Applied Psycholinguistics, page 1 of 29, 2012
doi:10.1017/S0142716412000550

Learning grammatical gender: The use of rules by novice learners

NORA PRESSON and BRIAN MACWHINNEY
Carnegie Mellon University

NATASHA TOKOWICZ
University of Pittsburgh

Received: June 19, 2011   Accepted for publication: March 20, 2012

ADDRESS FOR CORRESPONDENCE
Nora Presson, Learning Research and Development Center, University of Pittsburgh, 3939 O'Hara Street, Pittsburgh, PA 15213. E-mail: presson@pitt.edu

ABSTRACT
Two experiments successfully trained novices to categorize French nouns by grammatical gender, resulting in high levels of performance after delay. Training with a frequent exemplar and training with a more diverse exemplar set led to equivalent learning. However, providing explicit rules with correctness feedback led to better generalization and retention than did correctness feedback alone or feature focusing without explicit rule information. This suggests that, at least for some grammar tasks, explicit information about form–function mappings improves learning. Moreover, the advantage of rule instruction was robust to testing and training under time pressure. Thus, rule instruction may be helpful even when speeded performance is required, supporting the prediction that practice leads to proceduralization of declarative grammatical knowledge.

Learning the grammar of a new language is a relatively difficult task for adults (DeKeyser & Larson-Hall, 2005). Frequently, the structures to be learned involve complex and conflicting patterns that are low in salience and transparency. To deal with these problems, classroom instruction often provides direct instruction in grammatical rules. Meta-analyses of the literature on the teaching of grammar in the classroom context (Norris & Ortega, 2000; Spada & Tomita, 2010) have shown that explicit rule instruction improves grammatical performance. However, within this literature, it is often unclear whether these effects arise from explicit rule presentation or just from directing learners' attention to the relevant structures (Ellis, Loewen, & Erlam, 2006; Sanz & Morgan-Short, 2004). In the current study, we seek to separate out the effects of these two instructional treatments.

Another limitation of the existing literature is that it is difficult to measure the exact effects of explicit instruction in groups of classroom learners. This is partly because learners within a given class are often on very different levels. Moreover, learners may receive additional input that overlaps with the experimental instruction in unknown ways. For these and other reasons, researchers have begun to look at the learning of grammatical structures by complete novices outside of the classroom context (Ellis & Sagarra, 2010; Gullberg, Roberts, Dimroth, Veroude, & Indefrey, 2010; Rast, 2010; Robinson, 2010).

The current study consists of two experiments that use novice learners to examine how and when explicit instruction facilitates learning of grammatical categories. The first tests the effect of explicit instruction compared to practice alone and examines whether the breadth of exemplars changes that effect. In the second, we test whether the positive effects of explicit instruction can be attributed to simply focusing learner attention on the relevant structures.

As our framework for understanding the effects of rule instruction, we use the unified competition model (MacWhinney, 2011). In this model, the mapping of a given linguistic form (or cue) to a given function (or category) is called a *form–function mapping*. Language processing is viewed as a competition between alternate form–function mappings on the levels of lexicon, syntax, phonology, and meaning. When a form–function mapping is consistent, it is said to be high in *cue reliability*. If the mapping is both reliable and reasonably frequent, it is said to be high in overall *cue validity*, where validity is influenced by both reliability and the *availability* (or salience) of the cue. Cue validity is important in second language (L2) learning, particularly when cues cannot easily be directly transferred from the L1 to the L2 (MacWhinney, 2011).

Given the importance of the strength of form–function mappings in predicting the difficulty of L2 grammatical structures (DeKeyser, 2005; Ellis, 2002), explicit rule instruction that makes form–function mappings maximally reliable, transparent, and simple should improve learning. The theory predicts that, if low salience or lack of consistency in grammar leads to more difficult learning, then explaining those patterns simply and clearly without forcing the student to consider complex interlocking patterns of cues and other rules (MacWhinney, 1997) should reduce that difficulty. To the degree that explicit instruction directs the learner's attention to these simple, clear, and reliable cues, it benefits learning and performance (de Graaff, 1997; Hulstijn, 2001; Hulstijn & Laufer, 2001; MacWhinney, 1997; Pica, 1985; Robinson, 2005). In the unified competition model, explicit cue articulation through rules produces cue focusing, enabling better cue representation (Merriman, 1999). Focusing can be accomplished with explicit instruction in cue mappings. Cue focusing allows older learners to overcome the forces of entrenchment and transfer that can otherwise impede adult L2 learning (MacWhinney, 2011). According to the competition model, explicit rules serve a primarily transitional role, because they eventually give way to implicit control through proceduralization. The model holds that repeated exposure to reliable cues can produce memory consolidation in accord with the principle of graduated interval recall (Lindsay & Gaskell, 2010; Pavlik, Presson, & Koedinger, 2007; Pimsleur, 1967; Wittenberg, Sullivan, & Tsien, 2002).

One possible mechanism underlying the benefit of explicit instruction is *noticing* (Schmidt, 1993) or detecting cues in the input. Schmidt (1993) and others (e.g., Ellis, 1994; Robinson, 1995) have argued that acquisition of an element in L2 depends on learner attention to that element. Although it can happen spontaneously, noticing can also be triggered by various instructional and interactional contexts (Lyster, 2004a; Mackey, 2006). In the context of learning French gender in the classroom, both prompts (essentially giving correctness feedback) and recasts (providing the correct form itself) have been shown to improve learning equivalently (Lyster & Izquierdo, 2010), suggesting that both positive and negative evidence can lead to noticing. The theory of noticing supports the prediction that rule instruction that both focuses learners' attention on the rule and provides correctness feedback will produce better learning and retention compared to correctness feedback alone. However, we hypothesize that explicit statements of cue mapping rules can improve learning beyond noticing alone. Experiment 2 attempts to dissociate the positive effects of directing learner attention to relevant parts of the input from the positive effects of presenting an explicit rule.

Another mechanism that could support the formation of strong grammatical representations is generalization from a frequent or prototypical example to other forms through the process of analogy (e.g., Gentner & Medina, 1998) or example-based learning (Anderson & Fincham, 1994). Accounts of first language (L1) acquisition (Goldberg, 2006; Ninio, 1999) have emphasized the role of "leading forms" in the consolidation of grammatical categories and constructions. For example, Goldberg (1999) argues that children's learning of the double-object construction (*Tom mailed Bill the book*, where the indirect object, *Bill*, and the direct object, *book*, are adjacent) is centered around initial learning based on the high-frequency verb *give*. Benefiting from the consolidation of the construction around this core exemplar, learners then generalize the construction to additional transfer verbs such as *mail*, *send*, or *deliver*.

In the competition model (Bates & MacWhinney, 1982; MacWhinney, 2011), high-frequency forms that are closer to the center of a category promote higher levels of cue activation. This effect demonstrates itself in early syntactic acquisition, when both children (Chan, Lieven, & Tomasello, 2009) and L2 learners (Kilborn & Ito, 1989) depend on prototypical patterns to interpret grammatical relations. The unified competition model treats generalization effects in terms of a hierarchy of cue types, from item-based syntactic frames up through feature-based frames (MacWhinney, in press). The model holds that high-frequency exemplars can form centers of coherent correlation (McClelland & Rogers, 2004) around which new exemplars can cluster (Ellis, 2002; Lewis & Elman, 2001; Robinson, 2005). In Experiment 1, we test the efficacy of this possible mechanism of grammatical category learning by including a comparison between learning with a broad diversity of examples and learning in which a single example occurred frequently, along with other, similar forms that occurred less frequently.

Although exposure to high-frequency exemplars may improve learning, it is also possible that it could have a negative effect, particularly in the context of L2 instruction. In this context, each repetition of a single, frequent exemplar sacrifices a chance to expose the learner to a different exemplar. Learning might be fostered more by greater diversity of exemplars (Posner & Keele, 1970) than by consolidation about a highly frequent exemplar.

We also predict an interaction between explicit rule instruction and exposure to a highly frequent exemplar, such that a high-frequency exemplar may help, but only in the absence of explicit rule instruction. According to analyses such as those provided by Armstrong, Gleitman, and Gleitman (1983), rule-based category formation is not sensitive to the frequency of category members, meaning that a high-frequency exemplar should not improve learning in the context of explicit rule presentation. In the absence of rule instruction, however, learners can create the same categories described in rules through exemplar exposure (this process is modeled in the categorization literature). Providing a highly frequent exemplar is one way of centering the category by creating a salient leading form. Goldberg, Casenhiser, and Sethuraman (2004) showed that children's verb learning (in implicit context) initially generalizes from high-frequency exemplars, and Elio and Anderson (1984) found that implicit but not explicit learning conditions benefited from highly frequent exemplars. Thus, in the absence of rule instruction, exposure to a highly frequent exemplar should help to produce generalization to novel examples. However, in the context of rule instruction, the use of highly frequent exemplars may not help performance if those highly frequent exemplars lead to improvement in the same cue-focusing process as explicit instruction.

Our discussion so far has focused on achieving accuracy in grammatical decisions. However, the mastery of grammatical forms can also improve the fluency of communication in real-world L2 use. For example, if learners are confident about the gender of a given French noun, they will produce *un* or *une* without wasting valuable processing time. Thus, like all skill-learning domains, as accuracy improves, we expect that response latency will decrease with an eventual floor late in learning. Decreased response latency is important because real-world language production and comprehension occur under the time pressure of conversational turn taking, and robustness to time pressure is one marker of proficient performance.

Finally, it is important to ask whether training leads to long-term retention and speedup, because short-term performance is rarely the goal of language instruction. Successful encoding is related not only to immediate performance but also to less forgetting (Cahill et al., 1996; Ericsson, Chase, & Faloon, 1980). The more strongly encoded the representation, the more likely it will continue to be used correctly over time. Therefore, Experiment 1 examined learning and retention across three posttests: an immediate posttest, a delayed posttest (3 weeks later), and a long-term follow-up (3 months after training).

In these experiments, we examined the learning of grammatical gender mappings for French nouns. In French, nouns are typically either masculine or feminine. In addition, there are about 50 nouns that appear in both genders but with different meanings (L'Huillier, 1999, pp. 401–402; Price, 2008, pp. 75–77). For example, *physique* is masculine when it refers to a person's physique and feminine when it refers to the science of physics. In addition, the names of some (but not all) professions such as *president*, "president," or *auteur*, "author," can be either feminine or masculine, depending on whether they refer to a woman or a man. Compounds take the gender of the head noun, except for combinations of nouns with verbs and prepositions, which are often masculine even if the head noun is feminine (Price, 2008, pp. 72–75). Morphological endings such as *–esse* or *–tion* provide highly deterministic cues to gender, as long as they are truly independent morphemes. However, when these endings appear as simple phonological terminations, they are not reliable gender cues (MacWhinney, 2000). For example, *conversation*, "conversation," is feminine, because it ends with the morpheme

–tion, whereas *bastion*, “bastion,” can be masculine, because the final *–tion* is not a separate morpheme.

Native speakers will often agree on the hypothetical gender of novel words, even when they are not told the meanings of these new words (Tucker, Lambert, & Rigault, 1977). This is possible because native speakers rely heavily on cues associated with word endings. Cues can be formulated in terms of phonology (Lyster, 2006), orthography (Holmes & Segui, 2004), or morphology (Tucker et al., 1977). In practice, these multiple ways of representing gender cues overlap, because orthographic contrasts are usually reflected in phonological contrasts, and derivational morphemes could be interpreted in terms of their phonological constituents. For example, words ending in *–ie* tend to be feminine, and this cue is both audible and visible. However, there are some cues, such as silent final consonants, that are clear only in the orthography. For example, words ending in *–t* tend to be masculine, but the *–t* is not pronounced.

When judging the gender of a noun, literate adult learners have access to both phonological and orthographic representations (Perfetti, Bell, & Delaney, 1988). We chose to use orthographic representations in the current study for three reasons. First, the presentation of a written noun with its gendered article matches the common real-world presentation of new L2 vocabulary in textbooks. Second, novice learners have difficulty comprehending some of the sounds of spoken French (Ducroquet, 1979; Jafarpur & Yamini, 1993). Third, as Lyster (2006, p. 74) notes, “Orthographic representations surpass phonological representations in terms of predictive value.”

Despite the high reliability of some cues, there are inevitably exceptions to even the most reliable rules. For example, nouns that end in *–age* are generally masculine, but only when the *–age* is a derivational suffix. These irregularities lead to ongoing problems mastering French noun gender in even advanced L2 learners (Bartning, 2000; Dewaele, 1994; Dewaele & Véronique, 2001; Hardison, 1992). In the context of L2 learning, learning regularities and irregularities may involve different processes (Ullman, 2001a, 2001b). Because reliably using grammatical gender is difficult even for advanced L2 learners, and because our interest is in novice learners, the current experiments focus on mastering the regular application of the most reliable rules. These rules have practical utility for French instruction. For example, the rules presented in the current experiments correctly predict the gender of 85% of the 254 nouns in the online Beginning French course at Carnegie Mellon University. We taught these patterns (without exceptions) using computer-controlled instruction to college students who had not yet learned any French. This methodology has the advantage of providing a clear control over instructional variables and the sequence of instructional trials. Although this instruction is not situated in the context of actual classroom learning, it is similar in many ways to the presentation of new vocabulary items in textbooks.

Experiment 1 addresses two research questions: does explicit instruction improve naive learner categorization more than correctness feedback alone, and does providing a central, highly frequent exemplar accomplish the same effect? To that end, we compare the effects of explicit rule presentation plus correctness feedback to those of correctness feedback only within a sample of adult novice learners training to categorize nouns in French by grammatical gender, comparing performance at pretest to immediate posttest and performance after a delay. If rule presentation leads to greater learning, as it has in previous experiments, it will support the prediction that simple statements of relevant cue mappings can lead to more effective grammatical category acquisition, both by improving the transparency of relevant mappings in the input and the strength of learner cue representations and by directing learner attention to relevant cues. However, evidence that rules might impair performance would support Krashen's (1994) argument that explicit processing of grammatical category information harms future performance because of the implicit nature of L1 grammatical processing. In addition, Experiment 1 tests whether the use of a high-frequency exemplar will impact the advantage of rule instruction. We predict that, in the absence of rule instruction, exposure to a highly frequent exemplar should help to produce generalization to novel examples. However, when combined with rule instruction, the use of highly frequent exemplars may not show this effect.

EXPERIMENT 1

Participants

Participants were 41 (20 female, 21 male) students and community members recruited from online advertisements at Carnegie Mellon University. Two dropped out before the immediate posttest, 1 dropped out at the delayed posttest, and 14 did not complete the follow-up 3 months after the end of training. Only participants who completed at least one posttest were included in the analysis ($N = 38$). Exclusion criteria were any prior experience with French or more than 4 years of training or self-labeled fluency in any Romance language. Eleven participants had studied at least one semester of Spanish and 5 had studied at least one semester of Latin in high school.

Method

Materials. We used a computerized tutorial for instruction in French grammatical gender. Learners were asked to categorize French nouns by gender, using the information in the endings and their memory for trained words. To maximize the relevance of the training, we used the 24 orthographic cues (varying in length in terms of number of letters) with the highest validity and frequency, as determined by our processing of the Lexique lexical database (http://www.lexique.org). Our results align with similar cue analyses by Lyster (2006), Grevisse and Goosse (2007), and Walker (n.d.). For each of these candidate cues, we tracked all occurrences in the database and determined the frequency of the cue and its reliability as a predictor of the most probable gender. Following Lyster (2004a), we set a reliability cutoff of 90%. After using this cutoff to form a pool of highly reliable cues, we selected the 24 cues with the highest frequency. Many of the cues involved derivational markings that can also be found in English, (e.g., *–tion* or *–isme*). However, because English has no grammatical gender, there cannot be any general transfer of gender predictions from English to French.

For each of the 24 cues, we selected 14 training words and 4 testing words, for a total of 432 French stimulus nouns. Words to represent these cues were selected first from the Beginning French curriculum at Carnegie Mellon University and the University of Pittsburgh. In cases in which there were fewer than 14 examples in the curricula, we choose the remaining words from the Lexique word list. In these cases, the words would be unfamiliar to a beginning learner. However, given that no participant had any prior experience with French, lack of familiarity would be expected across stimuli, and familiarity is not needed for tests of the instructional conditions. No proper nouns were included, and none of the words had inherent biological gender, with the exception of the suffixes *–nne* and*–esse*, both of which attached to nouns of inherent feminine biological gender. Of the 432 nouns, 2 words (*musée*, "museum," and *fin*, "end") were incorrectly categorized as masculine, in accord with the *–ée* and *–in* cues, although they are actually feminine. Removing these items from the data does not change the results of any analyses. All cues and items are listed in Appendix A.

All materials were presented in a web-delivered Java-based program based on the Fact and Concept Training System (Pavlik, Presson, Dozzi, et al., 2007). All learning trials required a response, and for each participant, six cues were randomized to each of four within-subjects conditions (described below). Order of word presentation was random within each training block (each block contained two conditions).

Design. The experiment used a within-subjects repeated measures design with two crossed factors of diversity (frequent vs. diverse exemplars) and instructional explicitness (rule instruction plus feedback vs. feedback only). All participants experienced all four conditions: frequent exemplar with rule instruction, frequent exemplar with feedback only, diverse exemplars with rule instruction, and diverse exemplars with feedback only. Cues were randomly assigned to training conditions for each participant. Because the presence of the rule could influence the feedback only trials, each training session was split into two blocks: (a) rule instruction with feedback and (b) feedback only. Diversity conditions were present in each block. The order of the training blocks and thus the order of rule instruction and feedback only training was counterbalanced across participants and kept constant for each participant across training sessions.

Procedure. Participants had three training periods and took four tests. The four tests were a pretest, an immediate posttest, a 3-week delayed posttest, and a 3-month delayed posttest. The three training sessions were spaced 1 week apart. In the first week, participants completed the pretest with items from each cue that would be instructed in training and then began the first training period. In the second week, subjects continued training and were not tested. In the third week, subjects completed the training and then took an immediate posttest. Each training session took 20–35 min. The immediate posttest was a generalization task that used untrained items and did not include feedback. Three weeks after training had finished, participants took the delayed posttest with the same untrained items and again with no feedback given. Participants were compensated for each visit, with a bonus for participating in all four sessions. Between 2 and 3 months after the delayed posttest, we contacted the participants, asking them to take an additional posttest, again without feedback. Of the 38 participants who completed all three training sessions and the delayed posttest, 24 participated in this follow-up. All tests consisted of 96 words, 4 exemplars of each cue.

A training session consisted of 336 rule instruction trials and 336 feedback only trials, based on 28 presentations of each of the 24 cues in each training session (14 Exemplars $\times$ 2 Repetitions for the varied exemplar condition or 7×2 Repetitions + 1×14 Repetitions for the frequent exemplar condition). Across all three training sessions, there were a total of 2,016 trials; and the pretest and three posttests combined involved a total of 384 no feedback trials.

On each trial, participants were shown a noun in French, with its English translation. The subjects' task was to press M for masculine words and F for feminine words. During training, this response was followed by correctness feedback. In pretests and posttests, no feedback was given.

Each training session was divided into two blocks. In the feedback only block, the only feedback was *correct* (a check mark) or *incorrect* (an X; the correct response, M or F, was also displayed). The rule instruction block was identical on correct trials. However, on incorrect trials, the computer also displayed a prompt guiding participants' attention to the orthographic cues at the ends of words that predict grammatical gender. The cues appeared on the computer screen in this form: *–é* → M. This meant that a final long *–é* on a word was a cue to masculine gender. In some cases, the cues were trumped by other cues of greater specificity. In the given case, the full cue appeared on the computer screen as: *–é* → M (but *–té* → F). Cues were randomly assigned across blocks, so that no cue was trained in both blocks.

Within each block, another manipulation compared the effect of a frequent exemplar to the effect of diverse exemplars. For half of the cues, 14 equally frequent exemplars were shown in training; for the other half, 8 exemplars were used, with 1 (the frequent exemplar) occurring seven times more frequently and the other 7 exemplars equally frequent (and with the same number of occurrences as each exemplar in the 14-exemplar group).[1] Thus, the overall design was a 2×2 within-subjects design with 24 cues. In training, 12 cues were randomized to the rule instruction block and 12 were randomized to the feedback only instruction block. Half (6) of the cues in each block were randomized to the frequent exemplar condition and half to the diverse exemplar condition.

Results

Descriptive data. There was substantial improvement from pretest to all posttests. This is important, because differences between instructional manipulations are most meaningful in the context of successful learning. Table 1 shows the average d', accuracy, and latency (for correct responses only) for the instructional explicitness and diversity conditions at each time point.

PRETEST. As shown in Table 1, overall pretest accuracy was 62.7% (*SE* = 0.011), which is significantly above the chance correctness rate of 50% (95% confidence interval for mean pretest accuracy = 0.598, 0.648). Appendix A displays the pretest accuracies for each of the 24 cues. These results are similar to findings

Table 1. *Results from Experiment 1 on the effects of rule instruction, correctness feedback (only), frequent exemplar, and diverse exemplar training conditions*

	Time of Generalization Test							
	Pretraining		Immediate Posttraining		3-Week Delay Posttraining		3-Month Follow-Up	
	d′							
Training Condition	*d′*	*SE*	*d′*	*SE*	*d′*	*SE*	*d′*	*SE*
Rule instruction	0.59	0.10	2.74	0.11	2.39	0.13	2.08	0.14
Feedback only	0.79	0.11	2.45	0.13	1.88	0.11	1.77	0.19
Grand mean	0.69	0.08	2.60	0.10	2.14	0.10	1.95	0.12
	Accuracy							
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Rule instruction	0.607	0.016	0.915	0.017	0.870	0.022	0.829	0.021
Feedback only	0.647	0.016	0.889	0.017	0.821	0.016	0.804	0.027
Grand mean	0.627	0.011	0.902	0.014	0.845	0.013	0.816	0.018
	Latency (ms)							
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Rule instruction	1902	96	1061	53	1107	56	1227	85
Feedback only	1931	88	1125	63	1145	53	1243	89
Grand mean	1916	90	1093	57	1126	53	1235	86
	d′							
	d′	*SE*	*d′*	*SE*	*d′*	*SE*	*d′*	*SE*
Frequent exemplar	0.85	0.10	2.60	0.10	2.10	0.11	1.80	0.18
Diverse exemplars	0.52	0.11	2.58	0.13	2.19	0.12	2.10	0.17
Grand mean	0.69	0.08	2.60	0.10	2.14	0.10	1.95	0.12
	Accuracy							
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Frequent exemplar	0.660	0.014	0.901	0.013	0.841	0.015	0.809	0.019
Diverse exemplars	0.594	0.014	0.904	0.017	0.850	0.019	0.823	0.027
Grand mean	0.627	0.011	0.902	0.014	0.845	0.013	0.816	0.018
	Latency (ms)							
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Frequent exemplar	1917	90	1106	59	1136	55	1236	91
Diverse exemplars	1916	94	1080	58	1116	54	1234	85
Grand mean	1916	90	1093	57	1126	53	1235	86

for naive learners by Carroll (2005). Taking a 0.75 accuracy level as a rough cutoff, participants showed some pretest awareness of Cues 3, 7, 11, 12, and 17 as indicators of masculine gender and of Cues 16 and 19 as indicators of feminine gender. There were only nine nouns (all from Cues 16 and 19) with fixed biological gender. Therefore, these patterns appear to reflect some initial association of final *–n* with masculine and final *–e* with feminine. Supportive evidence of this is that Cue 22 with final *–n* was interpreted in the wrong direction 78% of the time and Cue 14 with final *–e* was in the wrong direction 64% of the time.

Given these preexisting (i.e., pretest) biases, we accounted for response bias in this binary classification task by using d' as our dependent measure of accuracy (calculated from the average percent correct and based on a chance response rate of 50%). The d' ranges from approximately –6.0 to 6.0, with 0 indicating naive performance, and d' above 1.0 a common threshold for inferring categorization. The d' value was calculated for all participants except those who had at least one condition randomized to all feminine or all masculine cues, making d' impossible to calculate ($N = 8$).[2] Mean d' at pretest was 0.69 ($SE = 0.08$), reflecting better than chance performance but with incomplete discrimination ability (confirming that these untrained participants do not have command of French gender).

POSTTEST ACCURACY: OVERALL LEARNING AND FORGETTING. After training, the mean d' across conditions at immediate posttest was 2.60 ($SE = 0.10$), which is well above the threshold for category discrimination. After a 3-week delay, mean d' was at 2.14 ($SE = 0.10$). For the 24 participants who completed all three posttests, mean d' at the 3-month follow-up was 1.95 ($SE = 0.12$), still indicating a high level of retention.

TEST OF LEARNING AND EFFECTS OF THE INSTRUCTIONAL CONDITIONS. To test our experimental predictions, we conducted a 3 (Test Time; pretest/immediate/3-week delayed posttest) $\times$ 2 (Feedback Only/Rule Instruction) $\times$ 2 (Frequent Exemplars/Diverse Exemplars) within-subjects analysis of variance (ANOVA). The average d' values for each level of rule instruction are shown in Table 1. The main effect of test time, F (2, 74) = 217, $p < .0001$, reflects the significant learning after training. In order to confirm that learning occurred in each instructional condition, we compared pretest d' to immediate posttest d' and to 3-week delayed posttest d'. We found significant learning immediately after training for all instructional conditions, *t*s (38) = 11.54–15.69, all *p*s < .001 (see Table 1 for means). Similarly, we found significant retention at the 3-week posttest for all instructional conditions, *t*s (37) from 7.96 to 11.97, all *p*s < .001 (see Table 1 for means). Finally, for those who participated in the follow-up after 3 months, accuracy remained above pretest levels for all instructional conditions, *t*s (23) = 5.30–12.56, all *p*s < .001. An additional analysis showed that learning was not a result of test item repetition.[3]

There was a significant interaction between test time and rule instruction, F (2, 44) = 10.59, $p < .001$, which was also found when including the 3-month follow-up, F (3, 51) = 4.34, $p < .005$. At baseline, cues assigned to the rule instruction condition showed lower d' than did cues assigned to the feedback only condition, F (1, 32) = 4.86, $p < .05$. However, at each posttest, there were higher

d' scores for rule instruction than for feedback only. The effect of rule instruction was marginally significant at immediate posttest, $F\ (1, 30) = 3.62$, $p < .07$, and significant at 3-week posttest, $F\ (1, 29) = 7.03$, $p < .02$, but no longer significant at 3-month follow-up, $F\ (1, 17) < 1$, $p > .2$, with a smaller sample.

The d' values for each level of the diversity condition are shown in Table 1. At baseline, there was a significant advantage for cues assigned to the frequent exemplars compared to the diverse exemplar condition, $F\ (1, 32) = 9.52$, $p = .004$, but there were no other effects of diversity of exemplars and no interactions between diversity of exemplars and test time or rule instruction.

LATENCY. There were no pretest differences in latency among the instructional conditions. We performed a 3 (Pretest, Immediate Posttest, 3-Week Posttest) × 2 (Feedback Only, Feedback Plus Rule) × 2 (Frequent Exemplar, Diverse Exemplars) analysis. The positive linear effect of test time showed evidence of learning, $F\ (1, 38) = 115.26$, $p < .0001$, and the negative quadratic effect of test time provided evidence of forgetting, $F\ (1, 38) = 103.08$, $p < .0001$. This main effect of learning reflects a speedup for each instructional condition, reflected in a significant speedup from pretest to immediate posttest, *t*s (38) = 10.05–10.85, all *p*s < .001, from pretest to 3-week posttest, *t*s (38) = 9.94–10.79, all *p*s < .001, and from pretest to 3-month posttest, *t*s (23) = 4.69–5.72, all *p*s < .001.

The frequent exemplar condition was faster than the diverse exemplar condition, $F\ (1, 38) = 5.11$, $p < .05$. There was no effect on latency of rule instruction, no interactions of test time with either rule instruction or diversity of exemplars, and no three-way interaction (*F*s < 2, *p*s > .1). For each time point, we performed a 2 (Feedback Only, Rule Instruction) × 2 (Frequent Exemplar, Diverse Exemplars) within-subjects ANOVA. There were no instructional effects on latency at any time point, with one exception. At immediate posttest, the rule instruction condition was faster than the feedback only condition, $F\ (1, 38) = 5.61$, $p < .05$ (see Table 1 for means).

Discussion

The first finding of Experiment 1 was that all of the instructional conditions produced learning. Having spent no more than 105 min learning these 24 cues, students who had no French training went from a baseline accuracy of 63% to a classification accuracy for untrained words at above 90% on the immediate posttest and, after 3 months without further training, still performed at over 80% accuracy on words for which they never received feedback. These findings support our earlier reports of effectiveness of instruction for French gender cues (Pavlik, Presson, Dozzi, et al., 2007). They are also consistent with previous studies of young learners over longer training periods (Harley, 1998; Lyster, 2004a). However, contrary to our findings, Harley (1998) found that young learners could not generalize the relevant cues to production of uninstructed nouns. Moreover, Manning (1996) reported cue generalization after computerized instruction but for intermediate learners, rather than for the naive learners trained here.

Instruction with rule presentation was more effective than was learning without rule presentation, consistent with the growing literature showing the benefit of explicit instruction (Norris & Ortega, 2000; Spada & Tomita, 2010). The advantage of rule instruction was seen in categorization accuracy, measured with d', showing that rule instruction produced more accurate discrimination of grammatical gender. Moreover, this advantage of rule instruction was maintained after a 3-week interval with no training or testing. Although effects at immediate posttest might be somewhat constrained by the very high level of overall performance, the effects at 3-week posttest show that rule instruction increases not only learning but also retention, which is certainly important for L2 learning. Effects weakened at 3 months, although the smaller sample for the third posttest makes this finding difficult to interpret.

The finding that rule instruction with correctness feedback outperforms correctness feedback alone supports the prediction of input-centered models of L2 acquisition such as the unified competition model. Within that model, explicit instruction through rule presentation produces cue focusing, encouraging learners to notice relevant cues and encode them efficiently, thus enabling better cue representation and allowing L2 learners to consolidate new knowledge (MacWhinney, 1997). Explicit cue presentation presents a clear mapping between spelling patterns and grammatical gender, and therefore produces more effective learning than correctness feedback alone. Moreover, the advantages of rule instruction were not significantly affected by variation in the frequency versus diversity of exemplars. It should be noted, however, that the rule instruction in this experiment involved both directing learner attention to the correct ending and explicitly describing the form–meaning mapping to gender. Thus, it is possible that the positive effects of rule instruction in Experiment 1 can be explained on the basis of the established benefits of feature focusing (Schmidt, 1994). Experiment 2 tests this possibility.

In contrast to the confirmation of the predictions regarding explicit rule presentation, our predictions regarding the beneficial effects of training with a highly frequent exemplar were not supported. Although this could mean that frequent exemplars are not helpful in structuring grammatical categories in the case of L2 learning, it may not reflect a failure of the frequent exemplar, but rather emphasize the importance of exposing learners to a wide variety of input form variations.

Furthermore, although there was the expected speedup after training, there was mixed evidence for instructional effects on latency. For rule instruction, there was an advantage at immediate posttest, but not on the delayed posttest. Although the frequent exemplar condition was faster on all posttests, there was no interaction with testing session and therefore no evidence of differential proceduralization for instruction with a frequent exemplar compared to diverse exemplars. The mixed evidence for instructional effects on latency may be because there was no time pressure during training or testing. Therefore, participants did not need to proceduralize the application of cues. The use of time pressure to force more automatic performance might provide evidence of proceduralization, and this issue is examined in Experiment 2.

EXPERIMENT 2

Experiment 1 showed that explicit rule instruction in combination with correctness feedback has an advantage over correctness feedback alone in a sample of adult novice learners. The unified competition model holds that the effect of rule presentation works by strengthening form–function mapping. However, the experimental manipulation of explicit rule presentation also inevitably involves directing learners' attention to certain properties or features of the input. Thus, it is possible that rule presentation has its effects merely through feature focusing. To test whether the mechanism of feature focusing accounted for all of the effects of rule presentation, Experiment 2 separated the effects of rule presentation and feature focusing by comparing the rule presentation treatment of Experiment 1 and a cue highlighting treatment. The rule presentation treatment should involve both feature focusing and rule learning effects, but the cue highlighting treatment should involve only feature focusing effects.

Feature focusing, broadly defined, is a common instructional intervention, ranging from input enhancement (Rutherford & Sharwood-Smith, 1985) to structured input and focus on form (Long, 2000). Feature focusing is thought to improve learning by directing attention to the correct part of the linguistic input, because noticing is a key factor in using input effectively (Mackey, 2006; Robinson, 1995; Schmidt, 1994). Experiment 2 compared the effects of rule presentation (i.e., creating conscious awareness of grammatical categories) and cue highlighting. In the cue highlighting condition, we direct attention to the correct features by highlighting the same orthographic endings that were taught in the rule instruction condition. The two instructional conditions were compared both immediately after training and in a 1-week follow-up. Consistent with Experiment 1, the predictions of the competition model, and the findings of the Norris and Ortega (2000) and Spada and Tomita (2010), we predict that rule instruction will produce better performance than cue highlighting.

In Experiment 1 there was no time pressure during training or testing, so participants did not need to proceduralize the application of cues. Time pressure interferes with the application of procedures that are incompletely learned or weakly represented. Testing how learners cope with time pressure is another way of testing the strength of their learned representations (Schneider & Chein, 2003). One potential drawback to explicit instruction is that it may lead learners to form representations that could be difficult to access during online (time pressured) usage. For example, the input hypothesis (Krashen, 1994) predicts that, even if instruction improves metalinguistic (conscious) competence, this improvement will not transfer to naturalistic, online performance. Similarly, researchers (e.g., Paradis, 2004) have argued that adult learners use explicit metalinguistic information to form declarative grammar representations, whereas fluent production in native speakers depends instead on procedural representations. If this is true, then the positive effects of rule instruction should be reduced or eliminated when learners are required to produce forms under time pressure, and rule instruction may prove ineffective as a support for naturalistic communication, which requires real-time (time pressured) comprehension and production. In contrast, a feature focusing intervention, without an explicitly presented rule, could avoid such problems by bypassing the declarative formulation of the rule and therefore could also be more robust to the interference of time pressure on more deliberative processing. Thus, Experiment 2 tested whether any advantage of explicit rule-based instruction would be lost when learners were tested under conditions of time pressure.

When measuring performance under time pressure, it is important to test participants who are accustomed to the task of responding with a deadline. The need to adapt to this task suggests that previous training with time pressure could be important preparation for performance under time pressure. Moreover, if a given instructional condition were easier to learn under time pressure demands, this would be important information for educators, even if slower practice did not show the same pattern. If representations created by explicit instruction are less effective when applied under testing time pressure (e.g., Krashen, 1994), then learning with time pressure should more negatively affect performance in the rule instruction condition than the cue highlighting condition. Accordingly, Experiment 2 also varied time pressure during training.

Finally, Experiment 1 showed better than expected pretest accuracy, which could constrain the demonstration of instructional effects. Moreover, the within-subjects design is subject to possible transfer effects across conditions (although order of condition was randomized in Experiment 1). Thus, in Experiment 2, the instructional conditions were tested in a between-subjects design, and we removed the English translations that might have produced higher than expected levels of pretest accuracy.

In summary, Experiment 2 compared rule instruction with cue highlighting and tested whether the predicted advantage of rule instruction would be modified by time pressure during training and during testing. Finally, we tested the possibility that training under time pressure could acclimate participants to the time pressure demands and encourage proceduralization, thereby eliminating the negative impacts of testing with time pressure. Consistent with the theory of noticing, we predict that highlighting cues to direct learner attention will lead to significant learning, as did rule instruction in Experiment 1. However, consistent with the unified competition model prediction that simple explicit mappings can create greater transparency and facilitate category formation, we predict that rule instruction will show greater improvement than cue highlighting.

Method

Participants. Participants were 92 (62 female, 33 male) students and community members recruited at Carnegie Mellon University using online advertisements on a university message board. Participants were excluded if they had any prior experience with French, any training in a Romance language, or if their native language had a declensional system based on gender. However, participants whose native language had no gender contrast, such as Chinese or Japanese, were included. Participants were randomly assigned to one of four instructional conditions: rule instruction + time pressure ($n = 20$); rule instruction – time pressure ($n = 21$); highlighting cues + time pressure ($n = 23$); highlighting cues – time pressure ($n = 28$). One participant was missing data in the untimed immediate posttest, 1 participant was missing data in the timed delayed posttest, and 13 participants did not complete the delayed posttest.

Design and procedure. The design was a 3 (Pretest, Immediate Posttest, Delayed Posttest) × 2 (Rules, Highlighting) × 2 (Training Time Pressure, No Training

Time Pressure) × 2 (Testing Time Pressure, No Testing Time Pressure) mixed design. The instructional explicitness condition and training time pressure were between-subjects factors, and testing time pressure was a within-subjects factor. The experiment consisted of a pretest, a training session, an immediate posttest, and a delayed posttest administered 1 week later.

TRAINING. The training session lasted 42 min for all conditions, and the session was ended based on the time elapsed rather than the number of trials. Thus, participants in the time pressure training conditions received more trials on average than did participants in the no time pressure training conditions, but the time on task was equal for all conditions. The same 24 cues and 14 word exemplars of each cue as in Experiment 1were presented in testing and training stimuli.

In training, participants were presented with a target noun shown both with the masculine article *le* and with the feminine article *la*, one above the other. Participants were asked, "Which version is correct?" The required response was to press either the M key for *le* or the F key for *la*. Feedback followed the answer in the form of green highlighting on the correct answer choice and sound feedback dependent on the accuracy of the response (a positive ding tone for correct and a negative buzz for incorrect).

As in Experiment 1, the explicit rule instructions were presented as statements in which an orthographic ending indicated a gender choice, but the gender choice presented was the form of the article, rather than the letters M or F (e.g., Experiment 1 taught *–age* → M, and Experiment 3 taught *–age* → *le*). The cue appeared after incorrect responses only, and feedback came on every trial.

The highlighting condition provided orthographic enhancement by capitalizing those letters at the end of the word that corresponded to the endings formulated in the explicit-rule condition. For example, the word "fromage" was presented as fromAGE. This capitalization was a basic property of the stimulus display that was visible both before and after learner response during training. In the highlighting condition, the only additional information provided after learner response was correctness feedback.

To introduce time pressure during training, we set a time limit for response at 1400 ms. This limit was chosen because it was 2 *SD* above the mean latency of 1050 ms for correct responses in Experiment 1. If the participant did not respond in the time allowed, feedback was presented as for incorrect answers. In the no time pressure condition, the time limit was 6000 ms, which was identical to Experiment 1.

TESTING. To minimize participant frustration, all participants, regardless of training condition, took the pretest with no time pressure. All participants took a total of four generalization posttests (two immediate posttests and two posttests after a 1-week delay, with and without time pressure). The order of the timed and untimed tests was counterbalanced across participants and constant for each participant at each test time. Test items were identical to Experiment 1. In the posttests, the time pressure variable was operationalized as in training: a 1400 ms deadline with time pressure and 6000 ms without time pressure. In each posttest, four exemplars of each of the 24 cues were presented. Posttest words were identical for the immediate and 1-week delayed posttests, but no words used in training were included in the posttests and no feedback was given during tests. Thus, posttest performance represents generalization to untrained items.

Data analysis. We first performed a 3 (Pretest, Immediate Posttest, Delayed Posttest) × 2 (Rules, Highlighting) × 2 (Training Time Pressure, No Training Time Pressure) mixed ANOVA for the timed and untimed posttests separately to compare rule instruction to cue highlighting and to test whether time pressure in training improved performance on timed (and also untimed) posttests. Then, to see whether time pressure during testing moderated the instructional condition effects (and therefore restricting the analysis to the posttests, which varied time pressure), we performed a 2 (Immediate Posttest, Delayed Posttest) × 2 (Testing Time Pressure, No Testing Time Pressure) × 2 (Rules, Highlighting) × 2 (Training Time Pressure, No Training Time Pressure) mixed ANOVA. Dependent measures were d' and latency for correct responses. Because it was restricted to the posttests, this analysis also tested forgetting.

We present raw accuracy data for descriptive purposes (see Table 2). However, as in Experiment 1, we analyzed the accuracy data in terms of d' scores because Experiment 1 showed preexisting response biases. We excluded trials on which the participant failed to make a response (within 1400 ms for timed tests or within 6000 ms for untimed tests), primarily because d' cannot be calculated on trials where no response is given.[4]

Results

Overall performance. Pretest accuracy for Experiment 2 was at chance (overall mean proportion correct = 0.51, $SD = 0.11$). The means for raw accuracy, d', and latency for all four training groups across testing intervals and levels of testing time pressure are presented in Table 2. The d' data indicate that participants had no substantive prior knowledge overall, with no group mean greater than 0.263 ($SD =$ 0.116) at pretest. Moreover, unlike in Experiment 1, there was no pretest difference in d' between groups for explicit rule instruction compared to highlighting, F (1, 88) = 0.53, $p = .47$, or training time pressure compared to no training time pressure, F (1, 88) = 0.16, $p = .69$, and no interaction, F (1, 88) = 0.24, $p = .62$. Pretest latencies showed no significant difference between randomized groups for explicitness, F (1, 88) = 2.64, $p = .11$, or training time pressure, F (1, 88) = 0.75, $p = .39$, and there was no interaction, F (1, 88) = 0.071, $p = .79$.

Training led to substantial learning for all groups at both immediate and delayed posttests. There was a large, significant change in d' over the testing sessions, as seen in the main effects of session in the 3 (Pretest, Posttest, Delayed Posttest) × 2 (Rules, Highlighting) × 2 (Training Time Pressure, No Training Time Pressure) ANOVAs for both untimed, F (2, 148) = 150.1, and timed, F (2, 148) = 96.72, testing conditions, both ps < .001. Post hoc comparisons confirmed that learners were more accurate at both immediate and delayed posttests compared to pretest in both timed and untimed testing conditions (ps < .001, see Table 2 for means).

Learning was also evident in reduced latencies for correct responses between pretest and posttests. In the overall ANOVA, there was a significant effect of session

Table 2. *Results from Experiment 2: Rule instruction and highlighting for training with and without time pressure*

	Pretest		Immediate Posttest				1-Week Delay Posttest			
	Untimed Test		Timed Test		Untimed Test		Timed Test		Untimed Test	
					d′					
Training Condition	*d′*	*SE*	*d′*	*SE*	*d′*	*SE*	*d′*	*SE*	*d′*	*SE*
Rule instruction	0.074	0.145	1.959	0.178	2.700	0.194	1.644	0.146	2.029	0.164
No time pressure	0.081	0.268	1.990	0.275	2.994	0.301	1.727	0.210	2.290	0.229
Time pressure	0.067	0.103	1.926	0.228	2.392	0.278	1.570	0.209	1.794	0.221
Highlighting	0.184	0.082	1.431	0.114	1.963	0.150	1.217	0.112	1.458	0.126
No time pressure	0.119	0.116	1.373	0.149	1.843	0.195	1.091	0.134	1.522	0.173
Time pressure	0.263	0.116	1.502	0.179	2.103	0.233	1.370	0.187	1.383	0.186
					Accuracy					
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Rule instruction	0.537	0.015	0.811	0.021	0.863	0.022	0.766	0.020	0.789	0.021
No time pressure	0.552	0.020	0.810	0.026	0.880	0.032	0.763	0.031	0.800	0.030
Time pressure	0.522	0.021	0.812	0.029	0.847	0.030	0.769	0.029	0.777	0.029
Highlighting	0.544	0.014	0.749	0.022	0.798	0.022	0.712	0.020	0.745	0.021
No time pressure	0.535	0.020	0.741	0.026	0.793	0.030	0.696	0.026	0.748	0.028
Time pressure	0.554	0.021	0.757	0.030	0.803	0.032	0.727	0.030	0.742	0.030
					Latency (ms)					
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Rule instruction	2083	106	890	25	1175	50	871	24	1159	57
No time pressure	2103	154	909	37	1251	73	863	34	1215	83
Time pressure	2062	146	871	35	1099	69	880	33	1102	79
Highlighting	1751	105	845	24	1207	50	854	22	1167	56
No time pressure	1728	143	868	32	1250	68	860	30	1193	77
Time pressure	1774	154	821	36	1163	73	848	34	1141	83

for the untimed posttest, F (2, 154) = 96.84, $p < .001$. Post hoc comparisons showed that responses on both immediate and delayed untimed posttests were significantly faster than were responses on the pretest (ps < .001, see Table 2 for means). There was also a significant speedup from pretest to the timed posttests, F (2, 154) = 226.68, ps < .001. However, given that there was no time pressure during the pretest, this particular comparison is less informative.

Forgetting. Although improvement from pretest was shown at both immediate and delayed posttest, there was also forgetting from the immediate to the delayed posttest. Across all conditions and both timed and untimed tests, average d' decreased between the immediate posttest and the 1-week delayed posttest, as seen in a main effect of test session in analyses comparing the immediate and delayed posttests, for both untimed tests, F (1, 74) = 52.62, $p < .001$, and timed tests, F (1, 74) = 18.15, $p < .001$. This pattern was also reflected in a significant quadratic effect of test session in the ANOVAs that included pretest for both untimed and timed testing conditions, Fs (1, 74) = 170.37 and 90.70, respectively, both ps < .001 (see Table 2 for means).

However, latency analyses did not show a similar change from the immediate to the delayed posttest. Although there was a significant main effect of test session in the latency analysis (described above), an analysis with the immediate and delayed posttest only showed that latency across all conditions did not increase from the immediate to the 1-week delayed posttest, F (1, 76) = 0.28, $p = .601$, and this lack of slowdown was equivalent for the timed and untimed posttests, F (1, 76) = 0.38, $p = .54$.

Effects of instructional conditions on learning (d'). Overall, rule instruction led to better performance than did highlighting. For the analyses that included pretest data, there was a main effect of training condition on d' scores for the untimed tests, F (1, 74) = 5.91, $p = .018$, and marginally for the timed tests, F (1, 74) = 3.92, $p = .052$. It is more important that, in both analyses, the main effect was qualified by significant interactions with test session, F (2, 148) = 7.86, $p = .001$, for untimed tests and F (2, 148) = 5.61, $p = .004$ for timed tests. The nature of the interaction was consistent, in that the rule instruction and highlighting groups did not differ at pretest, but the rule instruction groups showed higher d' at both the immediate and the delayed posttests.

Consistent with our predictions, at the immediate posttest, participants assigned to the rule instruction condition showed higher d' than did participants assigned to the highlighting conditions for both untimed ($d' = 2.70$ and 1.96 for rule and highlighting conditions, respectively), F (1, 87) = 8.24, $p = .005$, and timed test conditions ($d' = 1.96$ and 1.43 for rule and highlighting conditions), F (1, 88) = 6.37, $p = .013$.

The same pattern was found at the 1-week delayed posttest, as participants in the rule instruction condition showed better performance than did those in the highlighting conditions for both untimed ($d' = 2.03$ and 1.46 for rule and highlighting), F (1, 75) = 7.28, $p = .009$, and timed test conditions ($d' = 1.64$ and 1.22 for rule and highlighting), F (1, 74) = 4.46, $p = .038$.

Because rule application initially requires additional conscious processing, we predicted that the advantage of rule instruction relative to highlighting would be larger for untimed posttests than for timed posttests. This was reflected in the interaction between instructional condition and testing time pressure, F (1, 72) = 5.76, p = .002, though both test types showed a significant advantage for explicit instruction. The average rule advantage for untimed tests was 0.74 (mean d' = 2.70 and 1.96 for explicit and highlighting), whereas for timed tests that advantage was 0.53 (mean d's = 1.96 and 1.43).

To examine whether time pressure in training reduced the difficulty of time pressure during testing, we compared the effects of training with and without time pressure on performance in the timed and untimed tests. Surprisingly, training under time pressure did not affect learning measured in d' as a main effect or interaction in the analyses of all three sessions (pretest, immediate posttest, delayed posttest) for either timed or untimed posttests (all Fs < 1.4, *ns*). Moreover, there was no interaction between rule instruction and training time pressure on d' for either timed or untimed testing conditions (all Fs < 1.5, *ns*).

Effects of instructional conditions on latency. Rule instruction also showed a larger reduction in latency after training compared to the highlighting group, reflected in a significant interaction with test session for both the untimed, F (2, 154) = 5.48, p = .005, and timed, F (2, 154) = 6.27, p = .002, testing conditions (see Table 2 for means). It was surprising that training under time pressure did not affect testing response latencies, whether testing was with or without time pressure (all Fs < 1, *ns*).

Effects of instructional conditions on forgetting. We performed a separate analysis for forgetting, comparing performance from the immediate to the delayed posttests. There was a marginal interaction between test session and rule instruction, F (1, 72) = 2.99, p = .088, such that the gap in mean d' between rule instruction and highlighting overall was reduced from 0.63 at immediate posttest to 0.50 after a 1-week delay. Training with time pressure did not lead to significantly less forgetting from immediate to delayed posttest than did training without time pressure, F (1, 72) = 0.992, p = .323, and this was true for both timed and untimed tests, F (1, 72) = 0.441, p = .51.

For latency, there was no overall slowdown from immediate to delayed posttest, F (1, 76) = 0.28, p = .60, and no differences in slowdown from immediate to delayed posttest for training with and without time pressure, F (1, 76) = 0.58, p = .447, or with rule instruction versus stimulus highlighting, F (1, 76) = 0.093, p = .76. The lack of instructional effects on latency from immediate to delayed posttests was consistent for timed and untimed posttests.

Discussion

Experiment 2 compared the effects of rule instruction with a less explicit cue highlighting condition. In accord with the predictions of the unified competition model, we found that rule instruction produced better performance than mere feature focusing. We also tested the proposal that explicit instruction does not allow for procedural representation (e.g., Paradis, 2004), by testing whether the advantage of rule instruction would be absent or reversed under testing conditions that included time pressure. This prediction was not supported. Finally, we speculated that training with time pressure could eliminate the negative impact of testing with time pressure, which was not confirmed in the data.

The pretest data showed that, unlike in Experiment 1, participants initially performed at chance, with no pretest differences between instructional groups (indicating successful randomization). Starting from chance levels of performance, learners across all instructional conditions improved substantially after training and maintained accurate discrimination of gender categories a week after training. Moreover, in addition to improvements in accuracy, participants were faster after training both with and without time pressure, and this improvement was maintained over the 1-week interval. This suggests that the intervention was successful in producing associations between orthographic endings and grammatical gender through articles, and also that those associations were robust to a retention interval, which is one important way of testing the stability of mental representations.

The central finding was that rule instruction produced more accurate performance than did cue highlighting (feature focusing alone) and that this advantage was maintained over a 1-week delay. The advantage for explicit rule instruction replicates Experiment 1 as well as other studies showing positive effects of grammatical instruction (e.g., Norris & Ortega, 2000; Spada & Tomita, 2010). Moreover, these findings extend previous literature and Experiment 1 by demonstrating that the advantage of rule instruction was actually due to instruction in the rule and not the result of simply directing the learners' attention to the relevant forms, which in itself has been shown to be a powerful tool in improving language learning (Lightbown & Spada, 1990; Lyster, 2004a). Thus, these findings provide a more specific characterization of the benefits of rule instruction, suggesting that the explicitness could increase the transparency of cue mappings and strengthen the resulting representations. In addition, they provide a more stringent test of rule instruction by demonstrating its advantage over another active instructional technique rather than over correctness feedback only.

In the current study, cue highlighting showed significant and lasting learning after training, albeit not as much as explicit rule instruction, but it is possible that a different focusing manipulation could lead to even more improvement. Future studies can use different focusing methods, such as color highlighting or underlining, to test the generalizability of our finding to other stimulus modifications.[5] The most likely interpretation of the lesser, but positive, effects of cue highlighting is that noticing is the first step in the process of cue extraction. Participants use perceptual focusing to form mental representations of cues, first extracting the pattern and then associating it with (in this case) gendered articles. Participants who are given rules directly do not have to engage in this extra preliminary work because the association is explicitly presented; however, participants in both conditions end up formulating usable versions of the relevant rules.

In terms of forgetting, we did not find strong evidence that the advantage of explicit rule instruction over cue highlighting was lost over time. There was marginally more forgetting with rule instruction than with cue highlighting, but rule instruction still retained a significant advantage over cue highlighting at the

1-week posttest. Thus, based on the current findings, we do not predict that rule instruction would lose its advantage.

The strongest demonstration of the efficacy of explicit rule instruction is that the rule advantage was also present in time pressured posttests, when learners had to produce rapid responses, as required in more natural communication. Thus, although there was a narrowing of the advantage for explicitness when tested with time pressure, we found no evidence that rule instruction produced representations that were not compatible with online performance. In a model in which grammatical knowledge associated with explicit information is inherently incompatible with naturalistic processing (Krashen, 1994), presenting declarative or explicit category information should impair time pressured performance. The declarative/procedural model (Ullman, 2001a, 2001b) explains the lower levels of performance in L2 grammatical processing compared to L1 in terms of the costs of using declarative representations for a procedural process. The current study did not show such costs of explicit rule instruction. Perhaps a greater demand for automaticity (either from time pressure or dual-task contexts) would be needed to require procedural performance and thus eliminate the advantage of rule instruction.

As expected, adding time pressure during testing had an overall negative effect on accuracy, showing that time pressure does make the grammar categorization task more demanding. Further, there was no evidence that training under time pressure mitigated this decrement. However, cognitive models of skill acquisition predict automaticity only after accuracy is high (Anderson & Fincham, 1994; Segalowitz & Segalowitz, 1993). Thus, it may be more appropriate to reserve conclusions about response time floors and achieving procedural representations for more advanced learners who have mastered the basic categorization task. In future work, time pressure could be used as a testing variable to add demand, but training with time pressure during instruction might be more effective after initial training is already successfully achieved.

Responses were faster after training and remained faster than pretest after a 1-week delay. Rule instruction led to faster responding than did cue highlighting. This could reflect either the increase in accuracy or a difference in the amount of proceduralization between the two conditions. When compared to cue highlighting, rule instruction led to both greater accuracy and equal latency, which suggests that rule instruction did not result in a speed–accuracy trade-off.

GENERAL DISCUSSION

The current experiments extend the scope of the literature showing the benefits of explicit instruction. These experiments extend the results to participants who have no prior knowledge of the language. This shows that explicit instruction can be an effective aid to learning even at the very early stages of L2 learning. It is also possible that beginners will show even greater effects of explicit instruction because they need more of a scaffold to understand grammatical patterns in an L2. In addition, Experiment 2 tested one potential mechanism of explicit instruction, the powerful benefit of directing learner attention to relevant cues, showing that although directing attention led to sustained improvement, explicit rule presentation led to even greater improvement. Finally, the addition of time pressure during testing showed that explicit instruction does not cost learners the ability to perform the task quickly, as would be expected if declarative rule presentation discouraged skill learning.

Taken together, the current experiments provided clear support for the effects of rule instruction on both immediate learning and retention, showed that the benefits of rule instruction were not due to either repeated testing alone or to associated feature focusing, and showed that the benefits were retained (albeit reduced) when learners needed to produce responses under time pressure. Future experiments can expand the current results by testing whether the benefits of explicit instruction shown here generalize to more complex rule paradigms, such as verb conjugation, which have multiple levels of cue mapping (e.g., regularity, verb class, tense, person). In generalizing to more complex grammatical patterns, we can further explore how explicit instruction and noticing interventions improve performance: by affecting encoding, rehearsal, or generalization performance. These different mechanistic explanations will be more dissociable in a more complex linguistic task.

CONCLUSIONS

In conclusion, the present studies successfully trained novices to assign French nouns to their correct grammatical gender and retained high levels of performance up to a 3-month delay. Adding a frequent exemplar led to learning, but it did not lead to better performance than presenting a more diverse exemplar set. However, providing explicit rules and feedback led to better generalization and greater retention over time than simply providing feedback regarding correctness or highlighting the relevant word ending. Thus, the effect of rule instruction is not due to simply focusing the learner's attention. Our findings show that, in at least some grammar tasks, direct information about form–function mappings leads to better learning outcomes. Contrary to our expectations, this advantage remained (though smaller) even when testing occurred under time pressure. Thus, explicit instruction may be helpful even when the target performance task requires speeded performance. This finding supports the prediction of the declarative/procedural model (Ullman, 2001a, 2001b) that practice can lead to the proceduralization of declarative grammatical knowledge. However, it contrasts with predictions that explicitly instructed knowledge is not usable in online processing (e.g., Krashen, 1994) and that explicit and implicit knowledge are fully dissociable systems, such that explicit metalinguistic knowledge cannot become proceduralized over time (e.g., Paradis, 2004).

Our findings have extended the literature on the importance of explicit instruction in grammatical structures (Long, 2000; Lyster, 2004a; Norris & Ortega, 2000) in five important ways. First, we extend previous evidence of the utility of explicit instruction to a sample of novice learners. Second, we show that instruction that includes explicit statement of cues of rules is superior to cue highlighting (feature focusing without explicit rule information). Third, we show that providing highly frequent exemplars fails to yield the effects predicted by some models. Fourth, we show that, even for this complex cue set, a total of between 60 and 105 min of focused instruction with feedback can lead to significant retention across up to 3 months without further instruction. Fifth, we have presented a method based on computerized experiments conducted with novice learners that can be further extended and replicated to examine fine-grained details of the learning process. For example, using this framework, we could test parametrically for the effects of variations in speeded responding, changes in exemplar frequency, inclusion of irregular forms, and combinations with other tasks. Although findings derived from such tests cannot be applied to the classroom context without further testing, they provide a useful paradigm for comparison against more naturalistic but less controlled contexts of classroom learning.

APPENDIX A

Complete list of cues taught in the experiment with their pretest accuracy

Cue	Experiment 1 Pretest Accuracy (*SD*)	Training Words	Test Words
1. *–age* (M)	.52 (.50)	Garage, étage, fromage, paysage, recyclage, nuage, visage, potage, avantage, outrage, maquillage, ménage, repassage, stage	Apprentissage, tapage, clivage, otage
2. *–ance* (F)	.52 (.50)	Avance, confiance, outrance, aisance, béance, résistance, enfance, naissance, espérance, plaisance, puissance, résonance, souffrance, connaissance	Protubérance, abondance, croyance, tendance
3. *–d* (M)	.77 (.42)	Canard, billard, plumard, bord, traînard, nid, lard, retard, papelard, renard, froid, chaud, bled, accord	Nid, dard, bagnard, gond
4. *–é* (M)	.49 (.50)	Marché, congé, procédé, préjugé, péché, bébé, canapé, café, cliché, résumé, scellé, blé, défilé, ciné	Triplé, curé, envoyé, thé
5. *–ée* (F)	.62 (.49)	Poupée, fée, nausée, allée, chevauchée, coulée, vallée, invitée, fumée, année, idée, journée	Mosquée, armée, musée, buée
6. *–ère* (F)	.60 (.49)	Panthère, prière, colère, banquière, bière, verrière, lisière, infirmière, étagère, matière, bannière, ferrière, misère, rivière	Croisière, carrière, frontière, paupière
7. *–en/ien* (M)	.83 (.38)	Musicien, pharmacien, informaticien, chien, technicien, généticien, mathématicien, politicien, doyen, magicien, statisticien, moyen, terrien, gardien	Citoyen, rien, chirurgien, soutien
8. *–eur* (M)	.73 (.45)	Auteur, coiffeur, acteur, coeur, tricheur, ordinateur, traducteur, fugueur, dîneur, présentateur, sauteur, ingénieur, tailleur, joueur	Distributeur, transbordeur, entrepreneur, choeur
9. *–i/oi* (M)	.52 (.50)	Ami, cari, convoi, parti, abri, défi, céleri, roi, emploi, quai, pari, ennemi, essai, vrai	Canari, apprenti, alcali, frai
10. *–ie* (F)	.71 (.46)	Hystérie, analogie, confiserie, académie, monnaie, pharmacie, allergie, bougie, garderie, hypocrisie, anarchie, sortie, envie, infamie	Mercerie, jacasserie, agonie, apologie
11. *–ier* (M)	.75 (.44)	Officier, grenier, quartier, caissier, banquier, tablier, collier, gravier, papier, barbier, cavalier, fournier, chemisier, escalier	Terrier, fumier, lunetier, oreiller
12. *–in/ain* (M)	.82 (.39)	Pain, raisin, terrain, foin, gradin, fin, destin, étain, ravin, brin, chagrin, jardin, requin, bain	Parrain, tremplin, parchemin, puritain

13. *–ine/aine* (F)	.63 (.49)	Cuisine, migraine, usine, médecine, rétine, épine, gazoline, vitrine, porcelaine, bassine, piscine, farine, poitrine	Echine, crépine, voisine, insuline
14. *–isme* (M)	.36 (.48)	Capitalisme, judaïsme, cyclisme, criticisme, égotisme, prisme, charisme, futurisme, activisme, cubisme, ironisme, négativisme, autisme, baptisme	Cosmopolitisme, mutisme, journalisme, élitisme
15. *–me* (M)	.46 (.50)	Problème, rhume, calme, dilemme, drame, poème, sarcasme, fantôme, rythme, charme, thème, terme	Barème, diplôme, asthme, dôme
16. *–nne* (F)	.79 (.41)	Couronne, canne, ancienne, obsidienne, consonne, antenne, colonne, électricienne, musicienne, paysanne, magicienne, piétonne, lionne, nonne	Tonne, practicienne, collégienne, espionne
17. *–on* (M)	.83 (.38)	Pardon, faucon, sermon, poumon, avion, cordon, citron, dicton, manchon, savon, démon, pantalon, salon, menton	Capuchon, goudron, torchon, charbon
18. *–se* (F)	.46 (.50)	Défense, course, réponse, bosse, graisse, crise, danse, bourse, thèse, parenthèse, cuisse, ruse, anse, transe	Entorse, chausse, pause, crevasse
19. *–esse* (F)	.88 (.32)	Princesse, vitesse, politesse, vieillesse, hôtesse, déesse, presse, messe, promesse, maîtresse, adresse, justesse, finesse, sagesse	Prêtresse, grossesse, faiblesse, comtesse
20. *–t* (M)	.83 (.38)	Pont, bruit, trajet, droit, berlingot, commissariat, croyant, carnet, billet, point, mandat, syndicat, passeport	Achat, doigt, mot, saut
21. *–té* (F)	.50 (.50)	Beauté, santé, parité, habilité, densité, liberté, identité, humanité, proximité, impureté, agilité, publicité, égalité, cité	Avidité, société, rivalité, comptabilité
22. *–tion/sion* (F)	.22 (.41)	Condition, pression, pension, alimentation, discrimination, préoccupation, immigration, salutation, désorientation, signification, organisation, opération, détection, audition	Affirmation, attention, dénégation, inondation
23. *–tre* (M)	.53 (.50)	Ministre, centre, prêtre, pitre, âtre, antre, arbitre, monstre, ventre, orchestre, filtre, astre, pupitre, désastre	Cintre, peintre, théâtre, meurtre
24. *–u* (M)	.65 (.48)	Inconnu, tissu, trou, élu, tonneau, alu, faisceau, panneau, préau, cru, radeau, farfelu, parvenu, anneau	Essieu, flou, taureau, noyau

ACKNOWLEDGMENTS

This work was supported by a Graduate Training Grant awarded to Carnegie Mellon University by the Department of Education (R305B040063) and by the Pittsburgh Science of Learning Center, with funding from the National Science Foundation (SBE-0354420). Portions of the data were presented at the Second Language Research Forum (2010) and the International Symposium on Bilingualism (2009).

NOTES

1. For all analyses of the frequent-exemplar manipulation, data for trials that presented the highly frequent word itself were excluded to prevent confounding from familiarity with that token.
2. The d' scores were calculated with an adjustment for hit rates and false alarm rates of 0 or 1. If n is the number of possible hits or false alarms, the smallest nonzero hit or false alarm rate is $1/n$. Observed hit and false alarm rates of 0 are corrected midway between 0 and that lowest observable rate: $1/2n$. Likewise, hit and false alarm rates of 1 are adjusted to $1 – 1/2n$.
3. A comparison group of 20 participants who received no training but completed the same testing schedule as the trained participants showed no significant improvement over time, F (2, 32) = 1.82, p = .18, and a small effect size for pretest to delayed posttest improvement (partial $\eta^2 = 0.10$) compared to the learning gain effect size for the instructional conditions, which ranged from 0.631 to 0.763.
4. If the same analyses are run on raw accuracy or on d' with those trials marked as errors (assuming that learners would have eventually chosen the wrong answer), the overall pattern of data remains the same, but with increased variability for all tests.
5. It is possible that capitalization added cognitive load to the task. However, we do know that letters are not fully processed as seen; instead an abstract letter identity independent of case is extracted from visual letter forms, with experimental evidence that uppercase letters′ their lowercase counterparts (Besner, Coltheart, & Davelaar, 1984). Reading is possible even with within-word case mixing (e.g., “tHiS TeXt CaN bE rEaD”), and although the resulting lower form familiarity could lead to more errors in recognition (Jordan, Redwood, & Patchings, 2003), the equivalent accuracy during training between the two groups shows no problems with perceptual encoding.

REFERENCES

Anderson, J., & Fincham, J. M. (1994). Acquisition of procedural skills from examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1322–1340.

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition, 13*, 263–308.

Bartning, I. (2000). Gender agreement in L2 French: Preadvanced vs. advanced learners. *Studia Linguistica, 54*, 225–237.

Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 173–218). New York: Cambridge University Press.

Besnter, D., Coltheart, M., & Davelaar, E. J. (1984). Basic processes in reading: Computation of abstract letter identities. *Canadian Journal of Psychology, 38*, 126–134.

Cahill, L., Haier, R., Fallon, J., Alkire, M., Tang, C., Keator, D., et al. (1996). Amygdala activity at encoding correlated with long-term free recall of emotional information. *Proceedings of the National Academy of Sciences, 93*, 8016–8021.

Carroll, S. (2005). Input and SLA: Adults' sensitivity to different sorts of cues to French gender. *Language Learning, 55*, 79–138.

Chan, A., Lieven, E., & Tomasello, M. (2009). Children's understanding of the agent–patient relations in the transitive construction: Cross-linguistic comparisons between Cantonese, German, and English. *Cognitive Linguistics, 20*, 267–300.

de Graaff, R. (1997). The eXperanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition, 19*, 249–275.

DeKeyser, R., & Larson-Hall, J. (2005). What does the critical period really mean? In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*. Oxford: Oxford University Press.

Dewaele, J.-M. (1994). Variation synchronique des taux d'exactitude: Analyze de fréquence des erreurs morpholexiques des trois styles d'interlangue française. *International Review of Applied Linguistics, 33*, 275–300.

Dewaele, J.-M., & Véronique, D. (2001). Gender assignment and gender agreement in advanced French interlanguage: A cross-sectional study. *Bilingualism: Language and Cognition, 4*, 275–297.

Ducroquet, L. (1979). The dictation—An outdated exercise? *System, 7*, 125–129.

Elio, R., & Anderson, J. (1984). The effects of information order and learning mode on schema abstraction. *Memory and Cognition, 12*, 20–30.

Ellis, N. (1994). Consciousness in second language learning: Psychological perspectives on the role of conscious processes in vocabulary acquisition. *AILA Review, 11*, 37–56.

Ellis, N. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition, 24*, 143–188.

Ellis, N., & Sagarra, N. (2010). Learned attention effects in L2 temporal reference: The first hour and the next eight semesters. *Language Learning, 60*, 85–108.

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.

Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science, 208*, 1181–1182.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition, 65*, 263–297.

Goldberg, A. (1999). The emergence of the semantics of argument structure constructions. In B. MacWhinney (Ed.), *The emergence of language* (pp. 197–213). Mahwah, NJ: Erlbaum.

Goldberg, A. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Goldberg, A., Casenhiser, D., & Sethuraman, N. (2004). Learning argument struture generalizations. *Cognitive Linguistics, 15*, 289–316.

Grevisse, M., & Goosse, A. (2007). *Le bon usage: Grammaire Française*. Paris: Duculot.

Gullberg, M., Roberts, L., Dimroth, C., Veroude, K., & Indefrey, P. (2010). Adult language learning after minimal exposure to an unknown natural language. *Language Learning, 60*, 5–24.

Hardison, D. M. (1992). Acquisition of grammatical gender in French: L2 learner accuracy and strategies. *Canadian Modern Language Review, 48*, 715–728.

Harley, B. (1998). The role of form-focused tasks in promoting child L2 acquisition. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 156–174). New York: Cambridge University Press.

Holmes, V. M., & Segui, J. (2004). Sublexical and lexical influences on gender assignment in French. *Journal of Psycholinguistic Research, 33*, 425–457.

Hulstijn, J. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal, and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge: Cambridge University Press.

Hulstijn, J., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning, 51*, 539–558.

Jafarpur, A., & Yamini, M. (1993). Does practice with dictation improve language skills? *System, 21*, 359–369.

Jordan, T. R., Redwood, M., & Patchings, G. R. (2003). Effects of form familiarity on perception of words, pseudowords, and nonwords in the two cerebral hemispheres. *Journal of Cognitive Neuroscience, 15*, 537–548.

Kilborn, K., & Ito, T. (1989). Sentence processing in Japanese–English and Dutch–English bilinguals. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 257–291). New York: Cambridge University Press.

Krashen, S. (1994). The input hypothesis and its rivals. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 45–78). San Diego, CA: Academic Press.

Lewis, J. D., & Elman, J. (2001). *Learnability and the statistical structure of language: Poverty of stimulus arguments revisited.* Paper presented at the 26th Annual Boston University Conference on Language Development.

L'Huillier, M. (1999). *Advanced French grammar*. New York: Cambridge University Press.

Lightbown, P., & Spada, N. (1990). Focus-on-form and corrective feedback in communicative language teaching. *Studies in Second Language Acquisition, 12*, 429–448.

Lindsay, S., & Gaskell, M. (2010). A complementary systems account of word learning in L1 and L2. *Language Learning, 60*, 45–63.

Long, M. (2000). Focus on form in task-based language teaching. In R. Lambert & E. Shohamy (Eds.), *Language policy and pedagogy: Essays in honor of A. Ronald Walton* (pp. 179–192). Philadelphia, PA: John Benjamins.

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

Lyster, R. (2004b). Research on form-focused instruction in immersion classrooms: Implications for theory and practice. *French Language Studies, 14*, 321–341.

Lyster, R. (2006). Predictability in French gender attribution: A corpus analysis. *French Language Studies, 16*, 69–92.

Lyster, R., & Izquierdo, J. (2010). Prompts versus recasts in dyadic interaction. *Language Learning, 59*, 453–498.

Mackey, A. (2006). Feedback, noticing, and instructed second language learning. *Applied Linguistics, 27*, 405–430.

MacWhinney, B. (1982). Basic syntactic processes. In S. Kuczaj (Ed.), *Language acquisition: Vol. 1. Syntax and semantics* (pp. 73–136). Hillsdale, NJ: Erlbaum.

MacWhinney, B. (1997). Implicit and explicit processes. *Studies in Second Language Acquisition, 19*, 277–281.

MacWhinney, B. (2000). Lexicalist connectionism. In P. Broeder & J. Murre (Eds.), *Models of language acquisition: Inductive and deductive approaches* (pp. 9–32). Cambridge, MA: MIT Press.

MacWhinney, B. (2011). The logic of the unified model. In S. Gass & A. Mackey (Eds.), *Handbook of second language acquisition*. New York: Routledge.

MacWhinney, B. (in press). Item-based patterns in early syntactic development. In T. Herbst (Ed.), *Valency relations*. Berlin: Springer.

Manning, P. (1996). Exploratory teaching of grammar rules and CALL. *ReCALL, 8*, 24–30.

McClelland, A., & Rogers, T. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Merriman, W. (1999). Competition, attention, and young children's lexical processing. In B. MacWhinney (Ed.), *The emergence of language* (pp. 331–358). Mahwah, NJ: Erlbaum.

Ninio, A. (1999). Pathbreaking verbs in syntactic development and the question of prototypical transitivity. *Journal of Child Language, 26*, 619–653.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*, 417–528.

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

Pavlik, P. I., Jr., Presson, N., Dozzi, G., Wu, S., MacWhinney, B., & Koedinger, K. (2007). *The FaCT (Fact and Concept Training) System: A new tool linking Cognitive Science with educators.* Unpublished manuscript, Cognitive Science Society. Retrieved from http://csjarchive.cogsci.rpi.edu/proceedings/2007/docs/p1379.pdf

Pavlik, P. I., Jr., Presson, N. & Koedinger, K. (2007). Optimizing knowledge component learning using a dynamic structural model of practice. In R. Lewis & T. Polk (Eds.), *Proceedings of the Eighth International Conference of Cognitive Modeling*. Ann Arbor, MI: University of Michigan Press.

Perfetti, C. A., Bell, L. C., & Delaney, S. M. (1988). Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language, 27*, 59–70.

Pica, T. (1985). Second-language acquisition, social interaction, and the classroom. *Applied Linguistics, 8*, 3–21.

Pimsleur, P. (1967). A memory schedule. *Modern Language Journal, 51*, 73–75.

Posner, M., & Keele, S. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83*, 304–308.

Price, G. (2008). *A comprehensive French grammar*. Malden, MA: Blackwell.

Rast, R. (2010). The role of linguistic input in the first hours of adult language learning. *Language Learning, 60*, 64–84.

Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning, 45*, 285–331.

Robinson, P. (2005). Cognitive abilities, chunk-strength, and frequency effects in implicit artificial grammar and incidental L2 learning: Replications of Reber, Walkenfeld, and Hernstadt (1991) and Knowlton and Squire (1996) and their relevance for SLA. *Studies in Second Language Acquisition, 27*, 235–268.

Robinson, P. (2010). Implicit artificial grammar and incidental natural second language learning: How comparable are they? *Language Learning, 60*, 245–263.

Rutherford, W., & Sharwood-Smith, M. (1985). Consciousness raising and universal grammar. *Applied Linguistics, 6*, 274–282.

Sanz, C., & Morgan-Short, K. (2004). Positive evidence versus explicit rule presentation and explicit negative feedback: A computer-assisted study. *Language Learning, 54*, 35–78.

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics, 13*, 206–226.

Schmidt, R. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 165–210). San Diego, CA: Academic Press.

Schneider, W., & Chein, J. (2003). Controlled and automatic processing: Behavior, theory, and biological mechanisms. *Cognitive Science, 27*, 525–559.

Segalowitz, N., & Segalowitz, S. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics, 14*, 369–385.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language features: A meta-analysis. *Language Learning, 60*, 263–308.

Tucker, R., Lambert, W., & Rigault, A. (1977). *The French speaker's skill with grammatical gender: An example of rule-governed behavior*. The Hague: Mouton.

Ullman, M. (2001a). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research, 30*, 37–69.

Ullman, M. (2001b). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews: Neuroscience, 2*, 717–722.

Walker, J. (n.d.). *Le truc de genres.* Retrieved from http://www.fourmilab.ch/francais/gender.html

Wittenberg, G., Sullivan, M., & Tsien, J. (2002). Synaptic reentry reinforcement-based network model for long-term memory consolidation. *Hippocampus, 12,* 637–647.