Studies in Second Language Acquisition, 2016, **38**, 177–215.
doi:10.1017/S0272263115000388

THE EFFECTS OF IMPLICIT INSTRUCTION ON IMPLICIT AND EXPLICIT KNOWLEDGE DEVELOPMENT

Aline Godfroid
Michigan State University

This study extends the evidence for implicit second language (L2) learning, which comes largely from (semi-)artificial language research, to German. Upper-intermediate L2 German learners were flooded with spoken exemplars of a difficult morphological structure, namely strong, vowel-changing verbs. Toward the end of exposure, the mandatory vowel change was omitted, yielding ungrammatical verb forms (compare Leung & Williams, 2012). Two pre- and posttests—word monitoring and controlled oral production—gauged the development of learners' implicit and explicit knowledge, respectively.

Interviews revealed 33 out of 38 L2 learners remained unaware of the ungrammatical verbs in the input flood; however, they showed significant sensitivity during listening as evidenced by a reaction time slowdown on ungrammatical trials. The unaware learners also improved significantly from pretest to posttest on the word-monitoring task, but not the oral production measure, unless the verbs' salience in the input flood had resonated with them. Thus, implicit instruction affected implicit knowledge primarily, although prior knowledge and memory could potentially account for interactions between implicit processing, implicit knowledge, and explicit knowledge.

I owe a debt of gratitude to Jieun (Irene) Ahn, Sehoon Jung, and Jens Schmidtke for their dedication and help in programming the experiments and collecting and analyzing the data. I am also thankful to the faculty in the German program at Michigan State University for encouraging their students to participate in the study. This paper benefited substantially from the comments of two anonymous reviewers, Alex Housen, Le Anne Spino-Seijas, Kaitlyn Tagarelli, Koen Van Gorp, and the audiences at the 2014 conference of the American Association for Applied Linguistics and the symposium on New Directions in Implicit and Explicit Language Learning. Thanks also go to Gisela Granena for advice on the word-monitoring task. Despite all the helpful input from so many sources, I alone am responsible for any remaining errors.

© Cambridge University Press 2015

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

THE EFFECTS OF IMPLICIT INSTRUCTION ON IMPLICIT AND EXPLICIT KNOWLEDGE DEVELOPMENT

Understanding the scope of implicit, or unconscious, learning of second-language (L2) grammar is fundamental to understanding second language acquisition (SLA) and the relationship between child and adult language learning. As DeKeyser and Juffs (2005) put it, "nobody doubts that implicitly acquired procedural knowledge would be useful; the main question is to what extent it exists" (p. 441) and how it can be acquired. The current study aims to extend the support for implicit L2 learning obtained with artificial or semi-artificial languages (e.g., Leung & Williams, 2012) using a fully natural L2. Instruction focused on a subset of strong German verbs, while the effects of learning on knowledge were considered for these verbs as well as other, untrained verb exemplars. In so doing, the study investigated the interrelationship between instruction, learning, and knowledge for a target structure considered to be hard to acquire in terms of both item learning and system learning.

Implicit and Explicit Learning and Knowledge

Studying adult implicit and explicit language learning can help researchers understand the fundamental similarities and differences of SLA with other aspects of human cognition. Even though evidence for implicit adult L2 learning appears to be accruing (see Ellis, 2002; Williams, 2009, for reviews; but DeKeyser, 2003, and Leow, 2015, for more cautious positions), many questions remain and warrant an exploration of the topic beyond a narrow focus on whether learning without awareness is possible *in se*. For instance, to what extent and in what instructional contexts do implicit and explicit L2 learning occur? Are both types of learning beneficial to L2 speakers? Are there particular structures that lend themselves more readily to being learned either implicitly or explicitly? Answers to such questions can improve our understanding of how a L2 can best be learned and taught.

Stadler and Frensch (1994) equated implicit learning with unintentional learning, stating that "learning is implicit when the learning process is unaffected by intention" (p. 423). However, a lack of intentionality, which is most likely under *incidental* learning conditions, is necessary but not sufficient for learning to be implicit. A learner must also be unaware of what is being learned (DeKeyser, 2003) at the time of learning (Hama & Leow, 2010; Leow & Hama, 2013; Williams, 2009). Implicit learning, then, is "evident in the behavioral responses [learners] make" (Ellis, 2009, p. 3) without the learners' being aware of the changes in

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

their behavior (i.e., response speed or accuracy increases or decreases). This contrasts with explicit learning, which is both conscious and intentional (Rebuschat, 2013; Williams, 2009).

In spite of the terminological clarity, implicit learning has not been operationalized consistently. One well-known difficulty is how to understand and measure awareness of a language form (see Godfroid, Boers, & Housen, 2013; Leow, Johnson, & Zárate-Sández, 2010; Truscott & Sharwood Smith, 2011). A second issue is that the concept of learning is related to but distinct from knowledge and instruction (Schmidt, 1994a), yet the conceptual boundaries between these three notions are sometimes blurred in empirical research (see Ellis, 2009; Williams, 2009, for discussion). Hama and Leow (2010) and Leow and Hama (2013) addressed both points in their critiques of recent empirical work on implicit learning. The authors questioned implicit learning researchers' reliance on offline measures of awareness (e.g., retrospective interviews) to demonstrate implicit learning because this conflates the stages of memory encoding (i.e., learning viewed as a process) and memory retrieval (i.e., learning viewed as a product, or knowledge). While their critiques put methodological issues surrounding implicit learning research front and center (also see Shanks & St. John, 1994), the question seems to boil down to (a) whether retrospective measures suffer from memory decay and fabrication and (b) whether such validity concerns outweigh the risk of reactivity and nonveridicality inherent in measuring awareness concurrently, by means of think-aloud protocols. In my opinion, the jury is still out on these issues, as extant research has produced mixed findings (Bowles, 2010; Godfroid & Spino, 2015, in press; Hama & Leow, 2010; Leung & Williams, 2011, p. 37; Leung & Williams, 2012, p. 638; Rebuschat, Hamrick, Sachs, Riestenberg, & Ziegler, 2015).

Although implicit learning is assumed to result in implicit knowledge and explicit learning likely leads to explicit knowledge (e.g., Reber, Allen, & Reber, 1999; Rebuschat & Williams, 2012), one type of knowledge can also transform into another, for example through proceduralization and automatization (DeKeyser, 2015), analysis (Bialystok, 1994), or an insight (Williams, 2009). Of these processes, automatization (preceded by the more rapid process of knowledge proceduralization) is the only process to require extensive time and practice (DeKeyser, 2015). In contrast, analysis and insights, which both result in explicit knowledge, could in principle develop over the course of a single treatment.

Implicit and Explicit Instruction

A factor complicating the preceding description is that implicit and explicit knowledge may also develop simultaneously during instruction

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

(DeKeyser, 2009; Dienes & Perner, 1999; Paradis, 2009), and that instruction, like learning and knowledge, can be either implicit or explicit. Implicit instruction is characterized by an absence of rules or rule-search instructions (Hulstijn, 2005; Norris & Ortega, 2001); learners are exposed to relevant exemplars of a target structure in a meaning-focused task in the hope that they will infer patterns or rules. In explicit treatments learners are either provided with rules (deductive and metalinguistic; Norris & Ortega, 2001) or are instructed to look for them in the input (explicit induction; ibid.).

Although early comparative work on instruction relied heavily on explicit knowledge measures, which biased results in favor of explicit instruction (Doughty, 2003; Norris & Ortega, 2001), the recent development and validation of implicit knowledge measures (e.g., Andringa & Ćurčić, 2015; Ellis, 2005; Erlam, 2006; Godfroid et al., 2015; Granena, 2013; Jiang, 2007) has paved the way for more balanced studies of implicit and explicit instruction. Two studies are of particular relevance here. De Jong (2005) investigated the effects of implicit instruction on the development of receptive and productive knowledge of noun-adjective gender agreement in L2 Spanish. Participants were novice learners of Spanish who received either receptive training or receptive and productive training on the target structure. Training consisted of meaning-focused processing of spoken-language input and, in the case of the receptive plus productive group, spoken output. A battery of four auditory posttests assessed the learning effects. De Jong found "an effect of the direction of processing (comprehension or production)" (p. 227). The receptive group performed best on the receptive posttest measures, whereas the receptive plus productive group performed better on the production posttest. An interesting finding was that nearly all implicitly trained participants detected the target structure and more than half of them induced at least partial rule knowledge (compare De Graaff, 1997; DeKeyser, 1995). Therefore, De Jong felt that "despite the variety of measures, no firm conclusions can be drawn as to the type of knowledge—implicit or explicit—that was acquired" (p. 229).

Andringa, De Glopper, and Haquebord (2011) compared the effects of implicit and explicit grammar instruction in L2 Dutch, using two posttest measures targeting explicit and implicit knowledge, respectively. Explicit knowledge was measured by means of a grammaticality judgment test, whereas the proposed measure of implicit knowledge was a free written-response task. Of special interest was whether explicit instruction would yield larger gains on the free response measure than implicit instruction, given equal amounts of input to both groups. Results indicated explicit instruction and implicit instruction resulted in comparable amounts of improvement on the free written-response task. Explicit instruction had no apparent added value, except for a subset of learners with similar morphological structures in their L1s.

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

From a methodological viewpoint, the use of the free written-response task as a measure of implicit knowledge would be in need of further validation, given that performance on this test and the grammaticality judgment test correlated moderately strongly ($rs \geq .50$).

The present study shares with De Jong (2005) a focus on auditory implicit instruction. It also continues De Jong's, Andringa and colleagues', and other researchers' attempts to implement relatively separate measures of implicit and explicit knowledge in instructed SLA research. For that purpose, I draw on Granena's (2013) study on ultimate attainment. Granena investigated the relationship between early and late bilinguals' implicit learning aptitude and their L2 morphosyntactic attainment to clarify the role of individual differences in early learners' long-term L2 achievement. Granena administered two sequence learning tasks as measures of implicit learning aptitude and a word-monitoring task as a hypothetical measure of implicit language knowledge. Word monitoring is a crossmodal, dual-processing task that originated in L1 research (see Marslen-Wilson & Tyler, 1980) and drives participants to focus on meaning rather than form. Using both aptitude and knowledge measures, Granena (2013) found that sequence learning ability played a role in word-monitoring performance for late bilinguals and early bilinguals alike. Crucially, late bilinguals' word monitoring was significantly related to the nonverbal aptitude measure, which supported the validity of the word-monitoring task as an implicit knowledge measure (also see Suzuki & DeKeyser, 2015).

Input Flooding as a Type of Implicit Instruction

Input flooding is a type of instruction that falls under the broad category of input enhancement or consciousness raising (Sharwood Smith, 1981, 1991). Input enhancement consists of making elements in the input stand out so they become externally (perceptually) more salient, in hopes that this will promote the forms' internal (psychological) salience and acquisition (Sharwood Smith, 1991). In the case of input flooding, the enhancement is achieved by increasing the frequency of the target structure in the input. The assumption is that such externally induced salience may help learners to orient (i.e., allocate attentional resources) to the target form during meaning-focused processing, which in turn may boost the chances of detection, or memory registration (Tomlin & Villa, 1994).

Although it is not usually labeled that way, input flood has been one of the preferred instructional techniques in laboratory studies on incidental or implicit learning. For instance, Leung and Williams exposed their participants to large numbers of simple English sentences

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

(Leung & Williams, 2011; Williams, 2004, 2005) or English noun phrases (Leung & Williams, 2012, 2014) that participants had to process for meaning and repeat out loud. Each sentence contained one of four artificial determiners—*gi*, *ro*, *ul*, or *ne*—which were introduced to participants as denoting either distance (near/far) or the difference between adults and children. Unbeknownst to participants, the determiners also encoded another semantic distinction: animacy, thematic role, relative size, an orthographic-visual distinction, or a shape distinction. The authors wanted to know whether participants would acquire the hidden form-meaning regularity implicitly during meaning-focused processing. Using a reaction-time (RT) methodology, Leung and Williams (2011, 2012, 2014) found that even participants who could not state the rule, and therefore qualified as potential implicit learners, showed learning of thematic role, animacy, and the shape distinction as a result of input flooding. The linguistic encoding of objects' relative size or orthographic-visual differences could not be acquired implicitly, which suggested there are bounds on the human capacity for implicit learning.

The Use of Artificial and Semi-artificial Languages in Implicit Learning Research

Over the past 25 years, SLA researchers have extended artificial and semi-artificial language research in cognitive psychology to include more meaningful, but still artificial, languages (see Hulstijn & DeKeyser, 1997, for an early overview). A clear advantage of artificial and semi-artificial languages is that they allow controlling for participants' prior knowledge and input of the language to be learned (e.g., De Graaff, 1997; DeKeyser, 1995; Morgan-Short, Sanz, Steinhauer, & Ullman, 2012; Yang & Givón, 1997). Morgan-Short et al. (2012) argued that their artificial language Brocanto2 served as a "test tube [model] of natural language" (p. 934), the findings of which generalized to natural language learning (Morgan-Short, Faretta-Stutenberg, Brill-Schuetz, Carpenter, & Wong, 2014). Morgan-Short and colleagues based their claims on the finding that Brocanto2 could be used productively and that its predecessor, Brocanto, was found to elicit L1-like brain activity when learned to high proficiency (Friederici, Steinhauer, & Pfeifer 2002; Opitz & Friederici, 2003, 2004). Other artificial languages were used by De Graaff (1997), DeKeyser (1995), and Yang and Givón (1997). All emphasized that their artificial languages carried meaning, which, in the authors' minds, made their languages more similar to natural languages than to the artificial grammars used in cognitive psychology.

In other studies, semi-artificial languages have been used in order to replicate a natural language learning environment even more closely.

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

By and large, these semi-artificial language studies have focused on the acquisition of morphology (Faretta-Stutenberg & Morgan-Short, 2011; Hama & Leow, 2010; Leung & Williams, 2011, 2012, 2014; Rebuschat et al., 2015; Williams, 2004, 2005); syntax (Grey, Williams, & Rebuschat, 2014; Rebuschat & Williams, 2012; Robinson, 1997; Tagarelli, Borges Mota, & Rebuschat, 2015); or both (Williams & Kuribara, 2008). All of these studies showed learning under incidental or implicit, meaning-based exposure conditions, and some claimed to have found evidence for the implicit learning of adult L2 grammar (Leung & Williams, 2011, 2012, 2014; Rebuschat & Williams, 2012; Williams, 2005). Thus, artificial and semi-artificial language research has been informative and quite productive, as researchers have tended to reuse the same languages and have thereby added to a cumulative body of knowledge. Even so, the concerns about ecological validity in these studies are evident and the need to test the generalizability of findings to natural language learning is large (see Ettlinger, Morgan-Short, Faretta-Stutenberg, & Wong, 2015).

Artificial or semi-artificial language learning may be somewhat different from natural language learning because of the increased saliency[1] of the target language forms (R. DeKeyser, personal communication, October 31, 2013). This is especially so when artificial morphemes are inserted into natural language sentences, as is customary in semi-artificial language studies about morphology. Because saliency is a determinant of attention (e.g., Chun, Golomb, & Turk-Browne, 2011; Itti & Koch, 2000; Knudsen, 2007), the use of a semi-artificial language could influence and potentially alter the cognitive processes underlying natural language learning, for instance by enhancing attention and learning. In a similar vein, Morgan-Short et al. (2014) acknowledged that participants in artificial language studies are always aware that they are learning a new language (p. 63). Therefore, it is important to assess the generalizability of findings from these two paradigms to natural language learning, where the salience of a target form is not confounded with its synthetic nature (also see Robinson, 2010, Schmidt, 1994b, and VanPatten, 1994, for critical discussions of the relevance of findings from artificial grammar learning to SLA).

Strong German Verbs as a L2 Learning Problem

Research investigating the L2 acquisition of verb morphology indicates that adult learners have persistent difficulty acquiring and using inflectional verb morphology (for reviews, see DeKeyser, 2005; Ellis, 2006; Lardiere, 2006; Larsen-Freeman, 2010; Prévost & White, 2000). Compared to the extensive SLA literature on verb affixation and, to a lesser extent, verb suppletion, verb allomorphy has not been studied much (Krause, Bosch, and Clahsen, 2015; but see Leow, 1997, 2000; Murphy, 2004) despite

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

the frequency with which allomorphy occurs in the world's languages (Bickel & Nichols, 2007). In German, like in English, verbs that exhibit allomorphy are known as strong verbs. Table 1 presents the two types of verb allomorphy focused on in this study.

Strong verbs are considered marked (relative to weak verbs) because they have different stem allomorphs to encode person, number, tense, and mood information (Krause et al., 2015). Whereas English strong verbs differ from weak verbs only in their preterit and participle stems (e.g., *speak – spoke – spoken*), stem allomorphy also occurs in the present tense and some imperative forms in German (Table 1). These stem variants distinguish the German strong verbs from weak (unmarked or regular) verbs, which convey the same information through affixation only. It is argued that this information redundancy makes German strong verbs difficult to acquire (see DeKeyser, 2005; Ellis, 2006; VanPatten, 1996) because the subject, the verb stem, and the prefix or suffix all encode the same morphosyntactic information (Table 1). Adding to this learning problem is the fact that stem variants are perceptually nonsalient because the stem changes affect one or two letters or phonemes only (Godfroid & Uggen, 2013).

The vowel alternation that characterizes strong verbs was phonologically conditioned in older variants of German (Nübling, Dammel, Duke, & Szczepaniak, 2006); however, because the phonological trigger has disappeared from contemporary German, it is now difficult to predict which verbs will undergo a vowel change (see Bybee & Newman, 1995). As a result, the stem-change pattern is now only "partially productive" (Ullman, 2001, p. 41). Weak verb conjugation has become the default paradigm. This is why Ullman (2001), in his Declarative-Procedural model, holds that verb allomorphs are stored in the mental lexicon rather than the grammar. Even so, allomorphs are said to differ from

Table 1. Abbreviated verb paradigm of two strong verb types: *e* → *i* and *a* → *ä*.

sprech-en		*trag-en*	
"talk"		"carry"	
ich sprech-e	"I talk"	*ich trag-e*	"I carry"
du sprich-st	"you talk"	*du träg-st*	"you carry"
er/sie sprich-t	"he/she talks"	*er/sie träg-t*	"he/she carries"
sprich	"talk" (imperative)	*trag-e*	"carry" (imperative)
sprach	"talked" (preterit)	*trug*	"carried" (preterit)
ge-sproch-en	"talked" (past participle)	*ge-trag-en*	"carried" (past participle)

Note. ä is pronounced /æ/ or /ɛ:/; *i* is pronounced /I/; and *ie* is pronounced /i:/. The *ie* diphthong is a phonologically conditioned variant of the *e* → *i* change, as in *sehen* "to see" → *sieht* "sees."

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

lexical items in that allomorphs form structured representations that contain morphophonological and phonological information (Ullman, 2001). "Unlike a rote memory, this associative memory is productive, although the extent of its productivity remains unclear" (p. 42). In contrast to Ullman's Declarative-Procedural model, usage-based approaches reject the lexicon-grammar distinction, but agree with characterizing strong verbs in terms of associative networks (Bybee, 1995; Bybee & Newman, 1995; Ellis, 2002). In accordance with this view, DeKeyser (1995) argued that clusters of marked verb stems "form a prototype in the sense that, while there is no rule (i.e., no general, mechanical way to determine whether a verb will show [a stem alternation]), those [marked verb forms] are not merely exceptions either" (p. 385). Thus, although verb stem alternations are sometimes characterized as a purely lexical problem, in fact, different theories suggest that strong verbs form fuzzy grammatical systems in which phonological similarity is important (see Godfroid & Uggen, 2013, and Krause et al., 2015, for elaboration). If this is correct, we would expect strong verbs to lend themselves to both item learning and system learning. Put differently, the effects of strong-verb instruction should generalize to other verbs in the same category.

Krause et al. (2015) investigated how strong German verbs are represented in L1 and L2 speakers' mental lexicon. They conducted a crossmodal priming experiment, in which an auditory prime preceded a visual target word, and participants were asked to judge the word/nonword status of the target. Of interest was the amount of priming from a marked stem (*wirft* → *WERFEN*; "throws → THROW")[2] as compared to that of an unmarked stem (*werfen* → *WIRFT*; "throw → THROWS"). The authors found an asymmetric priming pattern between the native (NS) and nonnative speaker (NNS) groups, whereby the NSs responded faster following marked verb primes, but the highly advanced L2 speakers showed the opposite pattern. In addition, the L2 speakers responded more slowly to control primes with marked verbs (e.g., *wirft* → *WIRFT*) than unmarked verbs (e.g., *werfen* → *WERFEN*), which pointed to their general difficulties with verbs with changed stem vowels. Krause and colleagues interpreted these findings in light of L2 learners' more general problems with morphosyntax, "specifically with markers of finiteness" (p. 20).

Godfroid and Uggen (2013) used eye-tracking to investigate whether beginning learners of German distinguished between, or "noticed" (Schmidt, 1990), German stem variants, at a time when the learners had only been taught how to conjugate weak verbs. The participants' eye-movement records revealed that the beginning learners, who had not received formal instruction on strong verbs yet, paid more attention to strong verb pairs (alternating stems) than weak verb pairs that appeared in stacked sentences on a computer screen. Furthermore, two types of eye-movement behavior predicted the learners' improvement on a written

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

production posttest: (a) longer eye fixation times on the marked verb stems and (b) visual comparisons between the marked stems (e.g., *trägt*, "carries") and unmarked stems (e.g., *trage*, "carry"). However, such comparisons only benefited the acquisition of *a* → *ä* verbs, not *e* → *i* verbs, possibly because the umlaut was a salient written marker that attracted participants' eye gaze in the case of comparisons. On trials in which learners did not compare the two verb forms directly, *e* → *i* verbs were produced significantly more accurately than *a* → *ä* verbs on the posttest.

The current study aims to extend Godfroid and Uggen's research (a) into the auditory modality, (b) using multiple posttests that measure both implicit and explicit knowledge, and (c) with trained and untrained verbs. The following research questions (RQs) guided the study:

1. Do upper-intermediate L2 German learners show online sensitivity to inflection errors in strong verbs during input flooding?
2. Does input flooding facilitate the development of implicit and explicit L2 knowledge of trained strong verbs?
3. Do the effects of input flooding generalize to other, untrained strong verbs?
4. Are there any differences in sensitivity or knowledge depending on verb type (i.e., *a* → *ä* or *e* → *i*)?

Based on learners' well-attested difficulties with L2 morphology (DeKeyser, 2005; Ellis, 2006; Lardiere, 2006; Larsen-Freeman, 2010; Prévost & White, 2000), it is hypothesized that upper-intermediate L2 German learners will not be sensitive to inflection errors during input flooding (RQ1). Because sensitivity (increased RTs to ungrammatical forms) is taken as evidence for implicit knowledge (Granena, 2013; Jiang, 2007; Lim & Christianson, 2015) or implicit learning (Leung & Williams, 2011, 2012, 2014), a lack of sensitivity would suggest input flooding did not promote implicit learning or reveal the existence of implicit knowledge. In either case, it is expected that implicit instruction (i.e., the input flood) will not produce pre- to posttest gains in participants' implicit knowledge, as measured by a comparison of participants' word-monitoring performance before and after input flooding (RQ2). It is, however, predicted that implicit instruction will be favorable for L2 learners' explicit knowledge development because of the many strong-verb exemplars that participants will hear (RQ2). More specifically, it is hypothesized that even if learners do not show full-fledged implicit knowledge, item-based knowledge may still accrue during the input flood and will manifest itself under test conditions that allow for conscious monitoring. Learners may improve more on *e* → *i* verbs than *a* → *ä* verbs following Godfroid and Uggen's (2013) findings of verb-type-related differences with beginning learners (RQ4). However, because the improvement is hypothesized to be item-based, transfer to other, untrained strong verbs is not expected (RQ3).

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

METHOD

Participants

Eighty-two intermediate-level L2 German learners, who were completing 300- or 400-level German courses at Michigan State University, participated in this study. They were divided into two groups: an experimental group that received input flood (final n = 38) and a no-treatment, NNS control group (final n = 32). Twelve German heritage learners were excluded from data analysis. Information obtained from a background questionnaire showed that the majority of participants spoke English as their first language (L1). Three participants in the experimental group reported Albanian, Korean, and Kannada as their respective L1s.

In addition, 21 NSs of German served as a NS control group for this study. The NS controls completed the main task (the input flood) only. The purpose of including NSs was to obtain supporting evidence for the validity of the input flood as a measure that can elicit implicit knowledge. Specifically, for the input flood to work properly, the NSs had to show grammatical sensitivity on the violation trials (see "Input Flood" in "Materials" section for further details). Although all NS controls spoke German as their L1, four participants reported that they were bilingual in German and English. Table 2 summarizes other demographic and

Table 2. Demographic and linguistic profiles of the three participant groups

	NNSs with input flood (n = 38)	NNS Controls (n = 32)	NS Controls (n = 21)
Gender	20 females, 18 males	14 females, 18 males	12 females, 9 males
Age in years	20.53 (*SD* = 2.21)	20.47 (*SD* = 1.32)	35.36 (*SD* = 15.00)
Years of instruction	5.68 (*SD* = 1.83)	6.02 (*SD* = 1.92)	n/a
Self-rated German proficiency			n/a
Listening	3.53 (*SD* = 0.73)	3.63 (*SD* = 0.55)	
Reading	3.45 (*SD* = 0.69)	3.59 (*SD* = 0.56)	
Speaking	2.92 (*SD* = 0.82)	3.00 (*SD* = 0.62)	
Writing	3.24 (*SD* = 0.63)	3.34 (*SD* = 0.60)	
Number of participants who had studied other languages besides German and English	19 (50%)	16 (50%)	20 (95%)

Notes. Proficiency was rated on a 5-point Likert scale, with 1 = beginner and 5 = advanced. N/a = not applicable.

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

linguistic information for the three groups. A series of Mann-Whitney *U* tests on the data of the L2 learners with input flood versus the NNS controls revealed no significant differences in the NNS groups' self-rated listening (z = -0.67, p = .503, r = .08), speaking (z = -0.51, p = .609, r = .06), reading (z = -1.18, p = .237, r = .14), or writing (z = -0.71, p = .481, r = .09).

Materials

The research materials included three components: (a) a learning task (i.e., input flood with sentence-picture matching), (b) two assessment tasks (i.e., word monitoring and controlled oral production), and (c) a measure of awareness (i.e., debriefing interview). Both the picture matching task and the word-monitoring task were administered on a laptop computer using E-prime. The controlled oral production task and the interview were audio-recorded.

The Learning Task. The learning task consisted of an input flood with sentence-picture matching, henceforth referred to as the sentence-picture matching task or picture matching task. In this task, each sentence was presented aurally to participants along with two pictures on a computer screen (see Figure 1). Participants could preview the pictures for 2 s. Once the sentence began to play, their task was to select, as quickly as possible, the picture that best represented the sentence they heard. The instructions emphasized that participants did not need to

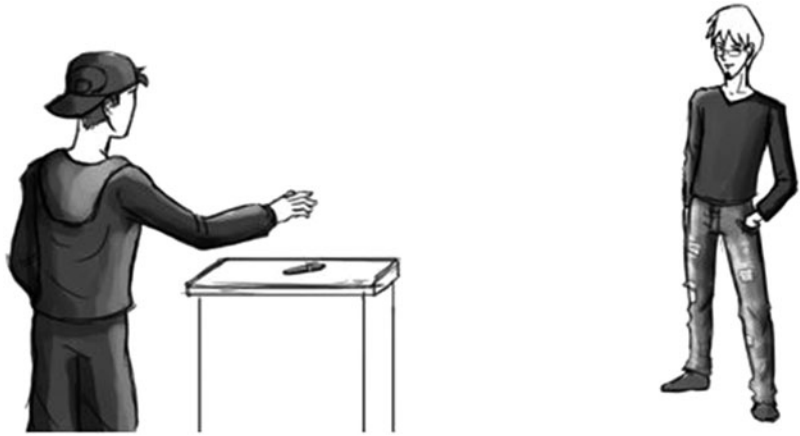


Figure 1. A sample trial of the sentence-picture matching task. Participants heard either *Er nimmt den Kuli* ("He takes the pen") in a control trial or * *Er nehmt den Kuli* in the violation block.
Note. Stimuli were counterbalanced for grammaticality.

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

listen until the end of the sentence to make a decision. All the sentences were simple sentences with subject-verb-object word order. The pictures, which were custom-made for the study, were paired such that on any given trial, both pictures depicted agents of the same gender: either a man and a man, or a woman and a woman. This made the verb the first reliable cue about the event denoted by the sentence. Consequently, RTs were measured from the onset of the verb onward.

The sentence-picture matching task produced two types of output, comprehension scores and RTs. The comprehension scores reflected participants' meaning comprehension ability and served as a check for accurate task completion. The primary data were the RTs, which I used to gauge participants' sensitivity to the targeted strong verbs. Essentially, evidence for sensitivity came from a significant slowdown in a violation block, which contained ungrammatical verb forms, compared to two control blocks, which contained grammatical verb forms.

Unbeknownst to the participants, the task consisted of three phases: a training phase, a transition phase, and a testing phase (see Table 3 and compare Leung & Williams, 2011, 2012, 2014). During the training phase, participants were flooded with auditory input (144 sentences) while their attention was directed to the meaning of the stimulus sentences for picture matching. This phase was followed by a brief transition phase (four sentences) and then the testing phase consisting of three blocks (24 sentences for each block): a first control block (henceforth, control-pre block), a violation block, and a second control block (henceforth, control-post block).[3]

In the training phase, participants were exposed to a mixture of weak and strong verbs in the first- and third-person singular, present tense. Of these, only the third-person singular, strong verb forms contained a vowel change: see Appendix for a list of verbs. From the transition block onward, participants heard only sentences with strong verbs in the third-person singular (i.e., marked verbs, requiring a vowel change). Verbs in the control blocks were conjugated correctly, whereas verbs in the violation block had not undergone a vowel change and were, therefore, ungrammatical. Figure 1 illustrates the difference in auditory input depending on whether a trial occurred in the violation block or in the training or control phases of the experiment. Three versions of the sentence-picture matching task were created, with sentences in the testing phase rotating between the control-pre, violation, and control-post blocks to control for any stimulus effects. Therefore, participants heard only one version of any given sentence.

Participants who possessed implicit knowledge of stem-changing verbs were expected to respond more slowly to violation trials than to grammatical control trials. In other words, by comparing participants' RTs for the violation block and the two control blocks, it could be inferred whether participants showed implicit knowledge of the target structure toward the end of the input flood. To determine the source of

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

Table 3. Distribution of verb types in the sentence-picture matching task

Phase	Verbs used	Grammaticality	Properties of verbs	No. of trials
Training*	24 verbs used in 1st and 3rd person singular forms	gramm.	[12 reg. Vs in 1st person sing.] × 2	24
		gramm.	[12 reg. Vs in 3rd person sing.] × 2	24
		gramm.	[12 irreg. Vs in 1st person sing.] × 2	24
		gramm.	[12 irreg. Vs in 3rd person sing.] × 6	72
			Subtotal	*144*
Transition**	4 verbs used in 3rd person singular	gramm.	[4 irreg. Vs in 3rd person sing.] × 1	4
Control Pre**	12 verbs used in 3rd person singular	gramm.	[12 irreg. Vs in 3rd person sing.] × 2	24
Violation**	12 verbs used in 3rd person singular	ungramm.	[12 irreg. Vs in 3rd person sing.] × 2	24
Control Post**	12 verbs used in 3rd person singular	gramm.	[12 irreg. Vs in 3rd person sing.] × 2	24
			Subtotal	*76*
			Grand total	*220*

Notes. Gramm. = grammatical; Ungramm. = ungrammatical; Reg. = regular or weak; Irreg. = irregular or strong; Sing. = singular; Vs = verbs; × [NUMBER] indicates how often a given verb occurred in that phase of the experiment.
* Items reused within block; ** All new items within block.

such sensitivity—that is, learning from input flood or prior knowledge—it was necessary to triangulate the results from the learning task with the assessment tasks. In particular, it was important to rule out the existence of prior implicit knowledge about strong verbs to be able to interpret the RT slowdowns in the input flood as real-time evidence for learning.

The Assessment Tasks. Two types of assessment task—a word-monitoring task and controlled oral production—were employed as pretests and posttests to gauge the learning outcomes brought about by the input flood. By incorporating two types of task, I sought to obtain a more complete and detailed picture of the resultant knowledge of the implicit instructional treatment (the input flood), while accounting for prior knowledge. In particular, the word-monitoring task was proposed to measure the development of learners' implicit knowledge (Granena, 2013; Suzuki & DeKeyser, 2015) whereas the controlled oral production task measured the development of learners' explicit, declarative knowledge.[4]

Word-Monitoring Task. A word-monitoring task was administered before and after the input flood to measure changes in learners' receptive, automatized, and implicit knowledge (Granena, 2013; Jiang, 2007; Suzuki & DeKeyser, 2015). The word-monitoring task imposed dual-task conditions on participants: word monitoring and sentence comprehension. For word monitoring, a content word was displayed on a computer screen for 500 ms, after which a spoken sentence, which might or might not contain the targeted content word, was presented. The participants were asked to monitor whether the target word occurred in the auditory sentence and to respond as quickly as possible if it did. For the comprehension component, a true/false question followed half of the sentences to ensure that participants also processed the sentences for meaning. To ascertain that participants engaged in dual-task processing as dictated by the task instructions, they needed to meet a 70% comprehension accuracy threshold to be included in the data analyses.

In the experimental items, the word to be monitored was always present. Crucially, it occurred immediately after a strong verb form, which might or might not have undergone a mandatory vowel change, as shown in the following examples:

(1) Der Reporter *fährt*$_{\text{(onset timer)}}$ seinen Jeep quer durch die afrikanische Wüste.
(2) Der Reporter * *fahrt*$_{\text{(onset timer)}}$ seinen Jeep quer durch die afrikanische Wüste.

"The reporter drives his jeep straight across the African desert."

RTs were measured from the offset of the verb. The rationale behind the word-monitoring task is that participants will unconsciously delay their

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

response to a target word (e.g., *Jeep*) that follows an ungrammatical form (e.g., *fahrt*), compared to a grammatical form (e.g., *fährt*), if they have integrated, implicit knowledge of the grammar form. By contrast, if a participant has incomplete or little implicit grammar knowledge, his or her response times will be unaffected by the preceding ungrammaticality.

The word-monitoring task consisted of 75 sentences, 45 fillers, and 30 experimental items. All experimental items contained one of the two strong verb types, either *a* → *ä* or *e* → *i* verbs. Half the verbs (k = 15) were grammatical (i.e., marked verb forms) while the other half (k = 15) was ungrammatical (i.e., unmarked verb forms), and the grammaticality of sentences was counterbalanced between participants. Twelve experimental items tested the verbs encountered in the input flood (the so-called *trained verbs*), whereas the remaining 18 items tested *untrained verbs* (see Appendix). The purpose of including untrained strong verbs was to test whether participants' sensitivity to rule violations would generalize to other strong verbs after participants had been flooded with a limited subset of stem-changing verbs.

Unlike the experimental items, the fillers contained an arbitrary regular verb, and target words occurred sometimes but not all the time. Importantly, the target word, if any, no longer appeared immediately after the verb in the fillers, so as to avoid strategic processing. The fillers consisted of a mix of grammatical sentences and sentences with case or gender agreement errors.

For scoring, a Grammatical Sensitivity Index (GSI) was calculated and used as the dependent variable, following Granena (2013). The GSI is obtained by subtracting the RTs on grammatical items from the RTs on ungrammatical items (i.e., $RTs_{ungrammatical} - RTs_{grammatical}$). A high GSI indicates a slowdown in the ungrammatical condition, which is assumed to reflect grammatical sensitivity. In contrast, GSIs close to zero and negative GSIs suggest a lack of sensitivity or reverse sensitivity, which point to a lack of implicit, integrated knowledge (no sensitivity) or a different interlanguage grammar rule (reverse sensitivity).

Controlled Oral Production Task. A picture-based oral production task was used as a pre- and posttest to assess the development of explicit, declarative knowledge. The test consisted of 36 pictures of events along with written prompts (content words and infinitives) that were meant to guide the description of the events. Participants were required to describe each picture in German, using the provided keywords (see Figure 2). They needed to add and inflect function words and conjugate the infinitives (including the stem-vowel change, when necessary) in order to produce correct sentences. Because the participants' output was constrained by the lexical descriptors accompanying each picture, this outcome variable was a "constrained constructed response task" (Norris & Ortega, 2001, p. 169) hypothesized to measure

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

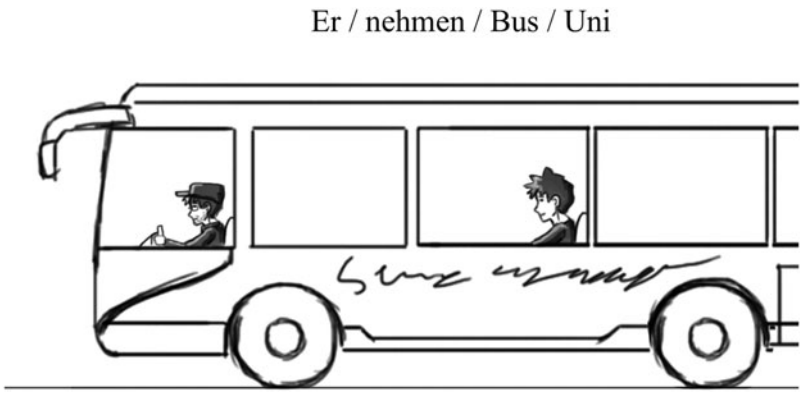


Figure 2. A sample trial of the controlled oral production task. The targeted sentence is *Er nimmt den Bus zur Uni*, "He takes the bus to campus".

"the application of explicit declarative knowledge under controlled conditions" (ibid.). Participants could monitor the accuracy of their own output (e.g., Krashen, 1985), given that the task was not time-pressured, had low communicative value, and made relatively small demands on memory because of the visual and verbal supports.

The pictures were presented in a random order for every participant. Twelve filler sentences contained regular (weak) verbs and 24 critical sentences contained irregular (strong) verbs. Among the 24 critical sentences, half (k = 12) included trained strong verbs, whereas the other half (k = 12) involved untrained strong verbs (see Appendix). Similarly to the word-monitoring task, the purpose of the untrained verbs was to measure participants' generalization ability, or system learning.

The controlled oral production data were scored in a binary fashion. Participants received one point if the target verb in their spoken sentence contained a correctly changed stem vowel and no credit otherwise (compare Godfroid & Uggen, 2013).

Debriefing Interview. All the L2 learners already possessed explicit, declarative knowledge of the strong-verb paradigm going into the study, given that strong verbs are taught halfway through the first semester in the German program at their university. Recruiting participants with some prior knowledge was a practical necessity because piloting had indicated that lower-level learners (e.g., German 101 students as in Godfroid & Uggen, 2013) could not perform a real-time listening task yet: they waited until the end of the sentence to make their judgments in the picture-matching and word-monitoring tasks, which made the RTs uninformative. However, possessing explicit rule knowledge does not guarantee that one is able to activate, let alone apply, that knowledge during real-time listening. Therefore, a debriefing interview was conducted to

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

probe whether participants had noticed (i.e., consciously registered) the occurrence of ungrammatical verbs in the sentence-picture matching and/or the word-monitoring tasks as a way to identify learners who were aware in the context of this study.

The debriefing interview was carried out in two phases: questions one and two immediately after the sentence-picture matching task and question three after the posttests. First, the researcher (a research assistant) prompted participants to report what they thought the experiment was about and whether they noticed anything odd at the end of the sentence-picture matching task. In the second phase, the researcher began by asking the same questions again and then led participants on their way by asking whether they felt there was anything special about the verbs that they had heard in the study. The purpose of the repeated questioning was to increase the likelihood that participants would also report instances of fleeting activation of conscious knowledge, thus increasing sensitivity of measurement (compare Rebuschat et al., 2015).

I classified as *aware* in this study the participants who reported hearing ungrammatical verbs in the learning or test tasks; *unaware* in this study were all other individuals, who did not activate or apply their conscious knowledge of strong verbs in the context of the present tasks.

Procedure

Each participant was tested individually in a quiet room. After filling out the consent form, the L2 learners took the two pretests: the word-monitoring task and the controlled oral production task, in that order. Then, they completed the learning task—the sentence-picture matching task, which consisted of the training phase as well as the testing phase. After finishing the sentence-picture matching task, the participants immediately answered the first part of the debriefing interview questions. After that, they took the posttests, the word-monitoring task, and the controlled oral production task. Both posttests were identical to their respective pretests, except that comprehension questions now appeared for the other half of the sentences in the word-monitoring task, so as to maintain the focus on meaning. (The two versions of the word-monitoring task were counterbalanced between participants.) Finally, the participants engaged in the second part of the debriefing interview and completed a background questionnaire. Unlike the L2 learners, the NSs completed only the main experiment, the interview, and the background questionnaire, not the pre- and posttests. The whole experiment lasted approximately 1h 15 min for the L2 learners and 30 min for the NSs.

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

Analysis

All the data from the learning task and the assessment tasks were entered into the Statistical Package for the Social Sciences, version 22. The main analyses consisted of three- or four-way mixed-design ANOVAs. First, the RTs from the sentence-picture matching task were analyzed to examine whether and to what extent the L2 German learners showed sensitivity to ungrammatical verb forms during input flooding (RQ1). Additionally, the L2 learners' pre- and posttest scores were analyzed to investigate the effects of input flood on the development of receptive/implicit knowledge as well productive/controlled knowledge (RQ2). Significant interactions between Time and Verb Training or Time and Verb Type would point to differential effects of instruction on trained and untrained verbs (RQ3) or on *a* → *ä* and *e* → *i* verbs (RQ4), respectively. The significance level α was set at .05 for all analyses. Assumptions were tested for each test and appropriate measures were taken in case they were not met. Effect sizes are reported with partial eta squared (η_p^2) for ANOVAs and Cohen's *d* for follow-up *t* tests.

RESULTS

Aware Participants, *Verbwise* Participants, and Fully Unaware Participants

Prior to the statistical analyses, I analyzed the retrospective verbal report data to identify the participants who became aware of the grammatical violations in the learning or test tasks. All NSs reported hearing the ungrammatical verb forms. In contrast, of the total of 70 NNSs, only seven participants reported awareness: five experimental participants and two controls.[5] Next, to acknowledge the long-standing criticism that not all unaware learners are alike (Berry & Dienes, 1993; Dienes & Berry, 1997; Ericsson & Simon, 1980), I examined the unaware learners' interview data more closely for potential traces of "incipient awareness" (Rebuschat et al., 2015, p. 328) or "fringe consciousness" (Norman, Price, Duff, & Mentzoni, 2007, p. 809). This revealed that 10 unaware learners (eight experimental, two controls) realized that the experiment focused on verbs; I will refer to these individuals as *verbwise* participants. The verbwise participants realized that many verbs were repeated in the study. Some verbwise participants who received input flood also realized that focusing on the verbs could help them respond faster. However, none of the verbwise participants reported hearing ungrammatical verbs in the experiment, and therefore, it was assumed that they did not activate their conscious knowledge of strong-verb conjugations during listening. The argument is that ungrammatical verb forms would clash with conscious

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

knowledge representations, leading participants to comment on the ungrammatical verbs' presence in the input when specifically asked about them. As a result, the verbwise and fully unaware participants will be treated as a single category of unaware learners for the main analyses reported in this study. The data for the aware learners and for the verbwise and fully unaware learners separately are available from the author upon request.

NSs and Unaware NNSs are Sensitive to Ungrammatical Verbs during Input Flood (RQs 1&4)

The first analysis was a comparison of participants' RTs in the picture matching task in the violation block and the control blocks. Any RTs falling outside of a subject's mean RT ± 2.5 *SD* for a given block were replaced by the respective cutoff values (i.e., mean + 2.5 *SD* or mean – 2.5 *SD*). This affected 1.41% of the data. The descriptive statistics for the cleaned RTs in each block are presented in Table 4.

To answer RQ1 (sensitivity), the main effect and interactions with Block are of interest. A 3 (Blocks) × 2 (Verb Type) × 2 (Language Status) mixed-design ANOVA on the RTs indicated the main effect of Blocks was significant with a large effect size, but the Block × Language Status interaction, the Block × Verb Type interaction, and the Block × Language Status × Verb Type interaction were not (see Table 5). The significant main effect of Blocks showed that for both *a* → *ä* and *e* → *i* verbs, NSs and unaware NNSs significantly slowed their responses to the violation

Table 4. NSs' and unaware NNSs' performance on the picture matching task (RTs in milliseconds)

		NS Controls (n = 21)		Unaware NNSs (n = 33)	
Blocks	Verb	*M*	*SD*	*M*	*SD*
Control Pre-	*a* → *ä*	1153	327	1288	317
	e → *i*	1170	351	1218	297
Violation	*a* → *ä*	1317	358	1349	264
	e → *i*	1321	425	1292	309
Control Post-	*a* → *ä*	1220	422	1299	360
	e → *i*	1285	386	1232	311
Mean Difference (Sensitivity)	*a* → *ä*	131	153	55	162
	e → *i*	75	127	67	137

Note. Sensitivity was computed as $RT_{Violation} - \left(\frac{RTControlPre + RTControlPost}{2}\right)$.

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

Table 5. Mixed-design ANOVA for the NSs' and the unaware NNSs' performance on the picture matching task

	SS	*df*	*MS*	*F*	*p*	η_p^2	power
Main effect (within-subject variable)							
Blocks	570054.88	1.78	320955.16	9.67	<.001***	.16	.97
Verb type	66263.94	1	66263.94	3.59	.064	.07	.46
Main effect (between-subject variable)							
Language status	95015.75	1	95015.75	0.16	.687	.00	.07
2-way interaction							
Block × Language status	122169.16	2	610084.58	2.07	.131	.04	.42
Block × Verb type	6886.70	2	3443.35	0.20	.817	.00	.08
Verb type × Language status	165877.99	1	165877.99	8.98	.004**	.15	.84
3-way interaction							
Block × Verb type × Language status	16214.29	2	8107.15	0.48	.622	.01	.13

* $p < .05$; ** $p < .01$; *** $p < .001$.

trials (M = 1320, SD = 315) compared to the control trials that included correctly inflected strong verbs (control-pre: M = 1217, SD = 303; control-post: M = 1260, SD = 350). The results, thus, suggest that the unaware L2 learners showed sensitivity to the ungrammatical verb forms. Furthermore, because the Block × Language Status interaction was nonsignificant, NNS sensitivity differed only in a small and statistically nonreliable amount from NSs' sensitivity. Finally, the nonsignificant Block × Verb Type and Block × Verb Type × Language Status interactions indicated that any verb-type-related differences in RT did not change over time (RQ4). In other words, the participants were equally sensitive to ungrammatical *a* → *ä* and *e* → *i* verb forms.

Changes in Unaware NNSs' Implicit Knowledge Representations after Input Flood (RQs 2-4)

To determine the source of the NNSs' implicit knowledge and to assess potential transfer to a different task, I compared word-monitoring pre- and posttest performance for unaware NNSs with and without input flood. As a reminder, these two groups are the experimental and NNS control groups, respectively. Across the two groups, four participants were excluded from the word-monitoring analyses due to their low comprehension rates (i.e., accuracy below 70%; n = 3) or because of a technical issue (n = 1). Therefore, the final number of unaware NNSs included

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

in the analysis was 59, consisting of 30 NNSs in the experimental group and 29 NNS controls. Only items that elicited correct answers to comprehension questions as well as correct monitoring responses were included. This amounted to 94.87% of the entire dataset. Finally, the raw RT data were inspected for outliers and values outside of the mean RT ± 2.5 *SD* were replaced by the corresponding cutoff values. This affected 1.4% of the data. The descriptive statistics for the cleaned RTs on the word-monitoring task are shown in Table 6. Test reliability was very good (Cronbach's alpha: pretest α = .90, posttest α = .84).

First, a 2 (Group) × 2 (Time) × 2 (Verb Type) × 2 (Verb Training) mixed-design ANOVA was carried out to examine the combined role of these variables on participants' GSI. The four-way ANOVA revealed a significant two-way, Time by Group interaction, but no other significant higher-order interactions. The lack of significant interactions with Verb Training (all *p*s > .114, all $\eta^2_p \leq .04$) suggests that input flood affected trained and untrained verbs similarly.

Therefore, the model was simplified to a 2 (Group) × 2 (Time) mixed-design ANOVA. This confirmed the significant Time by Group interaction ($F(1, 57) = 4.99$, $p = .029$, $\eta^2_p = .08$), as well as a significant main effect of Time ($F(1, 57) = 6.30$, $p = .015$, $\eta^2_p = .10$). The main effect of Group was not significant ($F(1, 57) = 1.32$, $p = .255$, $\eta^2_p = .02$).

To follow up on the Time by Group interaction, the change in participants' GSI (averaged over Verb Type and Verb Training) was evaluated for each group separately, using nonparametric Wilcoxon signed ranks tests. The analyses showed a large and significant increase in the experimental participants' GSI following input flood ($z = -2.75$, $p = .006$, $r = .50$) but no increase in the control group ($z = -0.13$, $p = .897$, $r = .02$). Together, these results suggest that the experimental participants' GSI values on the posttest (*M* = 20 msec, *SD* = 77) were reliably larger than their GSI values on the pretest (*M* = -82 msec, *SD* = 182), representing a shift toward targetlike sensitivity, from negative to positive GSI scores. However, while the negative GSI at pretest was significantly below zero (two-sided one-sample *t* test, $t(29) = -2.49$, $p = .019$, $d = 0.45$), the positive GSI at posttest did not differ from zero (two-sided one-sample *t* test, $t(29) = 1.42$, $p = .168$, $d = 0.26$). The change in participants' GSI also did not correlate with their sensitivity on the input flood. These findings suggest that the experimental participants entered the study with a non-targetlike interlanguage rule (i.e., a preference for unmarked strong such as * *er sprecht* "he talks" or * *er tragt* "he carries") that was undone, yet not reversed, after input flooding.

Finally, it should be noted that the control participants tended to differ from the experimental participants at the outset of the study, because they did not show a reliable preference for either marked (grammatical) or unmarked (ungrammatical) verb forms at pretest (control: *M* = -8 msec, *SD* = 98 versus experimental: *M* = -82, *SD* = 182, Mann-Whitney *U* test comparing the two groups: $z = -1.91$, $p = .056$, $r = .25$). However, as reported

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

Table 6. Unaware NNSs' performance on the word-monitoring task (RTs in milliseconds)

			NNS controls (n = 29)		NNSs with input flood (n = 30)	
			$a \rightarrow \ddot{a}$	$e \rightarrow i$	$a \rightarrow \ddot{a}$	$e \rightarrow i$
			M (SD)	*M (SD)*	*M (SD)*	*M (SD)*
Trained verbs	Pretest	Ungrammatical	1010 (200)	985 (308)	1002 (237)	908 (190)
		Grammatical	980 (170)	961 (265)	1066 (395)	988 (324)
		Difference (GSI)	30 (200)	24 (269)	-65 (312)	-81 (290)
	Posttest	Ungrammatical	975 (261)	910 (219)	852 (195)	775 (187)
		Grammatical	954 (206)	959 (297)	808 (138)	790 (189)
		Difference (GSI)	21 (274)	-49 (290)	44 (188)	-15 (171)
Untrained verbs	Pretest	Ungrammatical	955 (185)	962 (182)	939 (234)	997 (278)
		Grammatical	983 (246)	1019 (229)	1040 (405)	1079 (486)
		Difference (GSI)	-28 (189)	-57 (180)	-101 (293)	-83 (319)
	Posttest	Ungrammatical	901 (245)	972 (260)	787 (154)	889 (280)
		Grammatical	921 (223)	931 (211)	763 (152)	863 (186)
		Difference (GSI)	-21 (178)	41 (222)	24 (131)	26 (309)

in the preceding text, the control participants' GSI did not change over time, whereas the experimental participants' GSI did.

No Changes in Explicit Knowledge, Except for Verbwise Participants (RQs 2-4)

The final analysis considered the unaware learners' gains in controlled, productive knowledge, as indicated by a pretest-posttest comparison. Test reliability was acceptable, with Cronbach's alpha ranging from α = .74 on the pretest to α = .77 on the posttest. Table 7 presents the descriptive statistics for the unaware NNSs with and without input flood, respectively.

First, a 2 (Group) × 2 (Time) × 2 (Verb Type) × 2 (Verb Training) mixed-design ANOVA was performed to investigate whether production accuracy differed as a function of any of the variables of interest. This analysis revealed a borderline significant, three-way interaction between Time × Group × Verb Training ($F(1, 60) = 3.98$, $p = .051$, $\eta^2_p = .06$) and a significant Verb Type × Verb Training interaction ($F(1, 60) = 4.91$, $p = .030$, $\eta^2_p = .08$). Because Verb Training interacted with all other variables in the model, it was decided to analyze the data for trained and untrained verbs separately. Tables 8 and 9 present the results for the corresponding Time × Group × Verb Type ANOVAs. For both trained and untrained verbs, the analyses revealed significant main effects of Time and of Verb Type, with somewhat different effect sizes. Crucially, Time and Group did not interact, suggesting the experimental and control groups showed similar pre- to posttest improvements. In other words, the gains made by the experimental group could not be distinguished statistically from the test-retest effect observed for the control group. The main effect of Verb Type indicated a greater accuracy on *e* → *i* verbs than

Table 7. Unaware NNSs' test scores on the oral production task

		NNS Controls (n = 29)				NNSs with input flood (n = 30)			
		a → *ä*		*e* → *i*		*a* → *ä*		*e* → *i*	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Trained verbs	Pretest	2.76	1.86	4.36	1.47	2.33	1.49	4.39	1.39
	Posttest	2.99	1.69	4.32	1.51	2.94	1.58	4.70	1.21
Untrained verbs	Pretest	1.79	1.22	2.86	1.31	1.64	1.29	3.00	1.22
	Posttest	2.28	1.53	3.20	1.17	1.79	1.24	3.28	1.11

Note. The maximum total score in each cell was 6.

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

Table 8. Mixed-design ANOVA for the unaware NNSs' test scores on the oral production task: Trained verbs (item learning)

	SS	*df*	*MS*	*F*	*p*	η^2_p	power
Main effect (between-subject variable)							
Group	0.01	1	0.01	0.00	.958	.00	.05
Main effects (within-subject variable)							
Time	4.72	1	4.72	5.88	.018*	.09	.67
Verb Type	175.60	1	175.60	62.12	<.001***	.51	1.00
2-way interaction							
Group × Time	1.96	1	1.96	2.44	.123	.04	.34
Group × Verb Type	3.06	1	3.06	1.08	.302	.02	.18
Time × Verb Type	1.31	1	1.31	1.88	.176	.03	.27
3-way interaction							
Group × Time × Verb Type	0.00	1	0.00	0.00	.956	.00	.05

* $p < .05$; ** $p < .01$; *** $p < .001$.

a → *ä* verbs for pretest and posttest, and experimental and control participants alike.

Because verbwise participants' special insight could potentially serve them during oral production, I also investigated the role of Awareness (verbwise vs. fully unaware) in a post hoc analysis of the experimental group's data. As a reminder, the verbwise participants were those who realized the experiment focused on verbs, but did not report hearing ungrammatical verbs in the input flood or word-monitoring tasks. A 2 (Awareness) × 2 (Time) × 2 (Verb Type) × 2 (Verb Training) mixed-design ANOVA revealed a significant Awareness × Time × Verb

Table 9. Mixed-design ANOVA for the unaware NNSs' test scores on the oral production task: Untrained verbs (system learning)

	SS	*df*	*MS*	*F*	*p*	η^2_p	power
Main effect (between-subject variable)							
Group	0.65	1	0.65	0.20	.659	.00	.07
Main effects (within-subject variable)							
Time	6.20	1	6.20	10.92	.002**	.14	.90
Verb Type	91.04	1	01.04	44.64	<.001***	.43	1.00
2-way interaction							
Group × Time	0.60	1	0.60	1.06	.308	.02	.17
Group × Verb Type	2.88	1	2.88	1.41	.239	.02	.22
Time × Verb Type	0.00	1	0.00	0.00	.967	.00	.05
3-way interaction							
Group × Time × Verb Type	0.31	1	0.31	0.63	.430	.01	.12

* $p < .05$; ** $p < .01$; *** $p < .001$.

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

Training interaction ($F(1, 31) = 4.16$, $p = .050$, $\eta^2_p = .12$), pointing to different amounts of improvement in the verbwise and fully unaware participants' explicit knowledge for trained and untrained verbs. The results of the follow-up analyses for the verbwise participants only must be interpreted with caution, given the small sample size. Nonetheless, the ANOVA for the trained verbs showed large and significant improvements in the verbwise participants' production over time ($F(1,7) = 7.18$, $p = .032$, $\eta^2_p = .51$) whereas the ANOVA for the untrained verbs did not (Time: $F(1,7) = 0.70$, $p = .430$, $\eta^2_p = .09$). Averaging across $a \rightarrow \ddot{a}$ and $e \rightarrow i$ verbs, the gain scores were 1.13 ($SD = 1.19$) for trained verbs versus 0.25 ($SD = 0.85$) for untrained verbs, as compared to 0.24 ($SD = 0.72$) and 0.21 ($SD = 0.58$) for the fully unaware experimental participants. The controls' improvement was similar to that of the fully unaware participants, with 0.10 ($SD = 0.87$) and 0.42 ($SD = 0.87$) point gain scores on the trained and untrained verbs, respectively.

DISCUSSION

The current study supported previous findings of implicit L2 learning, by showing both real-time and short-term changes in unaware L2 learners' performance during and after input flooding. Although the unaware learners did not report hearing violations of the stem-change rule, they slowed down significantly when presented with nonchanging strong verbs in the penultimate block of the sentence-picture matching task. Because these same participants had shown the opposite pattern on the word-monitoring pretest, their grammatical sensitivity during input flooding likely reflected the effects of implicit instruction, that is, in their case, implicit learning. In support of this claim, part of the experimental group's sensitivity to ungrammatical strong verbs carried over to the word-monitoring posttest, thereby undoing the participants' original, nontargetlike preferences for "regularized" strong verbs. No evidence for transfer of implicit knowledge to an explicit knowledge test was found, except for a subgroup of learners who became aware of the special role of verbs in the study, but not of the change in their grammaticality. Results for untrained verbs followed those of trained verbs and provided evidence for implicit system learning, but not explicit system learning. Comparable amounts of learning and knowledge gains were observed for the two verb types.

Real-Time and Short-Term Effects of Implicit Instruction

This study replicated the findings from Leung and Williams (2011, 2012, 2014) that adults can show unconscious sensitivity to deviating

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

grammatical patterns provided they receive ample targetlike input. The L2 German learners in this study received an intensive, meaning-focused treatment seeded with strong verbs (172 tokens representing 12 different strong-verb types). The input flood focused on the two most frequently occurring persons in instructed SLA—*I* and *he/she*—which covary with the two verb stems—unmarked and marked—that distinguish strong verbs from weak verbs in the present tense in German. The goal of including both strong-verb and weak-verb pairs was to boost the number of verb exemplars stored in learners' memories and thereby facilitate comparisons, whether conscious or unconscious, between the two types of verb stem.

As became apparent throughout the study, verb allomorphs represent a complex learning problem that can be broken down into three levels. At the lowest level, there is the perceptual challenge that comes from the stem variants' low salience (Godfroid & Uggen, 2013). Indeed, a precondition for learning is that subjects consciously perceive, or notice (Schmidt, 1990), the difference between the allomorphs; overshadowing (Ellis, 2006) of the marked stem by the unmarked stem would compromise learning. The data from the word-monitoring pretest suggest that at least the experimental participants perceived the difference between the two verb forms. They responded more slowly to (grammatical) marked stems than (ungrammatical) unmarked stems. Although this is a reversal of the expected pattern whereby ungrammatical forms elicit longer RTs than grammatical forms, it does show that the participants distinguished between the two verbs.[6] Furthermore, this observation is consistent with Godfroid and Uggen's (2013) findings of noticing in written language, although, unlike in the present study, the beginning learners in Godfroid and Uggen may have benefited from seeing both verb forms presented together on the screen.

At a second, intermediate level, verb allomorphs are also prime candidates for form-form learning given that the focal elements in the target structure—the stem and its vowel—are concrete and contiguous (DeKeyser, 1995, 2003). DeKeyser (1995) hypothesized in a study on L2 morphology that "choosing the right allomorph for a verb suffix would be a likely candidate for item learning, because it is highly predictable from the concrete form of its immediate environment" (p. 384). The present study found evidence for item learning in both the input flood and the word-monitoring task. The experimental participants showed a preference for marked over unmarked verb stems toward the end of the input flood and, to a smaller degree, after that on the word-monitoring posttest. An interesting finding was that the effects of instruction extended beyond the verbs included in the input flood. The experimental group's GSIs underwent comparable shifts from pre- to posttest for trained and untrained verbs, as indicated by a lack of significant or practically meaningful interactions of Time and Group with the factor

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

Verb Training. Thus, gains in implicit knowledge observed on the word-monitoring task generalized to the wider strong-verb system. This result aligns well with DeKeyser's (1995) descriptive data that supported using implicit-inductive, rather than explicit-deductive, approaches for learning similarity-based (noncategorical) grammar patterns, such as allomorphy.

The acquisition of stem allomorphs does not end with form-form mapping, however. It also includes a form-function component. Krause et al. (2015) noted that verb stem allomorphs in German encode person, number, tense and mood information. This is illustrated by the existence of minimal pairs such as *spricht* ("talks$_{\text{Ind.Pres.3sg}}$") and *sprecht* ("talk$_{\text{Ind.Pres.2pl}}$" or "talk$_{\text{Imp.Pl}}$"), as well as *fährt* ("drives$_{\text{Ind.Pres.3sg}}$") and *fahrt* ("drive$_{\text{Ind.Pres.2pl}}$" or "drive$_{\text{Imp.Pl}}$"). Acquisition of the form-meaning mapping underlying these exemplars is difficult because the subject—respectively, *er/sie* ("he/she"), *ihr* ("you$_{2\text{pl}}$"), or ∅ for the imperative—conveys person, number, and, for the zero pronoun, mood information in a more reliable manner than the verb stem does. It may not be surprising, then, that the evidence for form-function learning in this study was limited. In this section, we focus on the word-monitoring data, while the results for controlled oral production will be considered in the next section.

The experimental group's performance on the word-monitoring pretest suggests that these L2 learners treated all verbs as weak or regular verbs in their interlanguage; that is, the participants in the experimental group did not like vowel changes, even in verbs that required them, and so when they encountered a marked verb stem, their monitoring responses were delayed. Table 6 indicated that the experimental participants' non-targetlike preference was undone, but not reversed, after input flooding. At posttest, response latencies were comparable for the grammatical and ungrammatical conditions. A possible interpretation of this finding is that the input flood had increased the activation levels of the marked verbs, perhaps to levels comparable with the control group's, but that all L2 learners were yet to establish the correct form-function mappings for each stem variant. On this account, input flood promoted experimental participants' form learning but not the induction of the distributional rules that govern the forms (i.e., unmarked stems with *I* and marked stems with *he* or *she*). This interpretation is consistent with Williams's (1999) conclusion that "knowledge of distributional rules does not simply emerge out of memory encodings of the relevant forms but depends upon appropriate allocation of attention over relationships between input elements at the time of encoding" (p. 1).

A promising approach to manipulate attention more effectively would be to modify the sentence-picture matching task so it forces participants to engage in task-essential practice (e.g., Loschky & Bley-Vroman, 1993; Sanz & Morgan-Short, 2004). Participants in the current

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

study processed the spoken sentences for meaning in order to select the correct picture; however, performance accuracy did not depend on whether participants processed the function of the stem alternation as opposed to the general semantics of the verb. By pairing pictures so they represent the same action, but show a different number of agents in each picture (*he/she* versus you_{2pl}), participants could be forced to process the number information that distinguishes minimal pairs such as *spricht* and *sprecht* "talk(s)." Ensuring the task-essentialness of the target structure would also make the research more comparable with Leung and Williams (2011, 2012, 2014), even though the concurrent measure of learning (i.e., RT increases) yielded similar findings in Leung and Williams's research and the present study despite the conceptual differences.

Verbwise Learners and the Controlled Oral Production Task

Implicit form learning and implicit knowledge development as measured during input flood and word monitoring, respectively, did not transfer to the controlled oral production task. On that task, the experimental and NNS control participants showed comparable, small gains in explicit, declarative knowledge, which could be accounted for by a test-retest effect. However, a subset of eight verbwise participants in the experimental group deviated from this general pattern and made substantial improvements on the verbs encountered in the sentence-picture matching task, but not on untrained strong verbs. Thus, item learning was contingent on learners' subjective experience of the experiment and system learning was absent.

The verbwise participants formed an interesting subgroup of learners who shared some characteristics with aware learners and most of their characteristics with fully unaware learners. To contextualize the analyses reported in this study, some of these characteristics will be described here. Despite the differences in awareness level, all three groups showed roughly comparable levels of sensitivity during input flood (fully unaware: 73 msec [*SD* = 101], verbwise: 24 msec [*SD* = 133], aware: 55 msec [*SD* = 239]). However, the verbwise and aware learners distinguished themselves by responding more than 1 *SD* faster than the fully unaware participants: an average of -115 msec (*SD* = 102) for the verbwise participants and -195 msec (*SD* = 166) for the aware participants. These response time differences did not fluctuate much as a function of grammaticality. Perhaps the differences in response speed could be accounted for by the verbwise and aware participants' higher cognitive aptitude, but a less speculative explanation is that these two learner groups used their insight into the experiment—namely, that it focused

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

on verbs—strategically to speed up their responses. Thus, by listening carefully to the verbs, both groups were able to select pictures more quickly than their fully unaware peers. For analysis purposes, however, it was crucial that verbwise participants did not report hearing ungrammatical verbs. Therefore, it seemed unlikely that their metacognitive knowledge about the task would affect grammatical and ungrammatical verbs differently; hence, there was no reason to assume it would affect the *difference* in RTs. Because RT differences were the focus in the input flood and word-monitoring analyses, it was decided to treat fully unaware and verbwise participants as a single group in these analyses.

Of course, the same argument did not apply to the controlled oral production task, which dealt with grammatical verbs only. A post hoc analysis for the different subgroups revealed that participants' subjective experience mattered for oral production: the verbwise participants, but not the fully unaware group, demonstrated item learning. This suggests that instruction impacted these two groups' knowledge representations differently. In Sharwood Smith's (1991) terms, the "externally induced salience" (p. 118) of the strong-verb flood brought about "internal salience" (p. 121) in the verbwise group only. This notion of internal salience—the verbs standing out to the participant—could be the result of an interaction between implicit learning, implicit knowledge, and explicit knowledge. Ellis (1996) and Williams (1999) both described the structuring effects of implicit knowledge on conscious perception (for variants of this argument, see Robinson, 1995, and Gass, 1997). The central idea is that even when L2 processing is data driven, unconscious, and inductive, the effects of processing on long-term memory may still be conscious in that memory effects may cause the learner to perceive subsequent input in a structured way. In the context of this study this could mean that the verbwise participants accumulated many instances of verbs unconsciously in memory, up to a point where this caused them to perceive the importance of verbs as structuring elements in the input. Therefore, their conscious experience of the experiment changed as a result of the implicit learning that had taken place.

Although similar processes may have been at work for the five aware learners in the experimental group, prior knowledge likely played a more important role for this subgroup. Of the 12 trained verbs, the five aware learners produced a mean of 10.20 verbs (SD = 1.30, range 9–12) correctly at pretest, compared to 6.25 verbs (SD = 2.38, range 3–9) by verbwise participants and 6.88 verbs (SD = 2.35, range 3–12) in the fully unaware group. Thus, the distribution of pretest scores between aware learners, on the one hand, and fully unaware and verbwise participants, on the other, showed virtually no overlap, even though this could not be confirmed statistically due to the small number of aware participants.

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

Challenges and Opportunities of Natural Language Research

Hulstijn and De Graaff (1994) and De Graaff (1997) promoted a "twin approach" to SLA research that combines "a more easily controllable artificial language learning study with a more valid and therefore more easily generalizable natural language learning study" (De Graaff, 1997, p. 255). This study could be seen as a natural language twin of Leung and Williams's experiments, provided one keep in mind a few differences (also see section "Real-Time and Short-Term Effects of Implicit Instruction"). In order to implement Leung and Williams's concurrent measure of learning (i.e., the RT methodology), it was necessary to recruit third- and fourth-year students of German, whose listening skills were sufficiently automatized to process spoken language rapidly (see "Debriefing Interview" in "Method" section). This only became clear when piloting the procedure. The change in population meant the new participants already possessed explicit, declarative knowledge of the strong-verb paradigm, unlike the German 101 students in Godfroid and Uggen (2013) and the participants in Leung and Williams (2011, 2012, 2014), who were exposed to novel form-meaning associations. To mitigate the effects of prior knowledge in this study, the main analyses were performed on those participants who did not report hearing violations of their conscious rule knowledge of strong verbs during the experiment. Furthermore, two pre- and posttests were added to trace participants' knowledge development relative to their baseline levels.

Although participants' prior knowledge made the current study design more complex, it also highlighted the fact that in the case of a natural language, implicit learning does not happen in a vacuum. In semi-artificial language research, the targets for learning are typically inserted in sentences in the participants' L1, such that the learners have a perfect command of the target language, except for the one structure under investigation. This is a simplified and idealized representation of the L2 learning process that bypasses any issues of developmental readiness and general language processing skills, such as L2 reading or listening proficiency. Learners must have adequate processing skills to be able to learn grammar implicitly. This will be especially important when learning is to take place from auditory input, which is arguably the main source of implicit language acquisition (Rebuschat & Williams, 2012), because the demands on working memory are larger during listening than reading (Hummel & French, 2010). As illustrated by the present study, for instructed learners to have good listening skills, this will often mean they have reached a point in their curriculum where many grammar topics have already been introduced in their classes. The present study, then, relinquished some experimental control, but in the process, uncovered interesting interactions between explicit and implicit knowledge in

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

the verbwise and aware learners that highlight the dynamic nature of L2 acquisition.

Measuring Implicit and Explicit Knowledge

The operationalization and measurement of implicit and explicit knowledge is an active area of research with direct relevance for the current study. In the present project, the word-monitoring task—a hypothetized measure of implicit knowledge (Granena, 2013; Suzuki & DeKeyser, 2015)—was implemented into a pretest-posttest design, with good results. Word-monitoring RTs captured gains in grammatical sensitivity, which is considered a hallmark of implicit knowledge (Granena, 2013; Jiang, 2007; Lim & Christianson, 2015), that occurred following an implicit instructional treatment. Although no direct relationship between sensitivity during input flood and sensitivity during word monitoring could be established, the findings support ongoing efforts to validate the test. In this regard, it is worth mentioning that the word-monitoring task in this study could not be used with the NS control group, who only completed the sentence-picture matching task. Although the sentences (mean length: 10.80 words, *SD* = 1.59) were of an appropriate difficulty for the intermediate-level L2 learners, they were too easy to impose dual-task conditions on the NSs. Specifically, all the German NSs who piloted the procedure reported hearing the ungrammatical verbs, which invalidated the task as a measure of implicit knowledge for this population. Therefore, researchers who intend to give beginning or intermediate-level L2 learners a word-monitoring task must plan on including a NNS, rather than a NS, control group.

The proposed measure of explicit knowledge in this study was a controlled oral production test. Although explicit knowledge measures tend to be more straightforward to develop than implicit knowledge measures (e.g., Goo, Granena, Yilmaz, & Novella, 2015), the explicit measures carry a risk of priming participants' subsequent performance and could contaminate implicit instruction and implicit knowledge measures, such as the ones implemented here. For this reason, I felt it was safer not to opt for a more traditional explicit knowledge measure such as an untimed grammaticality judgment or cloze test. Although the oral production test was constrained, untimed, and had low communicative value, it is acknowledged that it merely invited participants to use their explicit knowledge and did not force them to do so. As always, more data from more tests administered multiple times could enhance our understanding of the learning process.

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

CONCLUSION

This study combined research methodologies from (semi-)artificial language research and work on ultimate attainment to study the effects of implicit instruction on implicit and explicit knowledge development. Using Leung and Williams's (2011, 2012, 2014) RT methodology, I replicated the authors' findings of significant slowdowns in RT, which are an index of grammatical sensitivity, when L2 learners are exposed to ungrammatical forms during input flooding. Because no such sensitivity was attested on the word-monitoring pretest, the RT slowdowns were taken as evidence for implicit learning during the input flood.

Implicit learning coincided with implicit knowledge development on the word-monitoring task, which generalized to untrained strong verbs and was likely form based. This supported the idea that the acquisition of fuzzy grammatical categories such as strong verbs does not amount to rote memorization, like vocabulary learning, but involves the creation of associative networks (Bybee, 1995; Bybee & Newman, 1995; Ellis, 2002; Ullman, 2001) or prototypes (DeKeyser, 1995) of similarly behaving forms. Transfer of knowledge to an explicit knowledge test was limited to item learning for a subset of participants for whom verbs acquired special salience in the course of the experiment. Thus, implicit instruction affected implicit knowledge primarily. To explore interactions between implicit processing, implicit knowledge, and explicit knowledge, future research should focus on the roles of prior knowledge and memory to corroborate the suggestive evidence for interactions that was gleaned from this study's aware and verbwise learners.

Received 4 April 2015
Accepted 8 September 2015
Final Version Received 18 September 2015

NOTES

1. The salience or perceptibility of a stimulus is widely believed to be beneficial to language learning (e.g., Collins, Trofimovich, White, Cardoso, & Horst, 2009; DeKeyser, 2005; Ellis, 2006; Goldschneider & DeKeyser, 2001). Why then is increased salience presented as a negative factor here, whereas it was previously argued to promote language learning in input flooding? The argument is that targets in semi-artificial language studies owe their salience, at least in part, to being the odd one out in a familiar language environment, not unlike a single green dot among a field of red dots. Such salience, it is argued, is extraneous to the natural language learning process. I thank an anonymous reviewer for allowing me to clarify this issue.

2. Following standard practice in visual priming experiments, I present the prime in lowercase and the target in uppercase letters. However, because this was a crossmodal experiment, the prime was presented auditorily and the target appeared in lowercase.

3. As Leung and Williams (2012) suggested, a control-post block was added to their original design (which ended with the violation block) "to show that any difference

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

between the first control and the violation block is not due to general task practice effects, or fatigue" (p. 654).

4. It is acknowledged that these two tasks are not each other's implicit and explicit counterparts, because the word-monitoring task is a receptive measure and controlled oral production is not.

5. In the case of control participants, this meant they became aware of the ungrammatical verbs embedded in the word-monitoring task.

6. In contrast to the experimental participants, the control group had no preference for either marked or unmarked verb stems at pretest. This finding is difficult to interpret, because it could suggest that the control participants did not hear the difference or it could indicate that they were somewhat more advanced than the experimental group. For more information on the latter possibility, see the discussion of the experimental group's posttest data in the current section.

REFERENCES

Andringa, S., & Ćurčić, M. (2015). How explicit knowledge affects online L2 processing: Evidence from differential object marking acquisition. *Studies in Second Language Acquisition*, *37*, 237–268.

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

Berry, D., & Dienes, Z. (1993). *Implicit learning: Theoretical and empirical issues*. Hillsdale, NJ: Erlbaum.

Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition*, *16*, 157–168.

Bickel, B., & Nichols, J. (2007). Inflectional morphology. In T. Shopen (Ed.), *Language typology and syntactic description. Volume III: Grammatical categories and the lexicon*, 2nd ed. (pp. 169–240). Cambridge, UK: Cambridge University Press.

Bowles, M. (2010). *The think-aloud controversy in second language research*. New York, NY: Routledge.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*, 425–455.

Bybee, J., & Newman, J. (1995). Are stem changes as natural as affixes? *Linguistics*, *33*, 633–654.

Chun, M., Golomb, J., & Turk-Browne, N. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*, 73–101.

Collins, L., Trofimovich, P., White, J., Cardoso, W. & Horst, M. (2009). Some input on the easy/difficult grammar question: An empirical study. *The Modern Language Journal*, *93*, 336–353.

De Graaff, R. (1997). The *Experanto* experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition*, *19*, 249–276.

De Jong, N. (2005). Can second language grammar be learned through listening? An experimental study. *Studies in Second Language Acquisition*, *27*, 205–234.

DeKeyser, R. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Studies*, *17*, 379–410.

DeKeyser, R. (2003). Implicit and explicit learning. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 313–348). Oxford, UK: Blackwell.

DeKeyser, R. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning*, *55*, 1–25.

DeKeyser, R. (2009). Cognitive-psychological processes in second language learninga. In M. Long & C. Doughty (Eds.), *Handbook of second language teaching* (pp. 119–138). Oxford, UK: Blackwell.

DeKeyser, R. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 94–112). New York, NY: Routledge.

DeKeyser, R., & Juffs, A. (2005). Cognitive considerations in L2 learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 437–454). Mahwah, NJ: Lawrence Erlbaum Associates.

Dienes, Z., & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin & Review*, *4*, 3–23.

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences*, *22*, 735–808.

Digitales Wörterbuch der deutschen Sprache [Digital dictionary of the German language] [Language corpus]. (2003). Retrieved June 4, 2013 from www.dwds.de.

Doughty, C. (2003). Instructed SLA: constraints, compensation and enhancement. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 256–310). Oxford, UK: Blackwell.

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, *18*, 91–126.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*, 143–188.

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, *27*, 164–194.

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

Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3–25). Bristol, UK: Multilingual Matters.

Ericsson, K., & Simon, H. (1980). Verbal reports as data. *Psychological Review*, *87*, 215–251.

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics*, *27*, 464–491.

Ettlinger, M., Morgan-Short, K., Faretta-Stutenberg, M., & Wong, P. C. M. (2015). The relationship between artificial and second language learning. *Cognitive Science*. Advance online publication. doi: 10.1111/cogs.12257.

Faretta-Stutenberg, M., & Morgan-Short, K. (2011). Learning without awareness reconsidered: A replication of Williams (2005). In G. Granena, J. Koeth, S. Lee-Ellis, A. Lukyanchenko, G. Prieto Botana, & E. Rhoades (Eds.), *Selected proceedings of the 2010 Second Language Research Forum: Reconsidering SLA research, dimensions, and directions* (pp. 18–28). Somerville, MA: Cascadilla Proceedings Project.

Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences, U.S.A.*, *99*, 529–534.

Gass, S. (1997). *Input, interaction and the second language learner*. Mahwah, NJ: Lawrence Erlbaum Associates.

Godfroid, A., & Spino, L. (2015). Reconceptualizing reactivity of think-alouds and eye tracking: Absence of evidence is not evidence of absence. *Language Learning*. Advance online publication. doi: 10.1111/lang.12136.

Godfroid, A., & Spino, L. (in press). Under the radar: Triangulating think-alouds and finger tracking to detect the unnoticed. In A. Mackey & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS repository of instruments for research into second languages*. New York: Routledge.

Godfroid, A., & Uggen, M. S. (2013). Attention to irrregular verbs by beginning learners of German: An eye-movement study. *Studies in Second Language Acquisition*, *35*, 291–322.

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye tracking. *Studies in Second Language Acquisition*, *35*, 483–517.

Godfroid, A., Loewen, S., Jung, S., Park, J., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality judgments measure distinct types of knowledge. *Studies in Second Language Acquisition*, *37*, 269–297.

Goldschneider, J., & 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., 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. 443–482). Amsterdam: John Benjamins.

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

Granena, G. (2013). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning*, *63*, 1–39.
Grey, S., Williams, J. N., & Rebuschat, P. (2014). Incidental exposure and L3 learning of morphosyntax. *Studies in Second Language Acquisition*, *36*, 611–645.
Hama, M., & Leow, R. P. (2010). Learning without awareness revisited: Extending Williams (2005). *Studies in Second Language Acquisition*, *32*, 465–491.
Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning: Introduction. *Studies in Second Language Acquisition*, *27*, 129–140.
Hulstijn, J. H., & De Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review*, *11*, 97–113.
Hulstijn, J. H., & DeKeyser, R. (Eds.) (1997). *Testing SLA theory in the research laboratory. Studies in Second Language Acquisition*, *19*.
Hummel, K. M., & French, L. M. (2010). Phonological memory and implications for the second language classroom. *The Canadian Modern Language Review*, *66*, 371–391.
Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506.
Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, *57*, 1–33.
Knudsen, E. (2007). Fundamental components of attention. *Annual Review of Neuroscience*, *30*, 57–78.
Krashen, S. (1985). *The input hypothesis: Issues and implications.* London: Longman.
Krause, H., Bosch, S., & Clahsen, H. (2015). Morphosyntax in the bilingual mental lexicon: An experimental study of strong stems in German. *Studies in Second Language Acquisition* Advance online publication. doi:10.1017/S0272263114000564.
Lardiere, D. (2006). *Ultimate attainment in second language acquisition: A case study.* Mahwah, NJ: Erlbaum.
Larsen-Freeman, D. (2010). Not so fast: A discussion of L2 morpheme processing and acquisition. *Language Learning*, *60*, 221–230.
Leow, R. P. (1997). Attention, awareness, and foreign language behavior. *Language Learning*, *47*, 467–505.
Leow, R. P. (2000). A study of the role of awareness in foreign language behavior: Aware versus unaware learners. *Studies in Second Language Acquisition*, *22*, 557–584.
Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach.* New York, NY: Routledge.
Leow, R. P., & Hama, M. (2013). Implicit learning in SLA and the issue of internal validity. *Studies in Second Language Acquisition*, *35*, 545–557.
Leow, R. P., Johnson, E., & Zárate-Sández, G. (2010). Getting a grip on the slippery construct of awareness: Toward a finer-grained methodological perspective. In C. Sanz & R. P. Leow (Eds.), *Implicit and explicit conditions, processes and knowledge in SLA and bilingualism* (pp. 61–72). Washington, DC: Georgetown University Press.
Leung, J. H. C., & Williams, J. N. (2011). The implicit learning of mappings between forms and contextually derived meanings. *Studies in Second Language Acquisition*, *33*, 33–55.
Leung, J. H. C., & Williams, J. N. (2012). Constraints on implicit learning of grammatical form-meaning connections. *Language Learning*, *62*, 634–662.
Leung, J. H. C., & Williams, J. N. (2014). Crosslinguistic differences in implicit language learning. *Studies in Second Language Acquisition*, *36*, 733–755.
Lim, J. H., & Christianson, K. (2015). Second language sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics*, *36*, 1283–1315.
Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based learning. In G. Crookes & S. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 123–167). Clevedon, UK: Multilingual Matters.
Marslen-Wilson, W., & Tyler, L. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*, 1–71.
Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience*, *24*, 933–947.

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K. A., Carpenter, H., & Wong, P. C. M. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition*, *17*, 56–72.
Murphy, V. A. (2004). Dissociable systems in second language inflectional morphology. *Studies in Second Language Acquisition*, *26*, 433–459.
Norman, E., Price, M. C., Duff, S. C., & Mentzoni, R. A. (2007). Gradations of awareness in a modified sequence learning task. *Consciousness and Cognition*, *16*, 809–837.
Norris, J., & Ortega, L. (2001). Does type of instruction make a difference? Substantive findings from a meta-analytic review. *Language Learning*, *51*, 157–213.
Nübling, D., Dammel, A., Duke, J., & Szczepaniak, R. (2006). *Historische Sprachwissenschaft des Deutschen: Eine Einführung in die Prinzipien des Sprachwandels* [Historical grammar of German: An introduction to the principles of language change]. Tübingen, Germany: Gunter Narr Verlag.
Opitz, B., & Friederici, A. D. (2003). Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. *NeuroImage*, *19*, 1730–1737.
Opitz, B., & Friederici, A. D. (2004). Brain correlates of language learning: The neuronal dissociation of rule-based versus similarity-based learning. *The Journal of Neuroscience*, *24*, 8436–8440.
Paradis, M. (2009). *Declarative and procedural determinants of second languages.* Amsterdam: John Benjamins.
Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, *16*, 103–133.
Reber, A. S., Allen, R., & Reber, P. J. (1999). Implicit versus explicit learning. In R. J. Sternberg (Ed.), *The nature of cognition* (pp. 475–513) Cambridge, MA: MIT Press.
Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, *63*, 595–626.
Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics*, *33*, 829–956.
Rebuschat, P., Hamrick, P., Sachs, R., Riestenberg, K., & Ziegler, N. (2015). Triangulating measures of awareness: A contribution to the debate on learning without awareness. *Studies in Second Language Acquisition*, *37*, 299–334.
Robinson, P. (1995). Attention, memory and the "noticing" hypothesis. *Language Learning*, *45*, 283–331.
Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition*, *19*, 223–247.
Robinson, P. (2010). Implicit artificial grammar, and incidental natural second language learning: How comparable are they? In M. Gullberg & P. Indefrey (Eds.), *The earliest stages of language learning* (pp. 245–263). Chichester, UK: Wiley-Blackwell.
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. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129–158.
Schmidt, R. (1994a). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, *11*, 11–26.
Schmidt, R. (1994b). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 165–209). London: Academic Press.
Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*, 367–447.
Sharwood Smith, M. (1981). Consciousness raising and the second language learner. *Applied Linguistics*, *2*, 159–168.
Sharwood Smith, M. (1991). Speaking to many minds: On the relevance of different types of language information for the L2 learner. *Second Language Research*, *7*, 118–132.
Stadler, M. A., & Frensch, P. A. (1994). Whither learning, whither memory? *Behavioral and brain sciences*, *17*, 423–424.
Suzuki, W., & DeKeyser, R. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge. *Language Learning*. Advance online publication. doi: 10.1111/lang.12138.

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

Tagarelli, K. M., Borges Mota, M. & Rebuschat, P. (2015). Working memory, learning conditions, and the acquisition of L2 syntax. In W. Zhisheng, M. Borges Mota, & A. McNeill (Eds.) *Working memory in second language acquisition and processing: theory, research and commentary* (pp. 224–247). Buffalo, NY: Multilingual Matters.

Tomlin, R., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, *16*, 183–203.

Truscott, J., & Sharwood Smith, M. (2011). Input, intake, and consciousness: The quest for a theoretical foundation. *Studies in Second Language Acquisition*, *33*, 497–528.

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

VanPatten, B. (1994). Evaluating the role of consciousness in second language acquisition: Terms, linguistic features and research methodology. *AILA Review*, *11*, 27–36.

VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research.* Norwood, NJ: Ablex.

Williams, J. N. (1999). Memory, attention, and inductive learning. *Studies in Second Language Acquisition*, *21*, 1–48.

Williams, J. N. (2004). Implicit learning of form-meaning connections. In J. Williams, B. VanPatten, S. Rott, & M. Overstreet (Eds.), *Form-meaning connections in second language acquisition* (pp. 203–218). Mahwah, NJ: Lawrence Erlbaum.

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, *27*, 269–304.

Williams, J. N. (2009). Implicit learning in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 319–353). Bingley, UK: Emerald Press.

Williams, J. N., & Kuribara, C. (2008). Comparing a nativist and emergentist approach to the initial stage of SLA: An investigation of Japanese scrambling. *Lingua*, *118*, 533–553.

Yang, L. R., & Givón, T. (1997). Benefits and drawbacks of controlled laboratory studies of second language acquisition: The Keck second language learning project. *Studies in Second Language Acquisition*, *19*, 173–193.

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263115000388 Published online by Cambridge University Press

APPENDIX

THE TARGET VERBS

Verb Training	$e \rightarrow i$	Translation	Frequency per million	$a \rightarrow ä$	Translation	Frequency per million
Trained	*helfen*	to help	122	*empfangen*	to receive	275
	lesen	to read	149	*fahren*	to drive	351
	nehmen	to take	664	*fallen*	to fall	275
	sehen	to see	1148	*halten*	to hold	562
	sprechen	to speak	450	*lassen*	to let	1045
	sich treffen	to meet	224	*tragen*	to wear	295
	Mean (SD)		460 (395)	Mean (SD)		467 (303)
Untrained	*brechen*	to break	85	*einladen*	to invite	33
	*empfehlen**	to recommend	76	*fangen*	to catch	68
	essen	to eat	65	*gefallen**	to please	417
	geben	to give	1335	*graben*	to dig	9
	*geschehen**	to occur	241	*raten**	to advise	79
	schmelzen	to melt	6	*schlafen*	to sleep	66
	*sterben**	to die	145	*schlagen**	to hit	195
	vergessen	to forget	102	*wachsen*	to grow	87
	werfen	to throw	123	*waschen*	to wash	22
	Mean (SD)		242 (415)	Mean (SD)		108 (128)

Notes. Lemma frequencies are based on the core corpus of the *Digitales Wörterbuch der deutschen Sprache* (Digital Dictionary of the German Language) (2003), consulted in June 2013. Untrained verbs marked with * occurred in the word-monitoring task only and not the oral production task.