other hand, may have attempted to retrieve the rule from long-term memory in order to produce the dative form correctly.

Moreover, within the prompt group we found a trade-off between accuracy and fluency, the focus of RQ5. When students' accuracy and fluency scores were correlated, results indicated a negative correlation for the prompt group in the delayed oral posttest of the dative structure. In other words, students who performed the task more accurately showed less fluency, a result that seems to confirm Skehan's (1996, 1998) trade-off theory, which claims that increased attention to accuracy will lead to decreased fluency and vice versa. By contrast, for the morphologically simpler comparative that is also similar to the L1, no negative correlation between accuracy and fluency was found, suggesting that, under specific conditions, attentional resources are not in competition with each other.

Limitations and Future Research

Despite the positive findings of the present study, certain limitations need to be considered. First, the students of the control group followed the regular curriculum in which the target structures of the present study were embedded, but they did not participate in the communicative tasks. In other words, they had no previous exposure to the content of the task nor any experience with performing such a task, a factor that might have disadvantaged them on the oral tests. It may also explain why students in the control group were not motivated to participate in the delayed oral test on the comparatives. They stated that they were tired of participating in the tests, while their curriculum showed no differences with what they were used to. In future studies on CF, it might be advisable to let the control group participate in the treatment tasks, though without providing CF (see also Lyster et al., 2013).

The oral accuracy test for the comparative (i.e., compare two mobile phones) might also have limitations in that it did not always elicit the intended target structures: For example, some students used the same target structure for different elements, resulting in incorrect as well as correct utterances. We counted and judged these as different utterances.

In the current study we made use of humans raters to measure students' fluency. Although research demonstrates that several aspects of human-rated fluency (e.g., speech rate and pausing) correlate with more objective measures of fluency, we cannot rule out the possibility that the human raters' judgment was influenced by students' accuracy or accent.

Although the number of feedback moments did not differ between the two experimental conditions, the variation within the condition for this measure was sometimes quite large, indicating that some students received more feedback than others. We did not find a relation between the number of feedback moments and learning gain, but nevertheless, in a subsequent study we would control for the amount of feedback moments per hour to ensure equivalence between participants.

Obviously, because the study measured the effects of CF on only two different German target structures, additional research is needed to specify more precisely which linguistic structures benefit from which type of CF and in what way the L1 may contribute to this process.

CONCLUSIONS

To summarize, the results of the two experiments show that both recasts and prompts contribute to the acquisition of new structures, with prompts being more effective than recasts. They confirm earlier research on the effects of recasts versus prompts and contribute to a more nuanced understanding of the effect of feedback in relation to the difficulty and L1-relatedness of the linguistic structure: There exists an interaction effect between structure and feedback type for written accuracy and oral fluency. For written accuracy, recasts, as compared to prompts, had a larger effect on the comparative than on the dative structure. For oral fluency, results showed that, compared to the prompt group, the recast group performed the dative task far more fluently than the comparative task.

It seems appropriate, then, to conclude that prompts may be used in the classroom effectively for both complex as well as simple rules. But because prompts may interrupt the flow of communication quite obtrusively, correction of simple rules, which are related to the L1, may also be done effectively through recasts. The structure's relatedness to the L1 makes it easier for the students to notice the recasts, which may enable them to compare the target-like structure with their own false utterance and promote acquisition. Finally, with a complex syntactic structure it appears that attention for accuracy comes at the expense of fluency.

ACKNOWLEDGMENTS

This study is based on the first author's PhD research (van de Guchte, 2015). We gratefully acknowledge the cooperation of the participating teachers and students. We thank Heidi Byrnes and the anonymous MLJ reviewers for their valuable input and feedback on earlier versions of this article. This study was funded by the Research Institute of Child Development and Education of the University of Amsterdam, the Gooise Scholen Federatie, Bussum, the Netherlands, and the Roland Holst College, Hilversum, the Netherlands.

NOTES

[1] Translation: in, at, on, behind, next to, below, above, in front of, between

[2] Translation: on the bed (the) lies a pillow

[3] For the oral test on comparatives, no delayed posttests were administered in the control group because the students were not motivated to perform them. We did not force the issue in order to avoid a potentially negative influence on the reliability of the results.

REFERENCES

Ammar, A. (2008). Prompts and recasts: Differential effects on second language morphosyntax. *Language Teaching Research, 12*, 183–210.

Ammar, A., & Spada, N. (2006). One size fits all? Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition, 28*, 543–574.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum.

Anderson, J. R. (2000). *Learning and memory. An integrated approach*. New York: John Wiley.

Andringa, S., de Glopper, K., & Hacquebord, H. (2011). Effect of explicit and implicit instruction on free written response task performance. *Language Learning, 61*, 868–903.

Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition, 15*, 357–386.

Council of Europe. (2001). *Common European framework of reference for languages*. Cambridge: Cambridge University Press.

de Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language Learning, 46*, 529–555.

De Jong, N. H., Steinel, M. P, Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics, 34*, 893–916.

DeKeyser, R. (1993). The effect of error correction on L2 grammar knowledge and oral proficiency. *Modern Language Journal, 77*, 501–514.

DeKeyser, R. (2010). Practice for second language learning: Don't throw out the baby with the bathwater. *International Journal of English Studies, 10*, 155–165.

Dilans, G. (2010). Corrective feedback and L2 vocabulary development: Prompts and recasts in the adult ESL classroom. *The Canadian Modern Language Review, 66*, 787–815.

Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114–138). Cambridge: Cambridge University Press.

Egi, T. (2007). Interpreting recasts as linguistic evidence: The roles of linguistic target, length and degree of change. *Studies in Second Language Acquisition, 29*, 511–537.

Egi, T. (2010). Uptake, modified output, and learner perceptions of recasts: Learner responses as language awareness. *Modern Language Journal, 94*, 1–21.

Ellis, R. (1993). The structural syllabus and second language acquisition. *TESOL Quarterly, 27*, 91–113.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27*, 141–172.

Ellis, R. (2007). The differential effects of corrective feedback on two grammatical structures. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 339–360). Oxford: Oxford University Press.

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, *28*, 339–369.

Ellis, R., & Sheen, Y. (2006). Reexamining the role of recasts in second language acquisition. *Studies in Second Language Acquisition*, *28*, 575–600.

Goo, J., & Mackey, A. (2013). The case against the case against recasts. *Studies in Second Language Acquisition*, *35*, 127–165.

Han, Z. (2002). A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly*, *36*, 543–572.

Hulstijn, J. H., & de Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review*, *11*, 97–113.

Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford, UK: Pergamon Press.

Leeman, J. (2003). Recasts and second language development: Beyond negative evidence. *Studies in Second Language Acquisition*, *25*, 37–63.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, *60*, 309–365.

Lightbown, P., & Spada, N. (1999). *How languages are learned*. Oxford: Oxford University Press.

Lochtman, K. (2002). Oral corrective feedback in the foreign language classroom: How it affects interaction in analytic foreign language teaching. *International Journal of Educational Research*, *37*, 271–283.

Loewen, S., & Nabei, T. (2007). Measuring the effects of oral corrective feedback on L2 knowledge. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 379–377). Oxford: Oxford University Press.

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

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in cross-cultural perspective* (pp. 39–52). Philadelphia/Amsterdam: John Benjamins.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). New York: Academic Press.

Long, M. H. (2007). Recasts in SLA: The story so far. In M. H. Long (Ed.), *Problems in SLA* (pp. 75–116). Mahwah, NJ: Lawrence Erlbaum.

Long, M., Inagaki, S., & Ortega, L. (1998). The role of negative feedback in SLA: Models and recasts in Japanese and Spanish. *Modern Language Journal*, *82*, 357–371.

Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research and practice. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15–41). New York: Cambridge University Press.

Lyster, R. (1998). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition*, *20*, 55–85.

Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, *26*, 399–426.

Lyster, R., & Izquierdo, J. (2009). Prompts versus recasts in dyadic interaction. *Language Learning*, *59*, 453–498.

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

Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition*, *32*, 265–302.

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, *46*, 1–40.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 407–452). Oxford: Oxford University Press.

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

McDonough, K. (2007). Interactional feedback and the emergence of simple past activity verbs in L2 English. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 323–338). Oxford: Oxford University Press.

Miller, P. C., & Pan, W. (2012). Recasts in the L2 classroom: A meta-analytic review. *International Journal of Educational Research*, *56*, 48–59.

Nassaji, H. (2009). Effects of recasts and elicitations in dyadic interaction and the role of feedback explicitness. *Language Learning*, *59*, 411–452.

Nicholas, H., Lightbown, P., & Spada, N. (2001). Recasts as feedback to language students. *Language Learning*, *51*, 719–758.

Robinson, P. (2001). Task complexity, task difficulty and task production: Exploring interactions in a componential framework. *Applied Linguistics*, *22*, 27–57.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework

for second language task design. *International Review of Applied Linguistics, 43*, 1–32.

Robinson, P. (2011). Task-based language learning: A review of issues. *Language Learning, 61*, 1–36.

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

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 127–158.

Schoormann, M., & Schlak, T. (2012). Mündliche Korrekturstrategien im Zweit- und Fremdsprachenunterricht: Recasts und Prompts im Vergleich [Oral corrective feedback in second and foreign language education: A comparison of recasts and prompts]. *Journal of Linguistics and Language Teaching, 3*, 15–59.

Sheen, Y. (2011). *Corrective feedback, individual differences and second language learning*. Dordrecht, the Netherlands: Springer.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics, 17*, 38–62.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P., Xiaoyue, B., Qian, L., & Wang, Z. (2012). The task is not enough: Processing approaches to task-based performance. *Language Teaching Research, 16*, 170–187.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principles and practice in the study of language*. (pp. 125–144). Oxford: Oxford University Press.

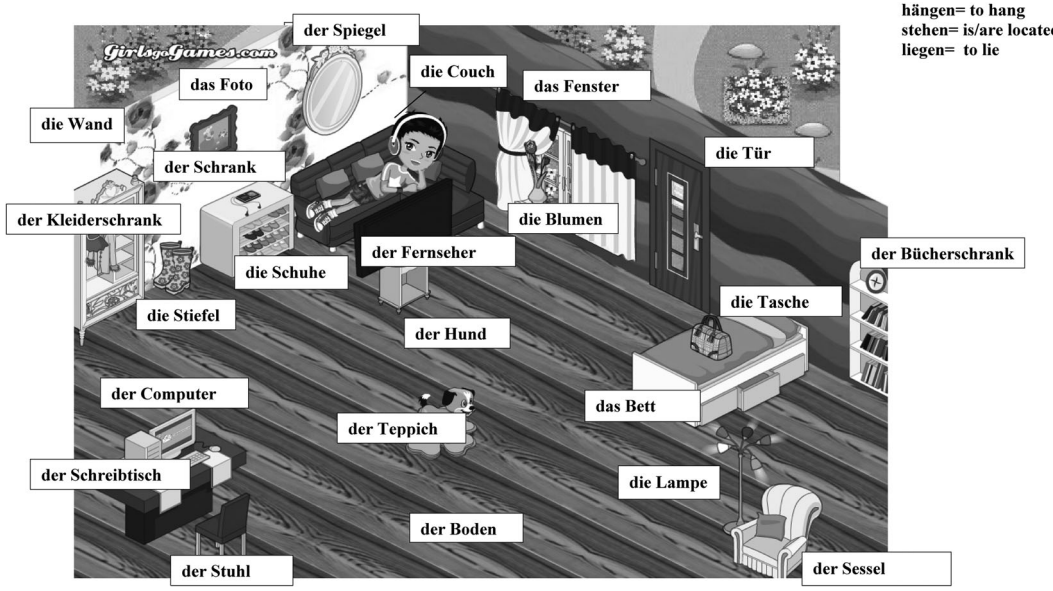
Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97–114). Oxford: Oxford University Press.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–483). New York: Routledge.

Yang, Y., & Lyster, R. (2010). Effects of form-focused practice and feedback on Chinese EFL students' acquisition of regular and irregular past tense forms. *Studies in Second Language Acquisition, 32*, 235–263.

APPENDIX A

Example Oral Accuracy and Fluency Test on the Dative



APPENDIX B

Example Oral Accuracy and Fluency Test on the Comparative

1. Vergleiche Preis, Größe, Speicher und so weiter von diesen zwei Handys miteinander.
   [Compare the price, size, memory, et cetera from these two mobile phones.]
2. Welches Handy würdest du kaufen und warum?
   [Which mobile phone would you buy and why?]

	Blackberry Bold 9900	**iPhone 4S**
Preis (price)	€695	€549
Länge (length)	115 mm lang	116 mm lang
Breite (width)	59 mm breit	66 mm breit
Dicke (thickness)	12 mm dick	10 mm dick
Gewicht (weight)	145 Gramm schwer	131 Gramm schwer
Touchscreen	Ja, 640 × 480 Pixel Auflösung	Ja, 960×640 Pixel Auflösung
Batteriedauer (battery length)	6 Stunden lang	5 Stunden lang
Speicher (memory)	14 GB groß	12 GB groß
Verfügbare Farben (colors)	Schwarz	Weiß
Downloadgeschwindigkeit (Download speed)	15.4 Megabit pro Sekunde	16.5 Megabit pro Sekunde