Author Manuscript

Yilmaz, Y. (2016). The linguistic environment, interaction and negative feedback. *Brill Research Perspectives in Multilingualism and Second Language Acquisition*, *1*, 45-86.

**Full title:** The Linguistic Environment, Interaction and Negative Feedback

**Author:** Yucel Yilmaz

**Affiliation**: Indiana University

**Address for correspondence:** Yucel Yilmaz, [yyilmaz@indiana.edu](mailto:yyilmaz@indiana.edu), Indiana University Memorial Hall 303, Bloomington, IN 47408

*Negative feedback* (NF) refers to the reactions language learners receive from their interlocutors indicating that the learners’ language production is not target-like. In recent years, the relationship between NF and *second language* (L2) acquisition has evolved into one of the most vibrant research areas in the field of *second language acquisition* (SLA). The vast number of journal articles, book chapters and books that review or meta-analyze the empirical research on the topic (Goo & Mackey, 2013; Goo, Granena, Yilmaz & Novella, 2015; Li, 2010; Long, 2007; Lyster & Ranta, 2013; Lyster & Saito, 2010; Lyster, Saito & Sato, 2012; Mackey & Goo, 2007; Nassaji, 2015; Nicholas, Lightbown, & Spada, 2001; Russell & Spada, 2006; Sheen & Ellis, 2011) indicates how much attention the topic attracts among scholars of SLA. This article reviews recent theoretical and empirical literature on the role of NF in L2 acquisition. The scope of the review is restricted to publications that have focused on the NF that is provided to a learner by a more proficient interlocutor (e.g., a native speaker, teacher or researcher) during real-time meaning-oriented communication. In an attempt to situate NF research within a larger theoretical framework, the review begins by providing an overview of research focusing on the role of the linguistic environment and interaction in L2 acquisition. This section is followed by a review of key empirical studies examining the extent to which NF is available, useful, used and necessary in L2 acquisition, as well as which factors moderate the efficacy of NF. The article ends with a discussion of some methodological issues and directions for future research.

**The Linguistic Environment and Conversational Interaction**

NF holds a central place in SLA research today as a result of a tradition that focused on the role of the linguistic environment in L2 acquisition. Contributing studies focused primarily on the analysis of a special register called foreigner talk, in which native speakers modify their speech to make themselves more understandable to non-natives. The findings from this literature indicated that the input available in the linguistic environment is generally grammatical but different from the kind of speech that takes place between adult native speakers. Some of these differences included shorter, syntactically/propositionally less complex utterances, slower speech, frequent pauses at constituent boundaries, clearer articulation, fewer irregular items, and a narrower range of higher frequency lexical items (Ferguson, 1981; Freed, 1981; Long, 1983). In this context, the work of two researchers, Hatch (1978) and Long (1980, 1981, 1983), drew the field’s attention to the interactional structure of the conversations between *native speaker* (NS) and *non-native speaker* (NNS) interlocutors and the possibility of a causal link between conversational interaction and L2 acquisition.

Hatch (1978) proposed a new perspective for exploring the relationship between conversational interaction and linguistic knowledge. The prevailing view of the time was that learners gain knowledge about L2 structures first, and then they put this knowledge to use during conversation. In Hatch’s view, learners first learn how to do conversation, and knowledge of L2 structures develops through these interactions. In other words, “learning evolves out of learning how to carry on conversations” (Hatch, 1978, p. 404). In addition to Hatch’s proposal, Long’s *Interaction Hypothesis* (IH; 1981, 1983) provided a framework for understanding how interaction could contribute to L2 acquisition and motivated a research agenda investigating the link between conversational interaction and L2 acquisition. Long (1980) showed that interactions between NSs and NNSs, in comparison to interactions between NSs and other NSs, involved significantly more instances in which NSs modified the interactional structure of the conversation to avoid conversational trouble or to repair communication breakdowns. Long (1983) claimed that these modifications (i.e., conversational adjustments) might aid L2 acquisition by making input comprehensible for learners. Since comprehensible input was considered the causal variable in L2 acquisition at that time (Krashen, 1978), the assumed link between comprehension and acquisition underlying Long’s claim had some theoretical support. The parts of interaction involving such modifications came to be known as negotiation, which was defined by Pica (1994) as the “modification and restructuring of interaction that occurs when learners and interlocutors anticipate, perceive, or experience difficulties in message comprehensibility” (p.494).

Early research inspired by the IH focused on the conditions that affected the frequency of negotiation. For example, it was shown that tasks in which information exchange was essential for task completion elicited more negotiation than tasks in which information exchange was not required (Doughty & Pica, 1986; Gass, Mackey, & Ross-Feldman, 2005; Long, 1980; Pica, 1987). It was also found that negotiation was more frequent in NNS-NS dyads than in NS-NS dyads (Long, 1980; Oliver, 2002), and that mixed rather than homogenous proficiency learners produced more negotiation (Doughty & Pica, 1986; Gass & Varonis, 1985a; Porter, 1986). Finally, mixed-gender dyads were found to negotiate more than same-gender dyads (Gass & Varonis, 1985b; Pica, Holliday, Lewis, & Morgenthaler, 1989).

Evidence of a direct link between negotiation and comprehension was provided by Pica, Young, and Doughty (1987). Their study showed that engaging in interaction in which there were opportunities for negotiation was more effective in assisting L2 learners’ listening comprehension than hearing premodified input (e.g., input with increased redundancy and simplified grammar) without any opportunities for interaction. Comprehension, in this study, was operationalized as the number of directions carried out accurately in an object-placement task. In another study using the same task, Pica (1991) showed that not only the learners who interacted with NSs, but also the learners who listened to the NNS-NS interaction, demonstrated improved performance on the task. Despite this empirical evidence of a relationship between comprehension and negotiated interaction, Pica (1994) argued that it was restricting to view comprehension of meaning as the main path to acquisition, and negotiation as the main way to ensure comprehension. According to Pica, this view relegates interaction to a secondary role in L2 acquisition because interaction can only be relevant to the extent that comprehension is relevant for L2 acquisition. She added that the role in L2 acquisition of comprehension itself was still contentious.

An important piece of evidence that led many researchers to be skeptical about the role of comprehension came from evaluations of French immersion programs in Canada where native English-speaking learners of French were exposed to massive amounts of comprehensible input. These evaluations (Swain & Lapkin, 1982, 1986) revealed that learners’ receptive skills (i.e., listening and reading) were indistinguishable from NSs’ receptive skills, but their productive skills (i.e., speaking and writing) were worse than NSs’ productive skills. In addition, these learners never achieved “native-like” levels of grammatical accuracy, continuing to make a wide variety of errors, including verb tense, prepositions, gender marking on articles, and basic vocabulary. Similar findings were reported for other meaning-focused, content-based classrooms (Pica, 2005).

In the updated version of the IH, Long (1996) argued that comprehensible input is necessary but not sufficient for learning certain problematic aspects of an L2 and that learners’ attention should be drawn to those aspects. The role of attention in L2 learning is theoretically supported by Schmidt’s Noticing Hypothesis (1990, 2001). According to Schmidt, “a second language learner will begin to acquire the target-like form if and only if it is present in comprehended input and ‘noticed’ in the normal sense of the word, that is consciously” (Schmidt & Frota, 1986, p. 311). In a later article, Schmidt (2001) explicitly drew attention to the fact that his noticing construct is different from metalinguistic awareness. Schmidt (2001) explained that metalinguistic awareness involves “comparisons across instances and metalinguistic reflection [thinking about what has been attended and noticed, forming hypotheses]” (Schmidt, 2001, p. 5), whereas noticing involves the cognitive registration of the surface structure of the instances of language. Long’s (1996) updated version of the IH discussed the potential role that implicit NF could play in drawing learners’ attention to problematic linguistic forms. He asserted that “negative feedback obtained during negotiation work or elsewhere may be facilitative of L2 development, at least for vocabulary, morphology and language-specific syntax, and essential for learning certain specifiable L1-L2 contrasts” (Long, 1996, p. 414). Long (1996) went on to describe the ways conversational interaction could contribute to L2 acquisition indirectly, e.g., by encouraging pushed output, facilitating comprehension, and increasing metalinguistic awareness. He admitted that, up to that point, evidence of a direct link between interactional modifications and acquisition was scarce.

Mackey (1999) provided the first convincing evidence that interaction and L2 acquisition are linked. In Mackey’s study, some learners, but not others, were allowed to interact with NSs. Of two groups that did not interact with NSs, one received premodified input, and the other was permitted to listen to the input that was given to the interactors. The study showed that developmentally ready learners who actively participated in interaction improved their performance from pretest to posttest, but the others did not. It is particularly important to note that this study, like many early interaction studies, treated interaction as a global variable, which included many different interactional features, e.g., positive evidence, comprehension enhancing interactional adjustments, and NF. The emphasis placed on NF in the updated version of the IH (Long, 1996) motivated a new line of research in which the effectiveness of one interactional feature, usually NF, is investigated in isolation from other interactional features.

**The Role of Negative Feedback in L2 Acquisition**

It is commonly accepted that two types of evidence are available for language learners: positive and negative. Positive evidence refers to what is possible in the target language and is obtained through well-formed exemplars of the target language available either in naturalistic or instructional settings. Negative evidence refers to what is not possible in the target language and can primarily be obtained through instruction informing learners about the incorrect uses of L2 forms or through NF.

In the child language acquisition literature, the role of negative evidence has been debated. Researchers have discussed the potential of negative evidence to prevent children from making incorrect generalizations (i.e., overgeneralizations) when learning linguistic rules. For example, Pinker (1989) has pointed out that there are cases in which internalization of rules without making overgeneralizations is possible on the basis of positive evidence alone. However, in cases in which learners’ grammar constitutes a superset of the target language grammar, no amount of positive evidence would allow learners to retreat from overgeneralizations. In these cases, learners’ hypotheses should be constrained either by negative evidence or innate linguistic knowledge. Pinker has maintained that the following conditions must be demonstrably satisfied before one can attribute a role to negative evidence: (1) negative evidence should be present in the input; (2) negative evidence should be useful: the child should construe the adult’s utterance as disapproving his/her utterance, and the adult’s utterance should be informative enough to allow the child to make a correct blame assignment (i.e., to identify which of the rules that led him/her to produce the non-target-like utterance has to be selected for change); (3) negative evidence should be used: it must be shown that negative evidence could lead to a measurable change in the child’s grammar; (4) negative evidence should be necessary: it should be shown that all children need negative evidence in order to learn a language.

Early child language acquisition research showed that explicit forms of negative evidence were rare (i.e., explicit rejections of children’s utterances, Brown and Hanlon, 1970). Other research, however, showed that more implicit forms, such as recasts (i.e., target-like reformulations of non-target-like utterances), occurred in response to children’s ungrammatical utterances (e.g., Demetras, Post, & Snow, 1986; Bohannon & Stanowicz, 1988; Saxton, Backley, & Gallaway, 2005). It was also demonstrated that recasts were useful and used by children in such a way as to promote the long-term acquisition of grammar (Farrar, 1990; Saxton, 1997; Saxton, Kulcsar, Marshall, & Rupra, 1998; Saxton, 2000; Strapp, Bleakney, Helmick, & Tonkovich, 2008). Despite this empirical evidence that seems to satisfy the first three conditions, there is an ongoing debate as to whether negative evidence is necessary for L1 acquisition. According to Pinker (1994), this condition cannot be satisfied because there are some cultures around the world in which parents do not modify their speech when talking to their infants (e.g., Heath, 1983), and, therefore, it is very unlikely that these infants receive negative evidence. There is, however, a strong objection to this position by other researchers (e.g., Saxton, 2010) on the grounds that the studies (e.g., Heath, 1983) used to support this argument lack crucial information that would help researchers decide whether children in those cultures experienced negative evidence either from their parents or their older siblings.

SLA researchers also disagree about the role of negative evidence. Some researchers (e.g., Krashen, 1985; Truscott, 1999; Schwartz, 1993; VanPatten, 2013) have argued that positive evidence is the only type of evidence learners need and that negative evidence, and therefore NF, cannot play a major role in L2 acquisition. These researchers believe that NF can only produce a superficial form of L2 knowledge, not affecting learners’ competence. Despite this opposition, the majority of SLA researchers (e.g., Ellis, 1991; Gass, 1997; Long, 1996; Mackey, 1999; Pica, 1988; White, 1991) claim that NF plays at least a facilitative role in L2 acquisition, while acknowledging the crucial role of positive evidence. Key empirical findings are reviewed below using Pinker’s (1989) four conditions as criteria to determine the exact nature of the role of NF in L2 acquisition.

Pinker’s (1989) first condition was that negative evidence should be available. Several studies (Chun, Day, Cheneworth, & Lupescu, 1982; Braidi, 2002; Iwasaki & Oliver, 2003; Oliver, 1995; Oliver, 2000) investigated the extent to which NF is available in oral NS-NNS conversations in non-classroom contexts. In some of these studies, the topic of the conversation was not controlled for (e.g., Chun et al., 1982), in others the researchers administered a communicative task to control for topic (Braidi, 2002; Oliver, 1995; Oliver, 2000). Chun et al. (1982) and Braidi (2002) focused on adult learners, Oliver (1995) on children. Oliver (2000) manipulated age as a variable and included both adults and children in her study. In Chun et al. (1982), only 8.9% of errors attracted NF. They reported that most feedback instances (66%) were unambiguously corrective (i.e., explicit), but others (34%) were ambiguous with respect to the message they conveyed (i.e., implicit). In other studies, the percentage of errors that attracted NF was much higher. These studies reported percentages ranging from 26% to 47% (26% in Braidi [2002], 41% in Oliver [1995], 47% for adults and 40% for children in Oliver, [2000]). Iwasaki and Oliver’s study (2003) showed that NS-NNS interactions through text-based *synchronous computer-mediated communication* (SCMC) also contained NF. In their study, 22% of all the NNSs’ errors attracted recasts or clarification requests (see Table 1 for the definitions and examples of these feedback types). Studies that were carried out in classroom contexts reported percentages ranging between 48% and 62% (61% in Lyster & Mori [2002], 62% in Lyster & Ranta [1997], 48% in Panova & Lyster [2002]). In these classroom studies, as well as in Chun et al. (1982) study, a wide variety of feedback types were used by the interlocutors. In the other non-classroom studies, NF included either recasts or negotiation (i.e., clarification requests or confirmation checks). Taken together, these results show that NF is indeed available to the L2 learner in a variety of different contexts. It is also clear that the percentage of errors eliciting NF differs from context to context. The exact percentage reported in the studies is not of central importance at this point, because it is simply not known whether a specific percentage of learners’ errors should be addressed with NF for them to begin to take advantage of NF.

Pinker’s second criterion was that negative evidence should be useful. In order to claim that learners benefit from the negative evidence contained in NF, it should be shown that learners can perceive negative evidence as such. If this cannot be shown, even if NF is eventually found to be effective, the effectiveness cannot be attributed to negative evidence and should be explained in other ways (e.g., enhanced positive evidence [Leeman, 2003]). Several studies using verbal protocols, such as stimulated and immediate recalls, have tried to determine whether learners can perceive the negative evidence conveyed through their interlocutors’ responses to their language productions. In the immediate recall technique, researchers give feedback and immediately ask learners either to repeat what they have just heard (e.g., Philp, 2003) or to verbalize their thoughts at the time of feedback (e.g., Egi, 2007). In the stimulated recall technique, researchers record feedback episodes and later ask learners to watch the recordings and to verbalize their thoughts at the time they were receiving the feedback (e.g., Egi, 2007). In Philp’s study (2003), *English as a second language* (ESL) learners performed communicative tasks with a NS, during which they received recasts on English question formation. After each recast, the learners were asked to repeat the last thing they heard. It was found that the learners were able to repeat 60% to 70% of recasts, indicating that they heard the recasts and were able to keep them in their memory for a short while. This finding, however, does not tell us whether learners can notice the negative evidence. Egi’s study (2007), which used both stimulated and immediate recalls, shed some light on this question. Japanese as a *foreign language* (FL) learners carried out communicative tasks with a NS of Japanese and received recasts on their morphosyntactic and lexical errors. The results showed that 60% of recasts targeting morphosyntax and 64% of recasts targeting lexical errors were noticed as either negative or as both negative and positive evidence. Kim and Han (2007) also used stimulated recalls to investigate the extent to which *English as a foreign language* (EFL) learners perceived recasts provided during communicative tasks. Their study showed that learners recognized 65% of all recasts as having a corrective intent. Gurzynski-Weiss and Baralt (2014) added to this line of research by investigating other feedback types, in addition to recasts. In their study, a variety of different feedback types were provided to Spanish FL learners during two communicative tasks, one through *face-to-face communication* (FTFC) and one through text-based SCMC. Learners perceived 68% of the feedback in SCMC and 71% of the feedback in FTFC.

Even though these findings indicate that learners can recognize the negative evidence in various types of feedback, perceiving negative evidence might not be enough for learners to take advantage of NF. According to Pinker (1989), learners should also be able to identify the source of the linguistic problem. Using the stimulated recall technique, Mackey, Gass and McDonough (2000) demonstrated that both ESL and Italian FL learners were the least accurate in perceiving the linguistic target of feedback (i.e., recasts or negotiation) provided on morphosyntactic errors (13% in ESL; 24% in Italian FL); however, they were more accurate in perceiving the linguistic target of lexical (83% in ESL; 66% in Italian FL) and phonological (60% in ESL; 21% in Italian FL) feedback. Similar results were reported in Gass and Lewis (2007) for heritage and non-heritage learners of Italian. Kim and Han (2007) reported that 45% of all the stimulated recall comments learners made indicated that they could identify the locus of the problem in the utterances that attracted the feedback. In Gurzynski-Weiss and Baralt (2014), learners perceived the linguistic target of feedback targeting morphosyntax 48% of the time in SCMC and 42% of the time in FTFC. In their study too, learners were more accurate in identifying the linguistic target of feedback on lexis (79% CMC; 80% FTF). Overall, then, there is evidence that learners can both notice the negative evidence in NF and identify the source of the error in their language production. They seem to be less successful in recognizing the linguistic target of feedback provided on morphosyntactic errors than on lexical and phonological errors. Future research, however, is needed to determine whether this finding is robust. Since none of the reviewed studies controlled for feedback type or morphosyntactic feature, one cannot rule out the possibility that the finding is due to the low salience of a specific feedback type or of certain morphosyntactic structures. It should also be noted that retrospective verbal reports may not always capture learners’ noticing behavior. For example, absence of comments about a feedback instance can stem from factors such as memory constraints or lack of ability to describe one’s own behaviors (Jourdenais, 2001), and therefore cannot be taken as unambiguous evidence that learners have failed to notice feedback. For this reason, it is important to complement verbal reports with data collected through other techniques (e.g., questionnaires, eye-tracking) in future research.

Pinker’s third condition is that negative evidence should be used by learners. Even though verbal reports provide insights into learners’ subjective experience of a NF situation, findings indicating that learners can neither perceive negative evidence nor do a correct blame assignment would not necessarily mean that NF is not useful. As long as it is shown that learners can use NF, as evidenced by a measurable change in learners’ developing system, it is possible to conclude that NF is still useful, but in ways that cannot be captured by the data-collection techniques used so far. In other words, evidence of the extent to which learners use NF can also be taken as evidence of its usefulness. Such evidence can take various forms.

**Learner responses.**  One way of measuring whether learners can use linguistic information provided in feedback is to analyze their responses to feedback. Different labels have been used to refer to such responses, with ‘uptake’ being the most general. Lyster and Ranta (1997) defined uptake as “a student’s utterance that immediately follows the teacher’s feedback and that constitutes a reaction in some way to the teacher’s intention to draw attention to some aspect of the student’s initial utterance” (p. 49). Uptake can capture a wide variety of discourse behaviors, ranging from simple acknowledgements of feedback (“Yes”, “Ok”) to reformulations of non-target-like productions. According to Lyster and Ranta’s (1997) taxonomy, uptake can be considered repair (i.e., target-like reformulations of learners’ off-target utterances) or needs repair (i.e., non-target-like attempts to improve initial off-target utterances or simple acknowledgements, e.g., “Yes”, “Ok”). A related term, ‘modified output’, similar to Lyster and Ranta’s (1997) repair category, refers to “learner production that reformulates an ill-formed utterance following feedback” (Loewen, 2013, p.675). Modified output is different from repair because modified output includes “more or fully target-like reformulations of the error (Egi, 2010, p. 18), whereas repair includes only target-like reformulations.

Sheen’s (2004) analysis of learner responses to NF in four instructional settings (i.e., New Zealand ESL in Ellis, Basturkmen, & Loewen, 2001; Canadian French Immersion in Lyster & Ranta, 1997; Canadian ESL in Panova & Lyster, 2002; Korean EFL in Sheen, 2004) revealed that learners responded to between 47% and 80% of all NF instances with uptake. In between 34% and 69% of these uptake instances, learners repaired their off-target productions. In a lab study carried out in an ESL setting, Oliver (1995) demonstrated that children correctly modified their original utterances more than a third of the time after recasts when the discourse context allowed them to do so. In other lab studies, adults in various settings (ESL, Italian FL, Japanese FL) were found to modify their utterances in response to recasts (Loewen & Philp, 2006; Mackey, et al., 2000; Oliver, 2000; Egi, 2010), negotiation (Mackey, et al., 2000; Oliver, 2000), prompts (Loewen & Philp, 2006), or metalinguistic feedback (Loewen & Philp, 2006) between 21% and 82% of the time. Gurzynski-Weiss and Baralt (2014) showed that FTFC was not the only communication mode in which learners modified their output. Their study revealed that Spanish FL learners modified their output in response to various feedback types both in FTFC (i.e., 68%) and in SCMC (i.e., 62%) when the discourse context allowed them to do so.

These results indicate that learners use a substantial proportion of NF to modify their initial non-target-like productions in the turn following feedback. The value of this finding is closely dependent upon whether modified output or repair can be taken as evidence of L2 development. Before one can claim that modified output or repair indicates a positive change in learners’ developing system, they should be shown to correlate with performance on other measures of development. The results of studies that have investigated this issue have been mixed. Two studies (Mackey & Philp, 1998; McDonough, 2007) showed that there was a disassociation between the frequency of modified output after recasts (Mackey & Philp, 1998; McDonough, 2007) or clarification requests (McDonough, 2007) and pretest-posttest development. In other words, more modified output did not mean more learning, as operationalized by pretest-posttest development. McDonough (2005), however, showed that modified output after repetition plus clarification requests predicted better posttest performance. Overall, previous research has not provided clear evidence as to whether modified output and pretest-posttest development are related.

There is also some evidence showing that learners’ noticing of the corrective intent behind NF is related to their production of modified output. Mackey et al. (2000) showed that 66% of the feedback instances to which learners’ responded with modified output were accurately perceived as NF by the learners. Egi (2010) also reported that when learners accurately perceived the negative evidence in recasts, they were more likely to modify or repair their utterances. These findings have led some researchers to conclude that even though learner responses to NF cannot be a reliable indicator of L2 development, the production of these responses can, nevertheless, indicate that the NF has been noticed (Ellis et al., 2001; Loewen & Philp, 2006; Mackey et al., 2000). One should also take into account that responding to NF is an optional move for learners (Loewen, 2015). This optionality, which indicates that there may be cases in which learners perceive the feedback, but choose not to respond to it, lowers the reliability of these responses as an indicator of noticing or language development.

**Tailor-made tests.** Another method that has been used to assess the effectiveness of NF is tailor-made tests (Swain & Lapkin, 1998). These tests are developed for each learner separately after their participation in interaction. Each item in these tests is designed based on a specific NF instance the test taker was exposed to during the interaction. The following NF episode and test item taken from Loewen and Philp (2006) illustrate how tailor-made items are created. In the NF episode, the teacher provides feedback on the learner’s non-targetlike use of the preposition “to” by saying “protection from”. The test item presents the error in the context of a sentence and asks the learner to correct it.

NF Episode:

S: when I was soldier I used to wear the balaclava

T: and why did you wear it S for protection from the cold or for another reason

S: just wind uh protection to wind and cold

T: protection from

S: uh from wind and cold

T: right (.) okay not for disguise

(Loewen & Philp, 2006, p. 543)

Test Item:

“The following sentence is incorrect or inappropriate. Please listen and tell me how you could make the sentence better: *I used to wear the balaclava for protection to wind and cold*” (Loewen & Philp, 2006, p. 545).

Loewen and Philp (2006) and Nassaji (2010) used tailor-made tests to assess the effectiveness of NF that occurred incidentally in ESL language classrooms. Tailor-made tests were administered in Loewen and Philp (2006) once as an immediate posttest (1 to 3 days after the original interaction) and once as a delayed posttest (13 to 15 days after the interaction). Learners’ accuracy on the test items that were based on errors that attracted a variety of feedback types (recasts, metalinguistic feedback and elicitation) ranged from 50% to 75% on the immediate, and from 50% to 59% on the delayed, posttest. In Nassaji (2010), the tailor-made tests were administered one week after the interaction as a posttest. In this study, as well, the test items were designed based on errors that were treated with a range of feedback types (recasts, repetitions, elicitations, or explicit feedback). The results revealed learners’ overall accuracy on these items was 53%. Together, the findings of the two studies suggest that, around 50% of the time, it is possible for learners to use the information presented in NF. When interpreting these findings, one should keep in mind that tailor-made tests have certain limitations. First, they usually lack a pretest. For this reason, Loewen and Philp (2006) cautioned that “in the absence of pretests, such measures …cannot differentiate between the acquisition of new knowledge and the consolidation of latent knowledge” (p. 542). Another limitation of these tests has to do with the rule-governed aspects of language. Since these test items target only the tokens on which learners receive feedback, they cannot be informative about whether the knowledge gained is generalizable to new tokens.

**Pretest/posttest experimental designs**. Another methodology to assess the extent to which learners use NF is experimental designs. The experimental designs used in the NF literature have included either both a pretest and a posttest, or only a posttest. In instances when only a posttest was employed, there was no need for a pretest because learners were entirely unfamiliar with the target language. Most designs also included a delayed posttest. Evidence for the effectiveness of NF in these studies is provided by comparing learners who receive NF to learners who do not receive NF (i.e., non-feedback group). Non-feedback groups have been of three types. 1) Control group: in some studies, the control group did not receive any sort of additional treatment and participated only in the assessment tasks (e.g., Goo, 2012), in others, the control group performed the same treatment task as the feedback group, but did not receive any feedback (e.g., Yilmaz, 2012, 2013). In a third type of control group, researchers made sure that learners received the type of instruction they would normally receive (Doughty & Varela, 1998) but without any feedback. 2) Model comparison group: these groups received a comparable amount of positive evidence that was not contingent on learners’ errors (e.g., Long, Inagaki, & Ortega, 1998). 3) Negative evidence comparison group: the learners in these groups received proactive negative evidence (i.e., before an error occurs) rather than reactive (i.e., after the error) negative evidence (e.g., Kang, 2010).

There is considerable variability across studies in methodological features. Some studies were carried out in language classrooms, others, in one-to-one tutoring (i.e., lab) settings. In most studies, the L2 under investigation was English (e.g., Mackey & Philp, 1998), French (e.g., Lyster, 2004) or Spanish (e.g., Leeman, 2003), but in a few studies, it was Japanese (e.g., Ishida, 2004), Korean (e.g., Kang, 2010) or Turkish (e.g., Yilmaz, 2012). Morphosyntactic aspects of language were the most popular linguistic target (e.g., Ammar & Spada, 2006; Loewen & Erlam, 2006). Only a few studies focused on other language domains (vocabulary, Dilans, 2010; pragmatics, Koike & Pearson, 2006; phonology, Saito, 2013; Saito & Lyster, 2012). Some of the experimental studies included a greater number of treatment sessions (e.g., Ammar & Spada, 2006; Han, 2002, Ishida, 2004), whereas the rest included a relatively smaller number of treatment sessions (e.g., Herron & Tomasello, 1988, Loewen & Erlam, 2006; Sauro, 2009). Results of the studies that compared NF groups with control (Ammar & Spada, 2006; Doughty & Varela, 1998; Ellis et al., 2006; Goo, 2012; Loewen & Nabei, 2007; Lyster, 2004; Mackey & Philp, 1998; Saito, 2013; Saito & Lyster, 2012; Sauro, 2009; Sheen, 2007; Yilmaz, 2012, 2013b), model (Herron & Tomasello, 1988; Leeman, 2003; Ortega & Long, 1997), or negative evidence comparison (e.g., Kang, 2010) groups have shown that at least one of the NF groups in at least one of the investigated target forms outperformed non-feedback groups. The overall magnitude of the effect of NF on L2 acquisition has been reported as medium-to-large (Li, 2010; Lyster & Saito, 2010; Goo et al., 2015; Mackey & Goo, 2007; Russell & Spada, 2006). In general, these findings suggest that NF is not only available and useful for L2 learners, but it can be used to aid learners’ short and mid-term language development.

Pinker’s fourth and final condition was that negative evidence should be necessary. In other words, it should be shown that learners can acquire an L2 only under conditions where negative evidence is available, useful and used. Long (1996, 2015) has proposed that, in some cases, NF is necessary, and, in others, it is facilitative. According to Long, its facilitative effects are empirically supported by studies demonstrating the effectiveness of NF. His necessity argument is based on White’s ideas on the role of negative evidence for certain language forms. White (1991) has claimed that when a specific L2 form is in a subset relationship with an L1 form, learners’ knowledge about how the L1 form operates would lead them to generate incorrect hypotheses about how the L2 form operates. In these cases, L2 learners cannot retreat from these incorrect generalizations unless they receive negative evidence. For example, there is a large overlap between French and English with respect to the position of adverbs in a sentence, with one notable exception. Even though *Subject Verb Adverb Object* (SVAO; e.g., John kisses often Mary) is a possible word order in French, it is not in English. Therefore, French-speaking learners of English would be expected to consider SVAO a possible word order in English, given the overlap between the languages. White (1991) and Trahey and White (1993) provided empirical evidence that L1 French learners of English need negative evidence to learn that SVAO is an impossible word order in English.

In general, it seems that NF is present in a variety of different learning contexts in different amounts. Verbal protocol data suggest that learners are able to construe their interlocutors’ responses as disapproving their productions and identify the sources of the problem in their non-target-like productions. Different types of evidence, e.g., learners’ responses to NF, tailor-made tests, and pretest-posttest measures in experimental designs, indicate that learners can use NF, as shown by a measurable change in their linguistic behavior. Finally, there is at least one theoretical argument, together with some empirical evidence, that suggests that NF is necessary for the acquisition of certain linguistic targets.

**Factors Moderating the Effectiveness of Negative Feedback**

The focus of NF research has gradually evolved and broadened from the investigation of the effectiveness of NF in general to the investigation of factors that moderate the effectiveness of NF. The section below reviews the literature on these factors.

**Feedback Characteristics**. Two research designs have been used to explore the moderating role of feedback characteristics in the effectiveness of feedback. Researchers following the first method have used learner responses to feedback or tailor-made tests as outcome measures and asked whether different characteristics of naturally occurring feedback are related to learners’ accuracy on these measures. This research has exclusively focused on the features of recasts. For example, Sheen (2006) found that shorter recasts and recasts that presented the correct form by substituting (rather than reordering, adding, or deleting) elements of the learner’s errone­ous utterance produced higher rates of learner uptake. Loewen and Philp (2006) reported that recasts bearing stress, recasts with declarative intonation and recasts that contained only one change predicted successful uptake (i.e., repair). They also found that recasts that had inter­rogative intonation, recasts that were short, and recasts that contained only one change predicted accuracy on tailor-made posttests.

The second method involves comparing the effectiveness of at least two feedback types (e.g., recasts versus metalinguistic feedback) that are different with respect to a specific dimension (e.g., implicit-explicit) using experimental or quasi-experimental designs. Two dimensions have been identified in the literature to classify feedback types (see Table 1 for definitions and examples of feedback types). The first of these is often referred to as the implicit-explicit dimension. Feedback types that refer to the language itself have been considered explicit. They do this by satisfying either one of the following conditions or both: (a) they include metalinguistic information (i.e., whether the feedback involves the use of metalinguistic terminology or the provision of a metalinguistic rule), and/or (b) they are direct about the accuracy of the learner’s production. Directness, here, means that the propositional content is unambiguous (Blum-Kulka, 1987). Feedback types lacking metalinguistic information and/or directness have been considered implicit. Accordingly, explicit correction and metalinguistic feedback are typically classified as explicit, whereas clarification requests, recasts and repetitions are classified as implicit feedback forms. It should be added that some researchers (Ellis & Sheen, 2006; Ellis, Loewen, & Erlam, 2006; Loewen & Erlam, 2010; Lyster, Saito, & Sato, 2012) have adopted the view that recasts cannot be regarded as purely implicit forms of feedback on the grounds that the negative evidence in some recasts can potentially be more easily noticed. Although it has been empirically shown that recasts can be more or less noticeable depending on various factors (Lyster, 1998; Sheen, 2006; Loewen & Philp, 2006), in this review, all forms of recasts are considered implicit because the definition of explicitness provided above refers to the overt linguistic indicators that encode NF rather than how NF is perceived by learners.

There is a second dimension along which feedback types can differ from each other. The values of this dimension have been assigned several different labels. At one end, there are feedback types that push learners to repair their own utterances, such as elicitation, metalinguistic feedback, clarification request, and repetition, which have been labelled output-prompting (Sheen & Ellis, 2011), self-repair (Loewen & Nabei, 2007), or prompts (Ranta & Lyster, 2007). At the other end, there are feedback types that reformulate learners’ erroneous production into target-like sentences, such as explicit correction and recasts, which are labeled as input-providing (Sheen & Ellis, 2011), other-repair (Loewen & Nabei, 2007), or reformulations (Ranta & Lyster, 2007). Reformulations provide both positive and negative evidence, whereas prompts provide only negative evidence.

Table 1

*Negative Feedback Types*

|  |  |  |
| --- | --- | --- |
| Feedback Type | Definition | Example |
| Recasts | target-like reformulations of learners’ non-target-like productions | Student: It means I am not familiar about that?  Teacher: No, it doesn’t mean that you’re not familiar with that. You’re familiar with that, but don’t like it.  Student: Oh, I see.  (Han & Kim, 2008, p. 35-36) |
| Explicit Correction | explicit rejection of the learner’s production, followed by the provision of the target-like form | Student: He has catch a cold  Teacher: Not catch, caught.  (Nassaji, 2015, p. 54) |
| Clarification Requests | expressions "designed to elicit clarification of the interlocutor's preceding utterance(s)" (Long, 1983, p. 137) | Student: Who has fin?  Teacher: Sorry, I beg your pardon?  (Oliver & Mackey, 2003, p. 519) |
| Metalinguistic Feedback | comments about the accuracy of the learner’s non-targetlike production including the provision of metalinguistic terminology, clues or rules | Learner: Who does John think that is reading a newspaper?  Researcher: Don’t use the conjunction “that” when you ask about the subject  of the subordinate clause.  (Goo, 2012, p. 455) |
| Elicitation | Feedback that attempts to push learners to provide the target-like form either “by repeating the learner utterance up to the error and waiting for the learner complete the utterance… or by asking the learner more directly to reproduce the utterance such as *can you say it again*?” (Nassaji, 2015, p. 54). | Student: And when the young girl arrive, ah, beside the old woman.  Teacher: When the young girl …?  (Nassaji, 2007, p. 539) |
| Repetition | repetition of the learner’s erroneous production with rising intonation | Student: Oh my God, it is too expensive, I pay only 10 dollars.  Teacher: I pay?  Sheen (2004, p. 279) |

Previous studies investigating the relative effectiveness of feedback types that differ with respect to the implicit-explicit dimension have mostly operationalized implicit feedback as recasts. Part of this interest stemmed from the theoretical arguments made in favor of the possible effectiveness of recasts. Some researchers (e.g., Doughty, 2001; Gass & Mackey, 2006; Long, 2007) have argued that recasts constitute an ideal form of NF because they are unobtrusive, that is, they provide negative evidence without compromising the meaning-based nature of the interaction, and they do so without being face-threatening. Long (2007) points out that the negative evidence presented through recasts during conversational interaction would potentially be effective because recasts provide information about the L2 precisely when the learner needs it. It is commonly accepted that learning through recasts (and other types feedback that include positive evidence) occurs when learners are engaged in a special type of monitoring, i.e., cognitive comparison, in which they mentally compare the intention, the input and the output in their working memory (Doughty, 2001). According to Doughty (2001), since target-like and non-target-like forms are juxtaposed in a recast instance, learners can cognitively compare the non-target-like form with the target language form. In addition to these theoretical arguments, the finding of descriptive classroom studies that recasts are the most frequently occurring feedback type in language classrooms (Ellis et al., 2001; Lyster & Ranta, 1997; Panova & Lyster, 2002; Sheen, 2004) has given SLA researchers another reason to investigate their effectiveness.

Not everyone has shared these positive ideas about recasts. Some other researchers (e.g., Carroll, 2001) have argued that learners can have difficulty noticing the negative evidence presented through recasts because they are not salient enough. Carroll (2001) has advocated for more explicit forms of feedback, stating that “the best feedback and correction is probably the most explicit—which is the least likely to occur” (p. 390). According to her, feedback is more likely to be successful if it is informative with respect to whether there is an error, the location of the error, and how it should be corrected.

Several empirical studies indicated that explicit feedback, operationalized as metalinguistic feedback, was more effective than implicit feedback, operationalized as recasts (e.g., Carroll & Swain, 1993; Ellis, 2007; Ellis et al., 2006; Sheen, 2007). For example, Ellis, Loewen, and Erlam (2006), following a pretest-posttest-delayed posttest quasi-experimental design, compared a special form of metalinguistic feedback (e.g., “kiss-you need past tense”), to recasts (i.e., reformulation of the erroneous segment), and a no-feedback control groups. ESL learners received feedback on past tense -*ed* according to their group assignment during a picture-guided narrative task performed in triads. The results showed an advantage for the metalinguistic group over the recast group in the delayed posttest. Conversely, other studies showed no difference between recasts and metalinguistic feedback (Goo, 2012; Kang, 2010; Loewen & Nabei, 2007; Loewen & Erlam, 2006; Sauro, 2009). For example, Loewen and Nabei (2007) followed a quasi-experimental pretest-posttest-control group design and compared the relative effectiveness of recasts, metalinguistic feedback (i.e., learners were asked to think about the form of their utterances), and clarification requests on the acquisition of question formation. The study showed no effect for explicitness.

Much less research has been conducted to compare the relative effectiveness of feedback types that differ in directness. Yilmaz (2012) investigated the relative effectiveness of explicit correction (i.e., explicit/direct) versus recasts (i.e., implicit/indirect), using an experimental design with two post-tests, immediate and delayed. The participants, who had not been exposed to the target language, Turkish, prior to the experiment, were asked to learn 50 Turkish words in order to carry out the rest of the study. Their errors on two Turkish morphemes (i.e., locative and plural) were corrected according to their group assignment during two communication games. One of the tasks was carried out through FTFC, the other through SCMC. The results revealed that explicit correction was more effective than recasts, as measured by oral production and comprehension tasks in the posttests.

In a later study, Yilmaz (2013b) replicated the findings of Yilmaz (2012) of the advantage of explicit correction over recasts. Yilmaz (2013b) also aimed to answer whether a similar advantage could be observed for a group that was corrected, first, through explicit correction, and then through recasts. The five conditions included in his study were explicit-only (i.e., only including explicit feedback), implicit-only (i.e., recasts), reduced-explicit (i.e., matched with the mixed group in the number of explicit feedback instances), and mixed (i.e., including both explicit and implicit feedback). The results showed that the explicit-only and mixed feedback groups outperformed the other groups, with no difference between the two. These results suggest that a mixed feedback treatment can be as effective as an explicit-only feedback treatment, and that it is not necessary to correct each learner error with explicit feedback in order to obtain the desired level of effectiveness.

The second line of feedback comparison studies focused on the relative effectiveness of reformulations and prompts. Lyster and Ranta (1997) showed that, although recasts were frequent (55% of all feedback turns), they were associated with a low rate of repair (18%). Lyster (1998) attributed this to the ambiguous nature of recasts. According to him, recasts are ambiguous because they perform both corrective and non-corrective functions in discourse and can easily be confused with non-corrective repetitions. Lyster (Lyster, 1998; Lyster & Ranta, 1997) claimed that prompts would be more effective than recasts because they are less ambiguous and cognitively more engaging. Prompts receive theoretical support from the output hypothesis (Swain, 1985, 1995) and skill acquisition theory (Anderson, 1980). According to Swain’s (1985, 1995) output hypothesis, if learners are pushed to express themselves more accurately, they cannot avoid paying attention to language forms, as they are likely to do when processing language for meaning. Since prompts also push learners to retrieve and produce target language forms, they can help strengthen associations in memory (Lyster, Saito, & Sato, 2012). Skill acquisition theory suggests that L2 learning, like any other skill learning, involves gaining declarative (i.e., factual) knowledge, first, and then proceduralizing and automatizing that knowledge at later stages through practice (DeKeyser, 2007). The retrieval and production opportunities provided by prompts, according to Lyster, Saito and Sato (2012), can help learners proceduralize the knowledge they already have about language forms.

The findings of the studies comparing the differential effects of prompts versus reformulations have been mixed. Ammar and Spada (2006) and Lyster (2004) demonstrated an advantage for prompts over recasts. These two studies followed a quasi-experimental pretest-posttest-delayed posttest design. The target form was French gender assignment in Lyster (2004) and English third-person possessive determiners in Ammar and Spada (2006). Other studies (Dilans, 2010; Lyster & Izquierdo, 2009; Yang & Lyster, 2010), however, reported no difference between the groups. Dilans (2010) and Lyster and Izquierdo (2009) followed a randomized experimental design, whereas Yang and Lyster (2010) followed a quasi-experimental design. Yang and Lyster (2010) and Lyster and Izquierdo (2009) focused on morphosyntactic development, and Dilans (2010) focused on vocabulary development in English. The linguistic target in Yang and Lyster’s study (2010) was the English simple past tense, whereas in Lyster and Izquierdo (2009), it was French gender assignment. In all of these studies, learners were exposed to form-focused instructional units including explicit instruction on the target structures.

Taken together, these results suggest that not all recasts are equal in terms of their impact on learner responses and tailor-made test results. In addition, although directness seems to play a positive role in feedback effectiveness, it is not necessary to provide direct feedback on each learner error. A small amount of direct feedback, followed by indirect feedback, might be enough to obtain the desired level of effectiveness. The current empirical evidence, however, is ambiguous with respect to whether prompting learners to repair their off-target productions or providing metalinguistic information play a role in feedback effectiveness. This is an area where much future research is needed to clarify the mixed results. It is important to note that caution should be applied when interpreting the results of studies focusing on feedback characteristics because there has been considerable methodological variability among them. Studies have differed from one another on many methodological features, such as the treatment tasks employed, operationalizations of feedback types, outcome measures, target structure choice, design features (e.g., whether learners are assigned randomly to experimental groups), number of treatment sessions, duration of the treatment sessions, and consistency in providing feedback on each error.

**Linguistic Target.** Several studies have shown that the effectiveness of recasts changes depending on the linguistic target. Ortega and Long (1997) reported that recasts were more effective for Spanish adverb placement than clitic pronouns. They attributed these results to the fact that Spanish adverbs were more salient than clitic pronouns because they are multi-syllabic, stressed, and meaning-bearing. Leeman (2003) showed that Spanish learners benefited more from recasts when they were provided on Spanish number agreement than on gender agreement. She also attributed her results to the fact that Spanish number agreement is more salient in the input than gender agreement because it is semanti­cally transparent and phonetically substantial, and it always occurs in word-final posi­tion. Mackey (2006) reported that English learners were more likely to notice English question forms than the past tense morpheme after receiving interactional feedback, including recasts. She suggested that a possible reason for this was that question formation is more salient than the past tense morpheme because it involves syntactic movement. Even though these researchers invoked saliency-related arguments to explain their findings*,* they did not manipulate salience as an independent variable to investigate the relationship between salience and feedback effectiveness.

In two studies (Yilmaz, 2012; Yilmaz & Yuksel, 2011), target structure salience was included as one of the independent variables. Both studies focused on the same two Turkish structures: the plural morpheme /-lAr/and the locative case morpheme /-DA/. Salience in these studies, following Goldschneider and DeKeyser (2001), was defined as the degree to which a linguistic form can be perceived and linked to an underlying meaning. Constructs such as L1-L2 similarity, perceptual salience, and morphophonological regularity contributed to the operationalization of salience. Perceptual salience was further divided into four components: phonetic substance, syllabicity, sonority, and suffix length. A salience score was calculated for each morpheme. The comparison of these scores revealed that the plural was more salient than the locative morpheme.

In Yilmaz and Yuksel (2011), native speakers of English who did not have any Turkish background learned some Turkish vocabulary items prior to the beginning of the experiment. During the treatment, they performed one communication game for each target structure with the experimenter in which they described pictures, receiving partial recasts after their errors. The results of the study showed that learners’ performance did not vary depending on target structure. The researchers explained that an uncontrolled difference between the structures with respect to their position in learners’ production might have nullified the effect of salience (i.e., plurality was marked on the first noun, whereas location was marked on the second). Yilmaz (2012) addressed this problem by controlling for the position of the structures in learners’ production by ensuring that learners followed a specific word order. Yilmaz (2012) included another group that received explicit correction in addition to the recast group. The study revealed an effect for salience: learners scored significantly higher on the salient morpheme than on the non-salient one. The study also showed that salience did not interact with feedback type, indicating that the relative effectiveness of feedback type does not depend on target structure salience.

Overall, the studies reviewed above suggest that target structure salience determines the extent to which learners can benefit from NF. It seems that learners perform better on salient forms than on non-salient forms in posttests regardless of the explicitness (i.e., directness) of the feedback they receive. One point that is worth noting is that categorizing linguistic targets according to their salience is not the only way in which one can distinguish between different linguistic targets. Future research might use other dimensions, such as communicative value (Lightbown & Spada, 2008), rule complexity (Hulstijn & De Graaf, 1994), scope and reliability (Hulstijn & De Graaf, 1994), or frequency (Lightbown & Spada, 2008) to predict which forms are more amenable to NF.

**Learner Factors.** The moderating role of factors such as developmental readiness, proficiency level and cognitive *individual differences* (IDs) have also been examined in previous NF research.

The basic idea behind the concept of developmental readiness is that learners should be at an appropriate developmental level in order to be able to benefit from the feedback provided to them (Mackey, 2012). It is based on Pieneman’s (1998) teachability hypothesis, which states that instruction would be most effective for those structures that are from the subsequent stage in the learner’s development sequence. For example, in Mackey and Philp’s study (1998) on the acquisition of English question formation, learners who were at stage 3 were considered unready to learn a stage 5 question, whereas those learners who did not have to skip a stage to learn stage 5 questions, that is, learners who were already at stage 4, were considered “readies.” Learners were assigned into one of five groups: interactor ready, interactor unready, recast ready, recast unready, and control. The recast groups received intensive recasts while performing communicative tasks with a NS of English, whereas the interactor groups performed the same tasks but did not receive any feedback. The control group completed only the pretest and posttest. The study revealed that the recast ready group outperformed the other groups, indicating that developmental readiness matters for the extent to which learners benefit from recasts.

Another learner factor that has been found to moderate NF effectiveness is general proficiency. Ammar and Spada (2006) included proficiency as a moderator variable in their study that aimed to examine the differential effects of prompts versus recasts. They used learners’ pretest scores to determine learners’ proficiency level (low or high). The results showed that the low-proficiency prompts group outperformed the low-proficiency recast group on most of the measures. The high-proficiency prompt and high-proficiency recast groups, however, did not perform differently. This result can be taken to mean that prompts were more effective than recasts only among low-proficiency learners.

The moderating role of cognitive IDs in the effectiveness of NF has also attracted considerable attention in recent years. The goal of this research has been to determine which cognitive IDs might constitute an aptitude for learning through NF. The majority of studies contributing to this line of research have focused on two variables: *language analytic ability* (LAA) (Li, 2013; Sheen, 2007; Trofimovich, Ammar, & Gatbonton, 2007; Yilmaz, 2013a) and *working memory capacity* (WMC) (Goo, 2012; Li, 2013; Mackey, Philp, Fujii, Egi, & Tatsumi, 2002; Yilmaz, 2013a).

LAA refers to ‘the capacity to infer rules of language and make linguistic generalizations or extrapolations’ (Skehan, 1998, p. 204), considered to be a component of Skehan’s (1998) language aptitude model, along with phonemic coding ability and memory. Several studies (Li, 2013; Sheen, 2007; Trofimovich et al., 2007; Yilmaz, 2013a) have investigated the relationship between LAA and learning under various feedback conditions. Trofimovich et al. (2007) included only one type (recasts), whereas Yilmaz (2013a), Sheen (2007) and Li (2013) included two feedback types. The feedback types compared were explicit correction and recasts in Yilmaz (2013a), and metalinguistic feedback (reformulation plus metalinguistic rule) and recasts in both Sheen (2007) and Li (2013). There was great variability across studies with respect to linguistic target (Chinese classifiers in Li [2013], English articles in Sheen [2007], English possessive determiners and intransitive verbs in Trofimovich et al. [2007], and Turkish plural and locative morphemes in Yilmaz [2013a]). Conflicting findings emerged from these studies. Trofimovich et al. (2007) and Li (2013) reported a relationship between LAA and language outcomes under the recast condition. Li (2013) also showed that there was no relationship between LAA and language outcomes under the explicit feedback condition. Sheen (2007) and Yilmaz (2013a), however, found a relationship between learning and LAA under explicit feedback conditions, not under recast conditions. It is difficult to attribute this discrepancy in findings to any specific factor, since the studies differed in so many ways. Nevertheless, the type aptitude task used has some potential for explaining the differences. Interestingly, the two studies that found a relationship between learning under the recast condition and LAA, i.e., Li (2013) and Trofimovich et al. (2007), used a task (the Words in Sentences subtest of the Modern Language Aptitude Test [Carroll & Sapon, 1959]) that involves identifying the function of words and phrases in sentences, but does not involve the learning of new material. Both Sheen (2007) and Yilmaz (2013a), however, used tasks (Otto [2002] in Sheen; Llama F [Meara, 2005] in Yilmaz) that measured the extent to which learners worked out the grammar of an unfamiliar language. A future study investigating the interaction between LAA and feedback types using both types of LAA tasks might shed light on whether the relationship between a specific feedback type and LAA is moderated by LAA task type.

WMC has been another popular cognitive ID among NF researchers. Kane, Conway, Hambrick and Engle (2007) define WM as “attentional processes that allow for goal-directed behavior by maintaining relevant information in an active, easily accessible state outside of conscious focus, or to retrieve that information from an inactive memory, under conditions of interference, distraction or conflict” (p. 23). Several studies have addressed the relationship between WMC and the extent to which learners notice recasts as a source of negative evidence. The results of these studies are mixed. Two studies showed that high-WMC learners tended to report more noticing of recasts (Lai, Fei, & Roots 2008; Mackey et al., 2002), whereas another study (Trofimovich et al., 2007) showed no significant relationship.

Li (2013), Trofimovich et al. (2007), and Yilmaz (2013a), which are reviewed briefly above, investigated the moderating effect of WMC, in addition to that of LAA, on the effectiveness of NF. Goo (2012) also investigated the relationship between WMC and the effectiveness of implicit and explicit feedback types. As in Li’s (2013) study, explicit and implicit feedback types were operationalized as recasts and metalinguistic feedback, respectively. Goo’s metalinguistic feedback, however, did not include the reformulation of the learner’s error. The target structure was the English *that*-trace filter. Révész (2012) also examined the relationship between WMC and the effectiveness of feedback. Her study focused on the acquisition of the English past progressive construction through recasts. The findings are mixed for both implicit and explicit feedback. While Li (2013) and Yilmaz (2013a) found a relationship between WMC and language outcomes under their respective explicit feedback conditions (metalinguistic feedback with reformulations in Li, and explicit correction in Yilmaz), Goo (2012) did not find a similar relationship between his explicit condition (metalinguistic feedback) and WMC. The findings of Goo (2012) and Révész (2012) revealed that WMC was related to learners’ posttest performance after recasts; however, such a link was not found by Li (2013), Trofimovich et al. (2007), or Yilmaz (2013a). This short review indicates that WMC does play a role in NF effectiveness. The question of whether it plays a role in the effectiveness of a specific feedback type, however, cannot be answered satisfactorily, given the conflicting findings. Future research is needed to clarify the nature of the relationship between WMC and specific processing demands of different feedback types.

Other variables, such as attention control (Trofimovich et al., 2007), phonological short-term memory (Mackey et al., 2002; Révész, 2012), and explicit language aptitude (Yilmaz & Granena, 2015) have also been investigated, but much less intensively. For example, Yilmaz and Granena (2015) looked at whether there were any significant interactions between learners’ *explicit language aptitude* (ELA) and a subset of the feedback types analyzed in Yilmaz (2013b): explicit-only, implicit-only, and control. ELA refers to a “broader type of aptitude, capturing various explicit cognitive abilities (Granena, 2013), such as explicit inductive learning ability, explicit associative learning, and rote memory ability” (Yilmaz & Granena, 2015, p. 1). The results indicated that ELA was predictive only for the explicit-only group’s performance on the immediate posttest. This suggests that explicit cognitive processes were at work during explicit feedback and that such feedback differentially affected learners’ acquisition and processing, depending on their ELA. Since Yilmaz and Granena (2015) was the first study to use the construct of ELA in a feedback study, future work is needed to determine the extent to which these findings are generalizable.

**Other Factors.** Several experimental studies have examined the influence of other factors on the extent to which learners take advantage of feedback. These factors are related to how (feedback timing), in which context (communication mode and tasks), and under what processing conditions (exposure condition) feedback is delivered.

*Feedback Timing.* One can distinguish between at least two kinds of feedback with respect to timing: immediate and delayed. Immediate feedback is provided immediately after learners make an error, usually during a communicative task. Delayed feedback is provided at the end of the task, the end of a lesson, or several days after a lesson. In one of the few attempts to compare the effectiveness of immediate and delayed feedback, Quinn (2014) followed a pretest-immediate posttest-delayed posttest design in a lab experiment. ESL learners were randomly assigned to one of three groups: immediate feedback, delayed feedback, and no-feedback control. All learners received a mini-lesson that involved the provision of explicit information on the English passive construction. They then participated in three communicative tasks, during which they received feedback according to their group. The feedback used in the study was a hybrid type, including both a prompt and a reformulation. The learners in the immediate group were provided feedback right after their errors, whereas the learners in the delayed group received feedback at the end of each task. The results showed no effect for feedback timing. It should be noted, however, that the study did not reveal any effect for feedback, either. That is, in none of the tests, did the two feedback groups outperform the control group.

*Communication Mode.* Communication mode refers to the channel through which interlocutors convey their messages. As operationalized to date, it can take two values: oral communication (or *face-to-face communication* [FTFC]) or text-based *synchronous computer-mediated communication* (SCMC). SCMC can be defined as real-time communication between people using text-based chat tools via computers. SCMC has been considered similar to FTFC in a number of ways, such as short turns, real-time communication, and informality of discourse. Other features of SCMC differentiate it from FTFC. These include visual saliency, re-readability of messages, and longer processing time, and have been claimed to facilitate L2 development by amplifying learners’ attention to linguistic form (e.g., Smith, 2004; Warschauer, 1997). The similarities (and potentially facilitative differences) between FTFC and text-based SCMC have motivated researchers to investigate the role of NF in text-based SCMC.

Although feedback has been provided through SCMC in a few studies (Loewen & Erlam, 2006; Sauro, 2009), the research has not manipulated communication mode as a variable. Only a handful of studies have focused on the role of communication mode in the extent to which leaners notice or use different feedback types. In Gurzynski-Weiss and Baralt (2014), Spanish as a foreign language learners carried out one information gap task with one of the researchers in each of the two modes and received various types of feedback addressing their errors with lexis, morphosyntax, semantic, and phonology. Stimulated recall protocols revealed that communication mode did not make a difference to learners’ accurate perception of feedback. The study also showed that FTFC provided more opportunities for modified output than SCMC, and learners modified their output more often in FTFC than in SCMC. As explained previously, Gurzynski-Weiss and Baralt (2014) did not focus on one feedback type, and feedback was provided on a variety of target features. Other experimental studies (Yilmaz, 2012; Yilmaz & Yuksel, 2011) showed that when a specific type of feedback was provided on a predetermined linguistic target, SCMC feedback led to more learning than FTFC feedback, as measured by recognition (Yilmaz, 2012) and oral production tests (Yilmaz, 2012; Yilmaz & Yuksel, 2011). This effect did not change depending on feedback type, explicit correction or recasts (Yilmaz, 2012). The researchers attributed the favorable results for SCMC feedback to the potential for greater processing time, due to the re-readability of the error-feedback pairs.

*Task Features.* Communicative tasks have often been used as a tool for eliciting language production from learners in interaction research. The relationship between various task features and the degree to which they facilitate interactional modifications was at the center of much early interaction research. For example, two-way tasks, in which both interlocutors’ contribution is necessary for task completion, have been found to lead to more negotiation of meaning than one-way tasks, in which one of the interlocutors’ contribution is enough for task completion (Long, 1980; Pica & Doughty, 1985). Recently, researchers have started to investigate the relationships between different task features and the effectiveness of a specific interactional feature, recasts.

For example, Révész and Han (2006) carried out a lab study that focused on the moderating role of task content familiarity in the effectiveness of recasts for improving adult ESL learners’ use of the past progressive. They hypothesized that NF would be more effective when provided in a task in which learners are familiar with the content, because carrying out a task multiple times should reduce the cognitive demands of tasks and free up cognitive resources that could be used to attend to NF. In their study, learners in each experimental group completed three oral narrative tasks in which learners were asked to describe a crime scene. Some learners performed the same version of the task, whereas others performed different versions of the same task type. During the tasks, learners received feedback in the form of recasts whenever they made an error in producing the past progressive. The results showed that learners who performed the same tasks made significantly higher written and oral production gains than did those who performed different versions of the same task. This shows that when learners are familiar with the content of the task, they are more likely to take advantage of recasts.

Révész (2009) investigated the relationship between task complexity and the effectiveness of recasts. Her study followed a pretest-posttest-delayed posttest design and was carried out in a lab setting. The target structure was the English past progressive. Her four experimental groups differed from one another with respect to (a) whether they received recasts during their task performance, or (b) whether they were provided with contextual support. Tasks in which learners were *not* provided with contextual support were considered more complex. Révész hypothesized, based on the Cognition Hypothesis (Robinson, 2001), that learners would benefit more from recasts provided during this type of complex task than from recasts provided during less complex tasks. All groups were initially presented with photos of crimes scenes and later asked to describe the scenes. In the plus-contextual-support conditions, learners were asked to describe the crime scene with the aid of a photograph on the screen as they did so. In the minus-contextual-support conditions, learners could only see a blank screen. The results showed that the group lacking contextual support outperformed all other groups. This lends supportive evidence to Robinson’s (2001) claim that “more and more complex tasks may prompt learners to look for more and more help in the input, attending to facilitative forms made salient by teacher intervention […] using focus on form techniques” (p. 304).

*Exposure condition.* Exposure condition refers to whether a learner’s exposure to NF is direct or indirect. In the majority of NF studies, learners were exposed to NF after producing a non-target-like utterance. This condition is not the only possible exposure condition under which learners could experience NF. In many language-learning contexts, there is another learner who has the opportunity to observe the interaction between the learner who produced the non-target-like utterance and the interlocutor who provides the NF. In these contexts, the observing learners have a chance to experience NF indirectly. In a randomized lab experiment, Yilmaz (2015) investigated whether the effectiveness of explicit correction in the acquisition of two Turkish morphemes (plural and locative) would change depending on exposure condition. Native speakers of Mandarin, who had not been exposed to Turkish, were assigned to one of three feedback groups: receivers, non-receivers, and control. Each learner performed three communication games with the experimenter, during which their plural and locative morpheme errors were treated according to their group. The receivers’ errors were reformulated through statements indicating what learners should have said (e.g., you should say “X”). Non-receivers did not receive any feedback on their errors. They were, however, allowed to hear the NF provided to the receivers group. The control group neither received feedback nor had a chance to hear the NF provided to other learners. The study revealed that the receivers outperformed the non-receivers on the plural morpheme, but the groups were not significantly different on the locative morpheme. This result hints at a possible interaction between target structure and exposure condition, which might be attributed to the less salient nature of the locative compared to the plural.

Overall, the above studies on feedback timing, communication mode, task features and exposure condition show that each of these factors has the potential to increase the impact of NF on linguistic outcomes. It should be noted, however, that very few studies have been conducted to investigate these factors as yet. In order to validate current findings, further investigation into the moderating role of these factors is strongly recommended.

**Methodological Issues**

Recently, methodological details of previous experimental studies have been subjected to serious scrutiny. Goo and Mackey (2013), Long (2007) and Yilmaz (2012) have pointed out methodological problems with previous experimental feedback research and questioned the reliability and validity of the findings of studies with those shortcomings. Because of these problems, it has been argued, the generalizability of findings of much published research is limited.

**Modified output opportunities**. As the above review has shown, there is some evidence, although not conclusive, indicating that the production of modified output is positively related to L2 acquisition (e.g., McDonough, 2005). Research has also shown that feedback types are not equivalent regarding the frequency with which they lead to modified output (e.g., Lyster & Ranta, 1997; Sheen, 2004). For this reason, it is essential for experimental research to control for modified output opportunities so that a possible change in learners’ linguistic behavior can be attributed to feedback alone, rather than modified output or a possible interaction between modified output and feedback.

**Inconsistent operationalizations.** Goo and Mackey (2013) identified three types of single-versus-multiple comparisons, all of which they considered problematic because they would lead to an unfair advantage for one of the treatment groups. However, from a replicability perspective, which Goo and Mackey (2013) did not argue, only the first two (a and b below) can be considered problematic. In all versions of Goo and Mackey’s single-versus-multiple comparisons problem, learners in one of the conditions receive the same feedback type (e.g., recasts) in one feedback move whenever they make an error, whereas learners in the other feedback conditions receive feedback in one of the following ways. a) Every time they make an error, they receive feedback from a range of feedback types (metalinguistic feedback, clarification requests, etc.). In other words, the feedback type they receive is not fixed (e.g., Ammar & Spada, 2006). b) They receive multiple feedback moves, where the delivery of the second move depends on whether the learner responds to the first move with a repair (e.g., repetition plus metalinguistic feedback). In this version, the number of feedback moves for each error is not fixed (e.g., Dilans, 2010). c) They receive a hybrid form of feedback which is a combination of two feedback types (e.g., reformulation plus metalinguistic feedback, [e.g., Sheen, 2007]). Cases a and b would affect the homogeneity of those treatment conditions because different learners in the same condition might get slightly different feedback treatments. This methodological ambiguity could make the study unreplicable. Case c, however, does not suffer from the same problem because learners would consistently receive the same hybrid form whenever they make an error.

**Form-focused instruction**. *Form-focused instruction* (FFI) has been considered necessary for prompt conditions in previous studies because learners cannot be expected to repair their utterances if they have no familiarity with the target structure. In order to remedy this potential problem, in some previous comparison studies (e.g., Ammar & Spada, 2006; Dilans, 2010; Lyster, 2004; Yang & Lyster, 2010), all conditions were exposed to FFI that included explicit information on the target forms and focused production activities. For some researchers (Li, 2010; Lyster, Saito & Sato, 2012), the provision of FFI is not a threat to internal validity because any difference between an FFI plus feedback condition and an FFI-only condition can be attributed to feedback. Goo and Mackey (2013) do not agree with this reasoning. For them, the use of FFI in addition to feedback creates interpretation problems because one cannot attribute the observed effect to the feedback factor alone. According to them, even though instruction is provided to both groups, it is not possible to know whether instruction interacts with each feedback type in the same way. Other than Goo and Mackey’s (2013) valid criticism, the fact that FFI has often been used in quasi-experimental designs might lead to an additional problem. Since learners are not randomly assigned to treatment groups in quasi-experiments, it is probable that groups differ on variables that can interact with the effectiveness of FFI. In other words, it would not be possible to assume that different groups would take advantage of FFI to the same degree. This is another reason why incorporating this design feature could hamper the possibility of attributing the observed effect to NF. To counter these problems associated with using FFI in a NF study, researchers could use a design that involves administering a pretest after the provision of FFI, classifying learners with similar scores into subgroups, and making sure that learners belonging to these subgroups are equally distributed across treatment groups.

**Prior knowledge.** Another important limitation of previous studies, which was previously discussed by Goo and Mackey (2013) Long (2007) and Yilmaz (2012), is the lack of control for learners’ initial level of knowledge of linguistic targets. Although several methods exist to guarantee that learners have only limited knowledge of a target structure, such as screening subjects into the study using a cutoff score (e.g., Révész, 2009) or selecting a target structure from a language learners are not familiar with (e.g., Yilmaz, 2012), these methods have not been used extensively. When researchers fail to control for learners’ preexisting knowledge, the following may result: 1) one cannot reliably attribute any potential effect to the NF factor, 2) the study may be biased toward explicit forms of feedback that provide metalinguistic rules or clues, because learners can evaluate their errors, as well as interpret the feedback, in light of their metalinguistic knowledge acquired prior to the study.

**Out-of-experiment exposure.** In some of the previous studies reported in Goo and Mackey (2013), learners had chances to be exposed to the target structure of the study outside the experiment. Once again, when this happens, it is not possible to attribute observed effects to the treatment conditions alone. This threat has usually been the outcome of a study that either has more than one treatment session or includes a posttest not administered immediately after the treatment finishes. In either case, learners can gain information about the target structure through the Internet or reference books, or they can simply be exposed to positive evidence that includes the target structure, either inside or outside the classroom. This problem can be dealt with by using an artificial language specifically created for the study or by choosing a time period for the experiment in which learners are not enrolled in any language course.

**Confounding individual and group feedback.** In several classroom-based studies, learners’ posttest scores cannot be attributed to the feedback they received on their own errors because feedback was not provided only to individual learners but to a group of learners. Since there is evidence suggesting that receiving feedback on one’s own errors might be more effective than being exposed to the feedback provided on someone else’s errors, for at least some salient target structures (Yilmaz, 2015), it is important for future studies to control for the number of feedback instances learners receive individually and as a group.

**Future Directions and Conclusion**

After listing a series of methodological problems (some of which were reviewed above), Goo and Mackey (2013, p. 22) stated “the recasts-versus-other-types-of-feedback comparison is one of apples and oranges.” They explained that they do not view comparative NF studies as a fruitful line of research because of the methodological problems that are almost inevitable in such comparisons. Instead, they proposed that researchers should select one feedback type and explore the conditions under which it is more or less effective. An alternative path to the one proposed by Goo and Mackey (2013) involves continuing to design new studies to compare feedback types following the methodological guidelines provided in the theoretical and methodological papers on NF (e.g., Long, 2007; Goo & Mackey, 2013; Lyster & Ranta, 2013). Since many of the previous feedback comparison studies suffered from methodological problems, the new research would be more likely to produce more reliable and valid results and enhance understanding of the effect of learning conditions in L2 acquisition. It should be added that findings from studies in which feedback types differ on a single well-defined, theoretically-motivated dimension would be more likely to help advance theory than findings from studies in which feedback types differ from each other on multiple dimensions.

This review has shown that the effectiveness of NF is moderated by factors other than feedback characteristics (developmental readiness, proficiency, etc.). The majority of these factors, however, have been studied in only a few studies so far. Further examination of each of them is warranted, therefore, to confirm existing findings and to understand the extent to which they are generalizable to other contexts. In addition, in order to establish a body of knowledge that can eventually be used in evidence-based pedagogical practice, it is necessary to identify other factors that may be related to the effectiveness of different NF types. One variable worth investigating is feedback frequency and its relationship to the effectiveness of NF. What is the minimum frequency of feedback necessary to create a short-term and/or long-term learning effect on specific linguistic targets? The study of the role of feedback frequency can elucidate the potential for language learning in different contexts where feedback occurs naturally. A related question that can be investigated by future research is the extent to which NF should be consistent across learners’ errors. Truscott (1999) claimed that “if teachers are inconsistent in their corrections, these corrections are as likely to be harmful as they are to be helpful” (p. 440). Should a learner receive NF whenever they produce an error? Or would it be sufficient for learners to receive feedback on some of their errors? Another interesting question is whether spaced or massed provision of feedback is more effective. A well-established finding in psychology (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006) is that information is retained better when it is studied in multiple spaced intervals (spaced distribution) rather than when it is studied intensively in a single training session (massed distribution). Some L2 researchers (e.g., Lightbown, 2008) predict that form-focused L2 instruction provided over time and in small portions would be more effective than instruction provided in a single session.

Finally, a greater emphasis on cognitive IDs in future NF research could produce interesting findings that may account for how learners interpret various forms of NF. If a cognitive variable is related to learning outcomes of a given feedback treatment, this can constitute indirect evidence of how that particular feedback type was processed by the learner. For example, in Yilmaz and Granena (2015) the fact that ELA was related to outcomes in the explicit feedback group, but not in the implicit or mixed groups, was interpreted as suggesting that explicit feedback had induced learners to process language explicitly by searching for and figuring out rules. As DeKeyser (2012) argues, a given treatment may interact with a given individual difference “because the treatment variable requires a mental process that is facilitated/hampered by the value of the individual difference variable” (p. 190). This is an alternative way of understanding learning processes that otherwise would be very difficult, if not impossible, to observe.

This article has explored the role of NF in L2 acquisition. An overview was provided of the historical background of the interactionist apporach, in which the majority of NF research has its roots. Then, current emprical evidence relevant to assessment of the role of feedback was reviewed with respect to Pinker’s (1989) four conditions: 1) NF is indeed available in various contexts and in different amounts; 2) NF is usable, because learners can perceive the negative evidence conveyed through NF and identify globally the source of the linguistic problem (morphosyntax, phonology, lexis) in their production that attracts the NF; 3) NF is used by learners, as evidenced by their responses to NF, peformance on tailor-made posttests, and the results of experimental research; 4) NF could be considered necessary for certain linguistic forms that are not learnable from positive evidence alone. Many factors enhance the effect of feedback, with feedback characteristics and cognitive variables being two that have attracted scholarly attention. Some recurrent methodological problems and ways to overcome them were discussed, and ideas for future research provided.There is a need for future research to open new avenues of investigation, as well as to pursue existing lines of work, and, above all, to produce credible evidence of a causal link between the factor under investigation and the observed outcome, while ruling out alternative explanations.

References

Anderson, J. (1980). Cognitive psychology and its implications. San Francisco, CA: Freeman.

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

Blum-Kulka, S. (1987). Indirectness and politeness: Same or different? *Journal of Pragmatics*, *11*, 145-160.

Bohannon, J.N., & Stanowicz, L. (1988). The issue of negative evidence: adult responses to children's language errors. *Developmental Psychology, 24*, 684-689.

Braidi, S. M. (2002). Reexamining the role of recasts in native-speaker/nonnative-speaker interactions. *Language Learning, 52*, 194-211.

Brown, R., & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In J. Hayes (Ed.), *Cognition and the development of language* (pp. 11-53). New York: Wiley.

Carroll, J. B., & Sapon, S. (1959). *Modern Language Aptitude Test: Form A*. New York: Psychological Corporation.

Carroll, S. (2001). *Input and evidence: The raw material of second language acquisition* Philadelphia, PA: J. Benjamins.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132, 354 -380

Chun, A. E., Day, R. R. Cheneworth, N.., & Lupescu, S. (1982). Errors, interaction, and correction: A study of native-nonnative conversations. *TESOL Quarterly, 16*, 537-547.

DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 97-113). New Jersey: Lawrence Erlbaum.

DeKeyser, R. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, *62*, 189-200.

Demetras, M.J., Post, K.N., & Snow, C.E. (1986). Feedback to first language learners: the role of repetitions and clarification questions. *Journal of Child Language, 13*, 275-292.

Dilans, G. G. (2010). Corrective feedback and L2 vocabulary development: Prompts and recasts in the adult ESL classroom. *Canadian Modern Language Review, 66*, 787-816.

Doughty, C. J. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 206-257). Cambridge, UK: CUP.

Doughty, C., & Pica, T. (1986). “Information gap” tasks: Do they facilitate second language acquisition? *TESOL Quarterly*, *20*, 305-325.

Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty and J. Williams (Eds.). *Focus on form in classroom second language acquisition* (pp. 114-138). New York: CUP.

Ellis, R. (1991). Grammar teaching practice or consciousness-raising? In R. Ellis (Ed.), *Second language acquisition and second language pedagogy* (pp. 232-241). Clevedon, UK: Multilingual Matters.

Ellis, R. (2007). The differential effects of corrective feedback on two grammatical structures. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp*.* 339-360). New York: OUP.

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

Ellis, R., Basturkmen, H., & Loewen, S. (2001). Learner uptake in communicate ESL lessons. *Language Learning, 51*, 281-318.

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition*, *28*, 339-368.

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. *The* *Modern Language Journal,* *94,* 1-21.

Farrar, M. J. (1990). Discourse and the acquisition of grammatical morphemes. *Journal of Child Language, 17*, 607-624.

Ferguson, C. A. (1981). 'Foreigner talk' as the name of a simplified register. *International Journal of the Sociology of Language*, *1981*, 9-18.

Freed, B. F. (1981). Foreigner talk, baby talk, native talk. *International Journal of the Sociology of Language*, *1981*, 19-39.

Gass, S. (1997). *Input, Interaction and the second language learner*. New Jersey: Erlbaum.

Gass and Lewis (2007). Perceptions about interactional feedback: Differences between heritage language learners and non-heritage language learners. In A. Mackey (Ed.), *Conversational interaction in SLA: A collection of empirical studies* (pp. 79-99). New York: OUP.

Gass, S. M., & Mackey, A. (2006). Input, interaction and output: An overview. *AILA review, 19*, 3-17.

Gass, S. M., & Varonis, E. (1985a). Task variation and nonnative/nonnative negotiation of meaning. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 149-161). Rowley, MA: Newbury House.

Gass, S. M., & Varonis, E.M. (1985b). Variation in native speaker speech modification to non-native speakers. *Studies in Second Language Acquisition*, *7*, 37-57.

Gass, S. M., Mackey, A., & Ross-Feldman, L. (2005). Task-based interactions in classroom and laboratory settings. *Language Learning,* *55*, 575-611.

Goldschneider, J.M., & DeKeyser, R. (2001). Explaining the ‘natural order of L2 morpheme acqui­sition’ in English: A meta-analysis of multiple determinants. *Language Learning*, *51*, 1-50.

Goo, J. (2012). Corrective feedback and working memory capacity in interaction-driven L2 learning. *Studies in Second Language Acquisition*, *34,* 445-474.

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

Goo, J., Granena, G., Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning: Norris & Ortega (2000) revisited and updated. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 443-483). Amsterdam: John Benjamins.

Granena, G. (2013). Cognitive aptitudes for second language learning and the LLAMA Language Aptitude Test. In G. Granena & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 105-129). Amsterdam: John Benjamins.

Gurzynski-Weiss, L., & Baralt, M. (2014). Exploring learner perception and use of task-based interactional feedback in FTF and CMC modes. *Studies in Second Language Acquisition, 36*, 1-37.

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

Han, Z., & Kim, J. (2008). Corrective recasts: What teachers might want to know. *Language Learning Journal, 36*, 35-44.

Hatch, E. M. (1978). Discourse analysis and second language acquisition. In E. Hatch (Ed.), *Second language acquisition: A book of readings* (pp.401-435). Rowley: Newbury House.

Heath, S. B. (1983). *Ways with words*. Cambridge: CUP.

Herron, C., & Tomasello, M. (1988). Learning grammatical structures in a foreign language: Modeling versus feedback. *The French Review, 61,* 910-922.

Hulstijn, J., & DeGraaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review, 11,* 97-112.

Ishida, M. (2004). Effects of recasts on the acquisition of the aspectual form -te i- (ru) by learners of Japanese as a foreign language. *Language Learning, 54*, 311-394.

Iwasaki, J., & Oliver, R. (2003). Chat-line interaction and negative feedback. *Australian Review of Applied Linguistics*, *17*, 60-73.

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007).Variation in working memory capacity as variation in executive attention and control*.* In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, and J. N. Towse (Eds.), *Variation in Working Memory* (pp. 21-48). New York: OUP.

Kang, H. S. (2010). Negative evidence and its explicitness and positioning in the learning of Korean as a heritage language. *The* *Modern Language Journal, 94*, 582-599.

Kim, J., & Han, Z. (2007). Recasts in communicative ESL classes: Do teachers intent and learner interpretation overlap? In A. Mackey (Ed.), *Conversational interaction in SLA: A collection of empirical studies* (pp. 269-297). New York: OUP.

Koike, D., & Pearson, L. (2005). The effect of instruction and feedback in the development of pragmatic competence. *System*, *33*, 481-501.

Krashen, S. (1978). Second language acquisition. In W. Dingwall (Ed.), A survey of linguistic science, 317-338. Stamford, CT: Greylock.

Krashen, S. (1985). *The input hypothesis: Issues and implications*. Oxford: Pergamon Press.

Leeman, J. (2003). Recasts and second language development. *Studies in Second Language Acquisition, 25*, 37-63.

Lai, C., Fei, F., & Roots, R. (2008). The contingency of recasts and noticing. *CALICO Journal, 26,* 70-90.

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

Li, S. (2013). The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *The* *Modern Language Journal, 97*, 634-654.

Lightbown, P. M. (2008). Transfer appropriate processing as a model for classroom second language acquisition. In Z. Han (Ed.), *Understanding second language process*(pp. 27-44)*.* Clevedon, UK: Multilingual Matters.

Loewen, S. (2013). Uptake. In P. Robinson (Ed.), *The Routledge encyclopedia of second language acquisition* (pp. 675-677). New York: Routledge.

Loewen, S., & Erlam, R. (2006). Corrective feedback in the chatroom: An experimental study. *Computer Assisted Language Learning*, *19*, 1-14.

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* (pp. 361-377). New York: OUP.

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

Long, M.H. (1980). *Input, interaction and second language acquisition.* Unpublished doctoral dissertation, University of California, Los Angeles.

Long, M.H. (1981). Input, interaction and second language acquisition. Annals of the New York Academy Sciences, *379*, 259-278.

Long, M.H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics*, *4*, 126-141.

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

Long, M. H. (2007). *Problems in SLA.* New Jersey: Lawrence Erlbaum.

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. New York: Wiley.

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

Lyster, R. (1998). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition*, *20*, 51-81.

Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition, 26*, 399-432.

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

Lyster, R. & H. Mori (2006). Interactional feedback and instructional counterbalance. *Studies in Second Language Acquisition*, 28, 269-300.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition, 19*, 37-66.

Lyster, R., & Ranta, L. (2013). Counterpoint piece: The case for variety in corrective feedback research. *Studies in Second Language Acquisition, 35*, 167-184.

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. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition, 21*, 557-587.

Mackey, A. (2006). Feedback and noticing in second language learning. *Applied Linguistics*, *27*, 405-30.

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

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed.), *Conversational interaction in SLA: A collection of empirical studies* (pp. 408-452). New York: OUP.

Mackey, A., Gass, S. M., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition,* *22,* 471-497.

Mackey, A., Philp, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson (Ed.), *Individual Differences and Instructed Language Learning*. Philadelphia, PA: John Benjamins.

McDonough, K. (2005). Identifying the impact of negative feedback and learners’ responses on ESL question development. *Studies in Second Language Acquisition*, *27*, 79-103.

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* (pp. 323-338). New York: OUP.

Meara, P. (2005). *LLAMA language aptitude tests*. Lognostics.

Nassaji, H. (2007). Elicitation and reformulation and their relationship with learner repair in dyadic interaction. *Language Learning, 57*, 511-548.

Nassaji, H. (2010). The occurrence and effectiveness of spontaneous focus on form in adult ESL classrooms. *Canadian Modern Language Review, 66*, 907-933.

Nassaji, H. (2015). *The interactional feedback dimension in instructed second language learning: Linking Theory, Research and Practice*. New York: Bloomsbury.

Nicholas, H., Lightbown, P., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning*, 50, 719-758.

Oliver, R. (1995). Negative feedback in child NS-NNS conversation. *Studies in Second Language Acquisition, 17*, 459-481.

Oliver, R. (2002). The patterns of negotiation for meaning in child interactions. *The* *Modern Language Journal*, *86*, 97-111.

Oliver, R. (2000). Age differences in negotiation for meaning in child interactions. *Studies in Second Language Acquisition, 17*, 459-481.

Oliver, R., & A. Mackey (2003). Interactional context and feedback in child ESL classrooms. *The Modern Language Journal, 87*, 519-543.

Ortega, L., & Long, M. H. (1997). The effects of models and recasts on the acquisition of object topicalization and adverb placement in L2 Spanish. *Spanish Applied Linguistics, 1*, 65-86.

Otto, L. (2002). Magyar Egyseges Nyelverzekmero Teszt. Unpublished material.

Panova, I. & R. Lyster (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly*, *36*, 573-595.

Philp, J. (2003). Constraints on “noticing the gap”: Nonnative speakers noticing of recasts in NS-NNS interaction. *Studies in Second Language Acquisition*, 25, 99-126.

Pica, T. (1987). Second-language acquisition, social interaction, and the classroom. *Applied Linguistics*, *8*, 3-21.

Pica, T. (1988). Interlanguage adjustments as outcome of NS-NNS negotiated interaction. *Language Learning, 38*, 45-73.

Pica, T. (1991). Classroom interaction, participation, & comprehension: Redefining relationships. *System*, *19*, 427-452.

Pica, T. (1994). Research on negotiation: What does it reveal about second language learning conditions, processes and outcomes? *Language Learning, 44,* 493-527.

Pica, T. (2005). Classroom learning, teaching, and research: Atask-based perspective. . *The Modern Language Journal, 89*, 339-352.

Pica, T., Holliday, L., Lewis, N., & Morgenthaler, L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition*, *11*, 63-90.

Pica, T., Young, R., & Doughty, C. (1987). The impact of interaction on comprehension. *TESOL Quarterly*, *21*, 737-758.

Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Philadelphia, PA: John Benjamins.

Pinker, S. (1989). *Learnability and cognition*. Cambridge, MA: MIT Press.

Pinker, S. (1994). *The language instinct*. New York, NY: Harper Perennial Modern Classics.

Porter, P. (1986). How learners talk to each other: Input and interaction in task-centered discussions. In R. R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 200-222). Rowley, MA: Newbury House.

Quinn, P. (2014). *Delayed versus immediate corrective feedback on orally produced passive errors in English* (Unpublished doctoral dissertation). University of Toronto, Canada.

Ranta, L. & R. Lyster (2007). A cognitive approach to improving immersion students’ oral language abilities: The awareness-practice-feedback sequence. In R. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 141-160). Cambridge, UK: CUP.

Révész, A. (2009). Task complexity, focus on form, and second language development. *Studies in Second Language Acquisition*, *31*, 437-470.

Révész, A. (2012), Working memory and the observed effectiveness of recasts on different l2 outcome measures. *Language Learning, 62,* 93-132.

Révész, A. & Han, Z. (2006). Task content familiarity, task type, and efficacy of recasts. *Language Awareness, 15*, 160-179.

Robinson, P. (2001). Task complexity, cognitive resources, and syllabus design: A triadic framework for investigating task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287-318). New York: CUP.

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

Saito, K. (2013). The acquisitional value of recasts in instructed second language speech learning: Teaching the perception and production of English /ɹ/ to adult Japanese learners. *Language Learning, 63*, 499-529.

Saito, K., & R. Lyster (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /r/ by Japanese learners of English. *Language Learning 62*, 595-633.

Sauro, S. (2009). Computer-mediated corrective feedback and the development of L2 grammar. *Language Learning and Technology, 13*, 96-120.

Saxton, M. (1997). The contrast theory of negative input. *Journal of Child Language, 24*, 139-161.

Saxton, M. (2000). Negative evidence and negative feedback: Immediate effects on the grammaticality of child speech. *First Language, 20*, 221-252.

Saxton, M, Backley, P., & & Gallaway, C. (2005). Negative input for grammatical errors: Effects after a lag of 12 weeks. *Journal of Child Language*, *32*, 643-672.

Saxton, M., Kulcsar, B., Marshall, G., & Rupra, M. (1998). The longer-term effects of corrective input: an experimental approach. *Journal of Child Language, 25,* 701-721.

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

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge, UK: CUP.

Schmidt, R., & Frota, S. (1986). Developing basic conversation ability in a second language: A case study of an adult learner. In R. Day (Ed.), Talking to learn: *Conversation in second language acquisition* (pp. 237-326). Rowley, MA: Newbury House.

Schwartz, B. D. (1993). On explicit and negative evidence effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition,* *15,* 147-163.

Sheen, Y., (2004). Corrective feedback and learners’ uptake in communicative classrooms across instructional settings. *Language Teaching Research, 8*, 263-300.

Sheen, Y. (2006). Exploring the relationship between characteristics of recasts and learner uptake. *Language Teaching Research, 10*, 361-392.

Sheen, Y. (2007). The effects of corrective feedback, language aptitude, and learner attitudes on the acquisition of English articles. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 301-322). New York: OUP.

Sheen, Y., & Ellis, R. (2011). Corrective feedback in L2 teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 593-610). New York: Routledge.

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

Smith, B. (2004). Computer-mediated negotiated interaction and lexical acquisition. *Studies in Second Language Acquisition, 26*, 365-398.

Spada, N., & Lightbown, P. M. (2008). Form-focused instruction: Isolated or integrated? *TESOL Quarterly, 42*, 181-207.

Strapp, C. M., Bleakney, D. M., Helmick, A.L., & Tonkovich, H. M. (2008). Developmental differences in the effects of negative and positive evidence. *First Language, 28*, 35-53.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp.235-253). Rowley, MA: Newbury House.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honor of H.G. Widdowson* (pp. 125-144). Oxford: OUP.

Swain, M., & Lapkin, S. (1982). *Evaluating bilingual education: A Canadian case study*. Clevedon, Avon: Multilingual Matters.

Swain, M., & Lapkin, S. (1986). Immersion French at the secondary level: The "goods" and the "bads". *Contact, 5*, 2-9.

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *The* *Modern Language Journal, 82*, 320-337.

Trahey, M., & White, W. (1993). Positive evidence and preemption in the second language classroom. *Studies in Second Language Acquisition, 15*, 181-204.

Trofimovich, P., Ammar, A., and Gatbonton, E. (2007). How effective are recasts? The role of attention, memory, and analytical ability. In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp.144-171). Oxford: OUP.

Truscott, J. (1999). What's wrong with oral grammar correction. *Canadian Modern Language Review*, *55*, 437-456.

VanPatten, B. (2013). Mental representation and skill in instructed SLA. In J. Schwieter (Ed.), *Innovations in SLA, Bilingualism and Cognition: Research and Practice.* Amsterdam: John Benjamins.

Warschauer, M. (1997). Computer-mediated collaborative learning: Theory and practice. *The Modern Language Journal, 81*, 470-481.

White, L. (1991). Adverb placement in second language acquisition: Some positive and negative evidence in the classroom. *Second Language Research, 7*, 133-161.

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

Yilmaz, Y.(2012). The relative effects of explicit correction and recasts on two target structures via two communication modes. *Language Learning, 62,* 1134-1169.

Yilmaz, Y. (2013a). Relative Effects of Explicit and Implicit Feedback: The Role of Working Memory Capacity and Language Analytic Ability. *Applied Linguistics*, *34*, 344-368.

Yilmaz, Y.(2013b). The relative effectiveness of mixed, explicit and implicit feedback. *System, 41*, 691-705.

Yilmaz, Y. (2015). The effectiveness of explicit correction under two different feedback exposure conditions. *Studies in Second Language Acquisition.* Advance online publication, doi:10.1017/S0272263115000212.

Yilmaz, Y., & Granena, G. (2015). The role of cognitive aptitudes for explicit language learning in the relative effects of explicit and implicit feedback. *Bilingualism: Language and Cognition*. Advance online publication, doi:10.1017/S136672891400090X.

Yilmaz, Y., & Yuksel, D. (2011).Effects of communication mode and salience on recasts: A first exposure study. *Language Teaching Research*, *15*, 457-477.