Towards ecological validity in research into input-based practice:

Form spotting can be as beneficial as form-meaning practice

Rowena Kasprowicz and Emma Marsden

University of York

PRE-PRINT VERSION

ABSTRACT

This study extends previous input-based grammar instruction research (for reviews, DeKeyser and Prieto Botana 2015; Shintani 2015) by comparing two types of input-based practice, each with the same explicit information, for learning L2 German definite article case-marking cues (*der, den*). Participants (*N*=138, aged nine to 11) received explicit information followed by either task-essential practice in making form-meaning connections (referential activities from Processing Instruction) OR task-essential practice in spotting the form (noticing activities). Both interventions yielded equivalent durable gains across six ecologically valid tests of comprehension and production (written and oral modalities), compared to negligible gains in a Control group. The findings revealed that, following explicit information, input practice requiring noticing of the target feature (as proposed by Svalberg 2012) was equally effective as task-essential form-meaning connection practice, shedding important light on previous claims in the research agenda on task-essential input practice (e.g. Marsden 2006; Marsden and Chen 2011). Responding to calls for ecologically valid effect-of-instruction research (Mitchell 2000; Spada 2015), this classroom study demonstrates the efficacy of grammar practice for young learners within input-poor foreign language classrooms.[1]

---

[1] Note and Acknowledgements

All intervention and test materials are freely available at www.iris-database.org. We are very grateful to the teachers and pupils who took part in this project, and to David O'Reilly and Anna Capinska. The project was funded by an ESRC PhD studentship.

Rowena Kasprowicz completed her PhD in Education at the University of York (under her maiden name, Hanan) and is now a post-doctoral research associate in the Digital Creativity Labs at the University of York.
Emma Marsden is a Senior Lecturer in the Department of Education at the University of York, director of the digital repository of instruments for research into second languages (IRIS) and Associate Editor of *Language Learning.*

INTRODUCTION

Explicit instruction for young language learners

Investigations into the effectiveness of explicit grammar instruction have typically been carried out with adult or teenage learners (noted by Norris and Ortega 2000; Spada and Tomita 2010; DeKeyser and Prieto Botana 2015). One reason for this is the proposal (originally put forth by Lenneberg, 1967) of a critical period in which children are able to acquire language "from mere exposure" (p. 176), i.e. via implicit learning mechanisms (DeKeyser and Larson-Hall 2005). This led to the perception that 'younger is better' in foreign language (FL) learning (Muñoz 2008). However, learning a language implicitly requires extensive exposure to input, which is very often unavailable. For example, most primary schools in England, the context of this study, offer 30 to 60 minutes FL instruction per week with negligible exposure outside school (Cable *et al.* 2012). Limited exposure is also common in other Anglophone contexts: in Australia, an average 60 minutes a week totals approximately 200 hours over seven years of primary schooling (Lo Bianco, 2009); and the majority of US elementary schools follow an exploratory language program offering limited exposure to the language over short periods (e.g. six to nine weeks) or in weekly lessons of less than 60 minutes (Rhodes & Pufahl 2009; Ingold & Wang 2010).

In middle childhood (age seven to 11), a child's L1 is generally highly developed, grammatically complex and children are becoming more cognitively mature, developing a greater propensity for logical thinking and increasing in language analytic ability and metalinguistic awareness, including of grammatical forms (Philp, Mackey and Oliver 2008). Although not yet developed to full adult capacity, this increased cognitive maturity and awareness of language has been associated with an increased capacity for explicit L2 learning (as observed by Tellier and Roehr-Brackin 2013 in their study of children aged 8 to 9 learning Esperanto). Investigations of young learners' developing metalinguistic awareness include: studies documenting children's L1 knowledge

(e.g. Bryant, Devine, Ledward and Nunes 1997; Sealey & Thompson 2009; Bryant, Nunes and Barros 2014); research into the role of young learners' L1 metalinguistic knowledge in L2 acquisition (e.g. Ammar, Lightbown and Spada 2010; Horst, White and Bell 2010); and evaluations of language awareness programmes, e.g. multilingual programmes inspired by Hawkin's (2005) proposal that in time-limited contexts FL education should serve as a 'language apprenticeship' (e.g. Jones, Barnes and Hunt 2005; Barton, Bragg and Serratice 2009). Drawing on these emerging (meta-)analytic skills (perhaps alongside more implicit mechanisms) could lead to more rapid progress in the acquisition of certain grammatical forms (White 2008).

Although research into FL learning for this age group is severely limited, a few studies have demonstrated the effectiveness of explicit instruction (e.g. White, Spada, Lightbown and Ranta 1991; Harley 1998; White 2008; Bouffard and Sarkar 2008; Serrano 2011; Kim et al 2015). Harley (1998) found that games requiring attention to French gender improved learners' (aged seven to eight) discrimination and production of articles for familiar nouns. White *et al.* (1991) found that input enhancement developed learners' (aged ten to 11) accuracy in L2 English questions. Similarly, White (2008) reports on studies with L1 French learners aged 11 to 12 (Studies 1 and 2) and 13 to 14 (White, Muñoz and Collins 2007), which found that metalinguistic, rule-based instruction developed understanding and use of L2 English possessive determiners; a finding replicated in Serrano's (2011) study with L1 Spanish learners aged 11 to 12. Bouffard and Sarkar (2008) demonstrated, through group discussions stimulated by video replays of communicative activities over three months, that learners aged 8 to 9 developed their ability to analyse L2 grammatical errors, showing heightened metalinguistic awareness of links between their L1 English and L2 French. Finally, Kim et al (2015) observed that the morphosyntactic awareness of bilingual (Spanish-English and Chinese-English) learners aged 9 to 10 was enhanced to the greatest extent by explicit focus on morphosyntactic knowledge.

Of relevance to the current context, England's current National Curriculum for English stipulates that from age 6 children are taught to demonstrate understanding of language with

metalinguistic terminology (e.g. *subject, object, tense etc*) and are tested on this at age 11 in the national standardised Spelling, Punctuation and Grammar test (DfE 2013a).

Although the above suggests that explicit instruction can be beneficial for these ages, research has tended to be conducted in immersion classrooms with extensive input (with the exception of White, Muñoz and Collins 2007 which, although in a limited-exposure classroom, was with teenage learners). In immersion contexts, the combination of extensive input with explicit instruction may play an important role, facilitating implicit and explicit tallying of instances within rich input (Schmidt 1990; Ellis 2002) – opportunities that are less available in FL classrooms.

Consequently, the current study sought to determine the effectiveness of explicit instruction for developing young learners' grammar within input-poor FL classrooms, in a context in which the curriculum demands grammar teaching (DfE 2013b). Despite a strong call for research in these contexts in Mitchell's (2000) anniversary *Applied Linguistics* article, very little research has yet been conducted to investigate effective grammar pedagogy in such environments, where, arguably, there is the greatest need for efficiency.

The role of input practice

Difficulty in learning linguistic features has been attributed to: a feature's low communicative value, saliency or frequency; learner characteristics, such as limited attentional resources; or L1 influence (e.g. Goldschneider and DeKeyser 2001; Ellis 2006; MacWhinney 2012; Robinson 2003; VanPatten 2012). Given these difficulties, positive evidence (instances in the input) is often thought to be insufficient and some form of practice is necessary (Leow, 2007). To date, discussion has focussed primarily on the efficacy of input- versus output-based practice (for review, see Shintani 2015) and the potential benefits of providing versus withholding explicit information (EI) and/or feedback (for review, see DeKeyser and Prieto Bontana 2015). The present study, however, is concerned with comparing the effects of different types of input-based practice, of particular relevance to input-poor FL classrooms where there is a need to optimise any exposure to input.

A role has been proposed for manipulating how learners attend to morphosyntax (VanPatten 2012; DeKeyser 2007; DeKeyser and Criado 2012), for example by providing repeated opportunities to interpret input through specific reading or listening activities. The following sections briefly discuss relevant theories and their pedagogical implications in this regard: i) input-processing theory and a pedagogic technique associated with it, Processing Instruction (PI) (VanPatten 2015); and, ii) the Noticing Hypothesis (Schmidt 1990, 2010) and iii) information processing/skill acquisition theory that accounts for a role for EI in learning.

Task-essential form-meaning connection practice

PI is based on the proposal that in order to establish a mental representation useful for learning, learners must process the input (VanPatten 2015), i.e. make "a connection between form and meaning/function" (VanPatten 2012: 269), such as whether a determiner denotes a noun as the subject or object. The practice component of PI provides repeated opportunities for relying on a form to derive meaning/function, i.e. form-meaning connections (FMCs) are task-essential, as defined by Loschky and Bley-Vroman 1993. Critical to this study is that the term processing does not refer solely to the noticing of a given form, but also to intentionally connecting that form with its meaning (VanPatten 2015).

PI, as originally conceived, contained four components: i) pre-practice EI; ii) information about a processing problem that learners often encounter and how to overcome it, iii) referential activities (which render FMCs task-essential); and iv) affective activities (which provide further exemplars but do not make perceiving the feature or its FMCs task-essential). Two strands of research have sought to determine the effectiveness of these components. One strand has investigated the effectiveness of providing EI with the input activities. VanPatten and Oikkenon (1996) and numerous replications (for review see DeKeyser and Prieto Botana 2015) observed that EI provided no additional benefits compared to input activities alone. Marsden and Chen (2011) also showed learning gains following referential activities (component ii) with no provision of EI. These studies suggested that repeated practice in attending to the FMC was accountable for learning gains,

rather than providing EI. It has therefore been argued that task-essential attention to the FMC is the "*necessary and perhaps sufficient* component of PI that leads to FMCs in SLA" (original emphasis, Farley 2004: 238). However, as will be discussed below, although such research has indicated that EI may not be *necessary* for learning gains, other studies have demonstrated that EI can be beneficial (DeKeyser and Prieto Botana 2015).

The second research strand, and its relevance to the current study, is discussed next.

FMC practice versus noticing practice

There are other accounts for the usefulness of repeated, task-essential attention to forms in the input. Schmidt's (1990) Noticing Hypothesis posits that learners need to notice (i.e. with awareness) a feature for that feature to be learned (or, at least, to help it be learned, Schmidt 2010). The crucial distinction between Schmidt's *noticing* and VanPatten's *processing* is that "unlike processing, noticing does not require the linking of form with meaning" (VanPatten 2015: 93) in the moment at which the noticing occurs. An important question, then, relates to the exact nature of task-essential practice that can lead to learning: Is repeated practice in attending to a grammatical form *and* its meaning necessary for learning to occur? Research comparing PI versus production practice, or provision versus absence of EI, does not seek to investigate the extent to which practice in making task-essential FMCs is required. For this, it is necessary to compare input-based practice where attention to FMCs is necessary with practice where it is not.

To date, two published studies have compared PI activities with an alternative input-based instruction to investigate this issue. First, in two nine-week classroom-based experiments with L1 English learners (aged 12-14) of L2 French, Marsden (2006) compared PI to Enriched Input. Enriched Input contained exactly the same EI and target exemplars (French verb inflections for tense, person, number) as the PI treatment, but attention to the FMCs was not task-essential. Tests of listening, reading (comprehension), writing and speaking (production) showed, in experiment 1, greater gains following PI than Enriched Input, maintained after several weeks. In experiment 2 (a different class and school) the PI group outperformed the Enriched Input group on the comprehension measures

and matched them on the production measures (which was argued to be due to frequent explicit grammar production practice in school 2, as even the control group made equivalent gains on production measures). These findings suggested that task-essential practice in making FMCs resulted in learning, whereas presenting numerous exemplars in the input did not reliably lead to learners processing the features "in a way that aided learning" (Marsden 2006: 544).

A second classroom study (Marsden and Chen 2011), with L1 Chinese learners (aged 12) of L2 English, isolated referential activities (where attention to the target FMC is task-essential) from affective activities (akin to enriched input where attention to the FMC is not task-essential). The target feature was the English past tense *–ed* verb inflection. Four conditions were compared: referential plus affective activities; referential-only; affective-only; and a control group. The referential+affective and referential-only groups made equivalent gains on a timed grammaticality judgement test and gap-fill test at post- and delayed post-test. The affective-only and control groups made no improvement. These findings indicated, in accordance with Marsden (2006), that simply exposing learners to a given grammatical form without pushing them to notice or process it (i.e. affective activities) did not benefit learning. This study also found that no gains were made in any condition on tests that exerted communicative pressure, an issue taken up by the current study.

The findings of Marsden (2006) and Marsden and Chen (2011) are in line with other studies (e.g. White, Spada, Lightbown and Ranta 1991; DeKeyser 1995; Harley 1998; Loewen *et al.* 2009; Reinders and Ellis 2009) that observed minimal (if any) effects for input-based instructional techniques such as enriched input. Although enriching the input aims to increase saliency and facilitate noticing (in line with Schmidt's Noticing Hypothesis), it is impossible to guarantee that learners will actually attend to the feature, as argued by Svalberg (2012). Indeed, greater effects have tended to be observed for instruction that orients the learner's attention on the feature (Norris and Ortega 2000; White 2008). Although some laboratory studies have demonstrated that learning without awareness of the target FMC *can* be observed, this has been with adult learners and on arguably less ecologically valid psycholinguistic measures (e.g. Leung and Williams 2011; Marsden,

Altmann and St Claire 2013). In sum, mere exposure in a classroom, with (Marsden 2006) or without (Marsden and Chen 2011) pre-practice EI, has not seemed sufficient for learning, at least with the relatively young learners and on the relatively ecologically valid measures used in those studies.

Thus it seems that some engagement with the target feature may enhance the effectiveness of an input flood (Svalberg, 2012), but it is not known whether task-essential practice in attending to the target form *and* its meaning is necessary, or whether attending to the form only could be equally effective. The latter approach is certainly easier to implement in classrooms via simple activities such as 'underline the form' or 'write a check mark when you see/hear x'. These techniques do not require any manipulation of the input or specially designed activities, reducing the burden on teachers. Furthermore, this kind of practice is more ecologically valid as it already regularly features in FL textbooks and in practice (Marsden 2005; Lanzer and Wardle 2011; Schicker Waltl and Malz 2011). Thus, as well as being of theoretical interest, it is clearly of applied interest to investigate the extent to which creating task-essential FMC activities is necessary, compared to simply asking learners to notice forms.

A role for explicit information (EI)

Some research has found beneficial effects of EI (reviewed by DeKeyser and Prieto Botana 2015). For example, Culman *et al.* (2009), Henry, Culman and VanPatten (2009) and VanPatten and Borst (2012) found that learners receiving EI prior to referential activities began correctly interpreting German definite article case-marking sooner than learners without EI. Farley (2004) found that "opaque" features (e.g. Spanish subjunctive) may make induction of FMCs more difficult, and so EI may help learners to "see the connections … more quickly" (p. 238).

The usefulness of EI may also be mediated by the practice it accompanies, with fewer benefits when FMCs are task-essential in the subsequent practice (reviewed by DeKeyser and Prieto Botana 2015). The perceived *in*effectiveness of providing EI in research to date may be because the correct/incorrect feedback in task-essential FMC practice enabled learners to induce the FMC and gain their own explicit knowledge, as suggested by DeKeyser *et al.* (2002) and demonstrated by

Marsden and Chen (2011). This leaves open the possibility that combining EI with a different kind of input practice may also provide benefits. Only one study to date, by Prieto Botana (2013) (discussed in DeKeyser and Prieto Botana 2015) has isolated EI when accompanying task-essential versus non-task-essential (where attention was oriented to a non-target feature) practice. Of relevance to the current study, two groups of (L1 English) adult learners received non-task-essential practice in which it was necessary to attend to a meaning carried by the sentence initial Spanish object pronoun (plurality; lo/los, la/las), but not its sentence initial position (OVS), which was the actual focus of the study. One of these groups also received, on five occasions, EI containing an explanation and examples of OVS. Of the two groups, only the group with EI made gains on OVS order, as measured on constrained response interpretation and production tests. Similar findings were observed for ser/estar (to be). This suggested a facilitative role for EI when combined with practice that did *not* force induction of the target rule but oriented attention to 'another' function carried by the target feature (however, cf. Williams, 2005 in which similar training, without EI, provided evidence of learning).

Finally, research into task-essential FMC practice has generally provided EI on a single, pre-practice occasion (with a few exceptions: Marsden 2006; Stafford *et al.* 2012). As noted by DeKeyser and Prieto Botana (2015), "this presents problems of ecological validity and makes any positive effects of EI contingent upon participants' memory and degree of attention" (p. 6). Providing repeated access to EI, then, may benefit learning.

The research above provides evidence that, arguably, aligns with an information processing (skill acquisition) account of learning – that awareness at the level of understanding (Leow 2015) can with practice become proceduralised knowledge, and eventually automatised (DeKeyser 2007). It seems possible that skill acquisition theory could account for at least some learning among even relatively young learners, and particularly in input-poor classrooms where opportunities for self-induced noticing are severely limited. Thus, the current study incorporated EI into both types of interventions under investigation, but isolated the type of practice, by comparing task-essential FMC

practice to noticing practice i.e. where learners were required to attend to the presence of a feature, but not its syntactic function.

Research question

The study compared two types of task-essential input-based instruction that differed in how the learners' attention was directed to the target grammatical feature, with both groups receiving identical pre-practice EI. We sought to address the question:

> Following EI, to what extent is task-essential practice in attending to the Form-Meaning Connection (TE-FM) more beneficial, on a range of tests, than task-essential practice in noticing the target Form (TE-F) for young FL learners in a low-input environment?

METHOD

Participants

138 monolingual L1 English speakers, aged nine to 11, were recruited from seven classes across three primary schools in England. As a new National Curriculum for English began in 2013, all participants were receiving explicit instruction about their L1 grammar. The first author taught German to all participants throughout the academic year in which the experiment took place, enabling control over exposure prior to, during and following the experiment. Previously, learners had received approximately two terms (2 x 12 weeks) of weekly 50 minute German lessons and were deemed to be beginners (Norris and Ortega 2000: 454). Learners from four classes were assigned to either the TE-FM (n = 46) or TE-F (n = 41) groups using matched pair randomisation based on their composite score on two written pre-tests. The three remaining classes formed a non-active control group (n = 52). It was not possible to randomly assign participants to the Control within class groups, as cross-contamination would have been unavoidable. Nevertheless, pre-test scores demonstrated baseline equivalence.

Given the experimental nature of this study, a number of ethical considerations were addressed: written consent was received from the class and head teachers; each participant's parent/carer gave informed consent (either "opt-in" or "opt-out"); materials were made available to

the Control group teachers following the study; and most of the study was delivered during regular German lessons, minimizing disruption (for discussion see Hanan 2015).

Target feature

In German, case-marking on determiners is the most reliable cue to grammatical role assignment and non-canonical (e.g. OVS) structures are permitted (Jackson 2007). However, mastery of case marking can be problematic due to over reliance on word order when assigning grammatical roles (e.g. Jackson 2007; Culman *et al.* 2009; VanPatten & Borst 2012). This difficulty can be explained by different theoretical accounts, for example, the First Noun Principle (VanPatten 2012), the Unified Competition Model (MacWhinney 2012) or models of learned attention in the L1 (Ellis 2006). The current study did not seek to test these, but sought to instruct learners to become sensitive to the case-marking system.

The target feature was accusative case-marking on the German definite article for masculine nouns (*den, the*). This was juxtaposed with nominative case-marking (*der, the*, which the learners encountered during the pre-experiment teaching, as shown by the pre-test results). Definite articles for feminine and neuter nouns (*die* and *das*), which mark *both* nominative and accusative case, were also juxtaposed with *der/den* in the final two intervention sessions. (For discussion of the juxtaposition of pairs of FMCs, see Marsden 2006; Marsden and Chen 2011).

Study design

The design is outlined in Table 1. For all materials, see www.iris-dababase.org.

Table 1: Overview of the study

Week	1	2 – 6	7	16
Stage	Pre-test	Intervention (TE-FM or TE-F)	Post-test	Delayed post-test (TE-FM, TE-F groups)

Interventions

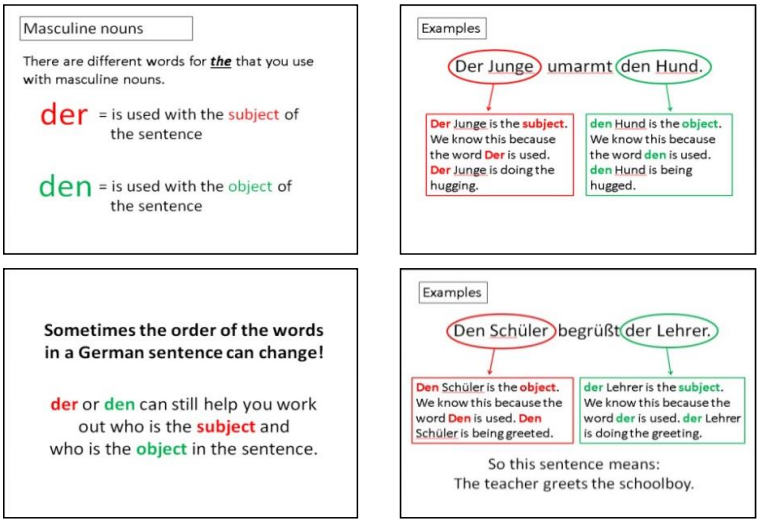
The two interventions (TE-FM, TE-F) were delivered in five weekly 50 minute sessions, totalling 4 hours 10 minutes. The interventions consisted of the same brief EI, followed by listening and reading activities in each condition, delivered via laptops with headphones (2 sessions, 96 items, using

Wondershare QuizCreator package) and via worksheets (3 sessions, 116 items). The total number (212 items) and type (SVO/OVS) of exemplars and delivery mode were identical across conditions. Materials used a list of noun and verb lexemes taught by the researcher prior to the experiment (Suppl. 1). During the experiment, the Control group continued their normal German lessons with the first author, including the vocabulary experienced by the intervention groups, but received no exposure to *den*.

Explicit information

EI was identical for the TE-FM and TE-F groups, provided visually via PowerPoint (Figure 1) and simultaneously read out to both conditions together at the start of each session in weeks 1, 2, 4 and 5.

Figure 1: Sample of EI



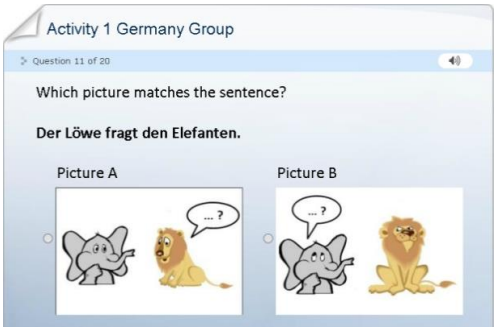
In designing EI appropriate for our young learners, we built on their L1 metalinguistic knowledge by first providing an explanation and examples of the key terms 'subject' and 'object'. Explanations and examples were then provided on the role-assigning functions of the German masculine accusative and nominative articles *den* and *der*. In weeks 4 and 5 it was explained that *der* and *den* can be used to identify the role of a feminine or neuter noun (which can be ambiguous because the same article, *die*/*das* respectively, denotes nominative and accusative).

TE-FM intervention

Akin to PI referential activities (VanPatten and Cadierno 1993; Culman *et al.* 2009; Marsden and Chen 2011), the TE-FM activities forced the learners to repeatedly make FMCs using article case-marking, because word order was not a reliable cue - the first noun was not always the subject (see Figure 2). Numbers of SVO and OVS items were equal.

Figure 2: TE-FM intervention

the-NOM lion asks the-ACC elephant (SVO)



the-ACC teacher greets the-NOM schoolboy (OVS)

In another activity, the learners read or heard a sentence (e.g. *Den Papageien fotografiert der Panda.* The-ACC parrot photographs the-NOM panda) and answered a question (e.g. Who is doing the photographing?).

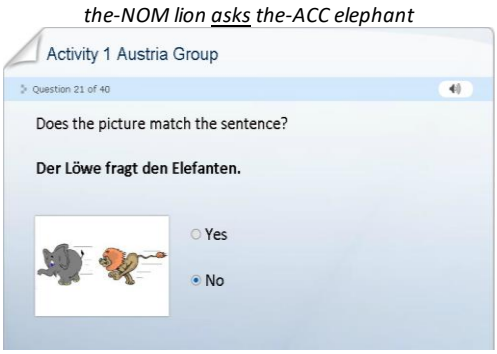
Learners were given 'correct/incorrect' feedback, including the sentence and the correct answer regardless of whether their response was correct. For 92 practice items (out of 212), another brief EI was provided (e.g. about the function of *den*) for both correct and incorrect responses. This contributed to ecological validity as it is unlikely that teachers would generally provide 'correct/incorrect' feedback without additional information/input (Lyster, Saito and Sato 2013).

TE-F intervention

The TE-F activities contained identical stimuli to the TE-FM activities, but made attention to the *form* essential, not the *FMC*. The activities did not intentionally orient attention to the FMC, but required learners to respond to lexical semantics of nouns or verbs, e.g. decide whether a sentence matched the picture (Figure 3) or was sensible/silly, which was never determined by word order or articles (see also DeKeyser 1995; Marsden 2006; White 2008; Reinders and Ellis 2009; Marsden and Chen 2011; Marsden, Altmann and St Claire 2013). Unique to the current study was that a noticing task

then required learners to identify the target form in each sentence: "click [for computer activities] / circle [for paper activities] the words for '*the*'".

Figure 3: TE-F intervention



As in the TE-FM condition, learners were given 'correct/incorrect' feedback, including the sentence and the correct answer regardless of whether their response was correct. For 92 practice items (out of 212), a brief explanation relating to the correct answer (i.e. lexical semantics) was provided, regardless of learners' responses. No further EI about *der* or *den* was provided.

In sum, the two conditions were identical in terms of their pre-practice EI, types and tokens of articles, word orders, and picture stimuli. They differed in that the TE-FM group practised matching articles to their role assigning function and received 92 instances of during-practice EI about *der* and *den*, whereas the TE-F group practised interpreting lexical items and spotting articles and received 92 instances of during-practice EI about lexical semantics.

Tests

Six measures, designed to take in to consideration time, focus of attention, and metalanguage (Ellis 2009; Marsden and Chen 2011), were administered at pre-, post-, and delayed post-test in the following order: Sentence Repetition; Act-Out Comprehension; Act-Out Production; a verbalisable knowledge test; untimed written Sentence Matching and Gap-fill. Three versions of each test were administered in a split block design. SVO/OVS word orders were counter-balanced and animacy was controlled.[2] Aural stimuli were delivered via a laptop and headphones.

[2] Animate+Animate, Animate+Inanimate, Inanimate+Animate appeared equally in SVO and OVS sentences in Act-Out Comprehension and Sentence Matching tests. No differences were found between the animacy

Oral modality tests

Three one-to-one tests exerted some communicative pressure and oriented attention to meaning. Only masculine nouns were used, thus *der* and *den* occurred in each sentence stimulus.

Act-Out Comprehension. Participants listened to a sentence (SVO, *k*=9; OVS, *k*=9) and acted out the meaning using toys (e.g. Chan, Meints, Lieven and Tomasello 2010). One point was awarded per correct response:

Participant hears: *Der Bär umarmt den Tiger* (the-NOM bear hugs the-ACC tiger)

Correct response: Participant makes the bear hug the tiger

Act-Out Production. The researcher acted out transitive sentences (*k*=12) with toys and participants produced German sentences. One point was awarded for each accurate article:

Researcher makes lion kiss monkey

Correct response: *der*-NOM *Löwe küsst den*-ACC *Affen*

Sentence Repetition. Similar to elicited imitation tests (e.g. Harley and Hart 2002; Erlam 2006), this test provided: 1) a meaningful context i.e. action involving two toys; 2) a German sentence describing the action (SVO *k*=3, e.g. *Der Bär umarmt den Elefanten*; OVS, *k=3*); 3) a two second beep; 4) an opportunity for the participant to repeat the sentence. The use of toys and a delay were intended to reduce the likelihood of pure reliance on phonological short term memory[3]. One point was awarded for each accurate article.

Verbalisable Metalinguistic Knowledge. In a one-to-one think-aloud sentence reconstruction task, participants ordered five words (printed on small pieces of paper) to create a sentence to describe a picture (*k*=3; Suppl. 2). The participant was asked "Why did you choose to put the words in that order?" (e.g. Roehr 2008). One point was awarded for correctly explaining the function of each

conditions at pre, post or delayed post-test, indicating that Animacy was not used for role assignment. Given this, we consider it unlikely that animacy influenced our data, perhaps due to our use of cartoon characters and objects. Further discussion is beyond the scope of this paper.

[3] This test was piloted with a comprehension question after each stimulus, but these young, beginner learners were unable to both answer the question and say a sentence.

article. The use of metalinguistic terms (e.g. subject/object) was not required. Correct answers included explanations such as "the dog is being chased so it's den Hund".

Tests in written modality (Suppl. 3)

Sentence Matching. Learners decided which sentence from a pair matched a picture (24 sentence pairs; SVO, *k*=12; OVS, *k*=12). Each sentence contained one masculine and one feminine or neuter noun; the masculine article determined the answer.

Gap-fill. Learners were presented with a picture and a sentence (SVO, *k*=12; OVS, *k*=12) with a masculine noun phrase omitted. Missing subjects/objects were counterbalanced. Participants had to write *der* or *den* and the noun. One point was awarded for each correct article.

Analysis

Data for all tests were non-normally distributed (Shapiro-Wilk[4]), requiring the use of non-parametric tests (Marsden 2006; Marsden and Chen 2011; Norris, Plonsky, Ross and Schoonen 2015).

Comparisons between the TE-FM, TE-F, and Control groups' scores at pre and post-test were made using the Kruskall-Wallis test. If significant, paired comparisons with a Bonferroni correction were carried out (i.e. alpha level divided by 3 for 3 comparisons, .05/3=.0167). Mann Whitney *U*-tests compared the TE-FM and TE-F groups' scores at delayed post-test (the Control group did not take delayed post-tests, both for practical reasons and because their results showed no change between pre- and post-tests, rendering gains at delayed post-test unlikely).

To compare test scores over time, the intervention groups' pre, post-, and delayed post-test scores were analysed using Friedman's ANOVA, and if significant, followed by pairwise comparisons with Bonferroni correction. The Control group's pre and post-test scores were compared using Wilcoxon's signed-rank test.

Effect sizes (Cohen's *d*) were calculated for each comparison between groups and time points (Norris *et al.* 2015).

RESULTS

[4] See Hanan (2015) for full details.

Checking for baseline parity and test effect

No statistically significant differences were found at pre-test between the TE-FM, TE-F, and Control groups' scores across all tests (Tables 2 and 3).

Table 2: Descriptive statistics for comprehension of OVS items, production of den and metalinguistic knowledge

Modality	Task	Group	N	Pre		Post		Delayed	
				M	*SD*	*M*	*SD*	*M*	*SD*
Written	Sentence Matching (OVS) (*k* = 12)	TE-FM	45	0.97	1.41	8.16	4.29	7.60	5.54
		TE-F	38	0.61	0.79	8.53	4.06	7.92	4.68
		Control	50	1.48	2.01	0.92	1.75	-	-
	Gap-fill (SVO) (*k* = 12)	TE-FM	45	0.05	0.21	4.11	2.10	3.24	2.50
		TE-F	38	0.24	0.88	4.03	2.30	3.18	2.52
		Control	50	0.50	1.54	0.62	1.52	-	-
	Gap-fill (OVS) (*k* = 12)	TE-FM	45	0.14	0.55	3.56	2.41	3.22	2.55
		TE-F	38	0.26	1.16	4.03	2.19	3.37	2.50
		Control	50	0.44	1.40	0.40	1.25	-	-
Oral	Act-Out Comprehension (OVS) (*k* = 9)	TE-FM	45	0.50	0.70	3.90	3.70	4.24	3.59
		TE-F	41	0.70	0.90	3.80	3.40	3.71	3.38
		Control	46	0.26	0.61	0.39	1.00	-	-
	Act-Out Production (Object) (*k* = 12)	TE-FM	45	0.00	0.00	6.20	5.81	6.69	5.31
		TE-F	41	0.15	0.69	5.32	5.19	6.27	5.26
		Control	46	0.00	0.00	0.00	0.00	-	-
	Sentence Repetition (SVO) (*k* = 3)	TE-FM	45	0.30	0.60	1.70	1.10	1.69	1.28
		TE-F	41	0.30	0.70	1.40	1.10	1.59	1.14
		Control	46	0.20	0.60	0.10	0.40	-	-
	Sentence Repetition (OVS) (*k* = 3)	TE-FM	45	0.80	1.10	2.10	1.10	2.31	1.12
		TE-F	41	0.40	0.60	1.80	1.20	2.15	1.22
		Control	46	0.80	0.90	0.70	1.00	-	-
Metalinguistic	Sentence Reconstruction (*k* = 6)	TE-FM	45	0.07	0.33	4.78	1.51	3.50	2.20
		TE-F	41	0.00	0.00	4.32	2.02	2.90	1.80
		Control	46	0.00	0.00	0.04	0.21	-	-

Table 3: Results of Kruskall Wallis comparing TE-FM, TE-F and Control groups' scores at pre-test

Task		*H*	*p*
Written	Sentence Matching (OVS)	4.309	0.116
	Gap-fill (SV<u>O</u>)	1.040	0.595
	Gap-fill (<u>O</u>VS)	1.135	0.567
Oral	Act-Out Comprehension (OVS)	5.829	0.054
	Act-Out Production (Object)	4.470	0.107
	Sentence Repetition (SV<u>O</u>)	1.810	0.405
	Sentence Repetition (<u>O</u>VS)	2.870	0.238
Metalinguistic	Sentence Reconstruction	3.900	0.143

The Control group showed no improvement between pre- and post-test on any test (Table 4); the statistical significance observed on the Sentence Matching task was due to a *decrease* in scores (see

Table 2). These findings indicated that any changes over time in the intervention groups' scores were unlikely to be due to a test effect.

Table 4: Results of Wilcoxon signed-rank test comparing Control group's pre- and post-tests

Task		Pre-post		
		T	*p*	*d*
Written	Sentence Matching (OVS)	117.50	0.017	-0.29
	Gap-fill (SV<u>O</u>)	37.50	0.684	0.08
	Gap-fill (<u>O</u>VS)	29.50	0.838	-0.03
Oral	Act-Out Comprehension (OVS)	61.00	0.244	0.16
	Act-Out Production (Object)	0.00	1.000	0.00
	Sentence Repetition (SV<u>O</u>)	23.00	0.353	-0.20
	Sentence Repetition (<u>O</u>VS)	197.00	0.881	-0.11
Metalinguistic	Sentence Reconstruction	3.00	0.157	0.004

Analyses confirmed equivalence of pre-test performance across potentially extraneous variables of age (9-10 vs 10-11 years) and school (1, 2, 3) (see Hanan 2015).

Learning over time

Comprehension of SVO and production of der

All groups (TE-FM, TE-F, Control) performed at, or close to, ceiling in their comprehension of SVO test items and production of *der* at pre-test. Overall there were minimal, non-significant changes over time in these (Suppl. 4 and 5)[5].

Comprehension of OVS and production of den

Participants' comprehension of OVS items and production of *den* revealed a highly consistent pattern of results for both the TE-FM and TE-F groups between pre-, post-, and delayed post-test (Table 2). Friedman's ANOVAs revealed significant differences for the written modality tests (**Sentence Matching OVS**: TE-FM, $X^2(2)=45.436$, $p=.001$; TE-F, $X^2(2)=40.358$, $p=.001$. **Gap-fill SV<u>O</u>**[6]: TE-FM, $X^2(2)=53.737$, $p=.001$, TE-F, $X^2(2)=38.032$, $p=.001$. **Gap-fill <u>O</u>VS**: TE-FM, $X^2(2)=47.091$, $p=.001$; TE-F, $X^2(2)=42.017$, $p=.001$) and oral modality tests (**Act-Out Comprehension OVS**: TE-FM, $X^2(2)=14.504$, $p=.001$; TE-F, $X^2(2)=15.250$, $p=.001$. **Act-Out Production Object**: TE-FM, $X^2(2)=47.328$,

[5] Although three Kruskall Wallis tests suggested some change may have happened, no paired tests indicated statistical change over time in any group or measure. For more details, see descriptive statistics (Suppl. 4) and Hanan (2015).

[6] The underlining indicates which noun phrase was missing, i.e. subject or object (see Suppl. 3)

p=.001; TE-F, X^2(2)=36.845, *p*=.001. **Sentence Repetition SV<u>O</u>**: TE-FM, X^2(2)=48.439, *p*=.001; TE-F, X^2(2)=35.504, *p*=.001. **Sentence Repetition <u>O</u>VS**: TE-FM, X^2(2)=41.392, *p*=.001; TE-F, X^2(2)=48.123, *p*=.001). Pairwise comparisons between time points showed improvement on all measures pre-post and pre-delayed, but no change post-delayed (Table 5), indicating that the substantial gains yielded by both interventions were sustained nine weeks post intervention.

Table 5: Pairwise comparisons (Dunn-Bonferroni) of time points following Friedmans' ANOVA for comprehension of OVS and production of den

Modality	Task	Group	Pre-post		Pre-Delayed		Post-Delayed	
			z	*d*	*z*	*d*	*z*	*d*
Written	Sentence Matching (OVS)	TE-FM	-6.061**	2.52	-4.849**	1.91	1.212	-0.11
		TE-F	-5.277**	3.27	-5.047**	2.67	0.229	-0.14
	Gap-fill (SV<u>O</u>)	TE-FM	-6.430**	3.60	-4.954**	2.35	1.476	-0.38
		TE-F	-5.105**	2.38	-4.531**	1.73	0.574	-0.10
	Gap-fill (<u>O</u>VS)	TE-FM	-5.376**	2.31	-4.743**	1.99	0.632	-0.14
		TE-F	-5.277**	2.25	-4.531**	1.70	0.746	-0.28
Oral	Act-Out Comprehension (OVS)	TE-FM	-2.530*	1.55	-3.162**	1.74	-0.632	0.09
		TE-F	-3.092**	1.44	-2.871**	1.41	0.221	-0.03
	Act-Out Production (Object)	TE-FM	-4.796**	2.13	-5.007**	2.52	-0.211	0.09
		TE-F	-4.086**	1.76	-4.693**	2.06	-0.607	0.18
	Sentence Repetition (SV<u>O</u>)	TE-FM	-5.060**	1.65	-4.902**	1.48	0.158	-0.01
		TE-F	-3.920**	1.22	-4.693**	1.40	-0.773	0.17
	Sentence Repetition (<u>O</u>VS)	TE-FM	-4.164**	1.18	-5.007**	1.36	-0.843	0.19
		TE-F	-4.252**	1.56	-5.522**	1.92	-1.27	0.29

Verbalisable knowledge over time

Friedman's ANOVA revealed change in the TE-FM (X^2(2)=65.790, *p*=.001) and TE-F scores (X^2(2)=59.842, *p*=.001) on the metalinguistic task (Table 2). Pairwise comparisons indicated significant improvement for both groups pre-post (**TE-FM** *z*=-7.537, *p*=.001, *d*=5.12; **TE-F** *z*=-7.012, *p*=.001, *d*=4.28), and pre-delayed (**TE-FM**, *z*=-5.112, *p*=.001, *d*=2.71, **TE-F**, *z*=-4.583, *p*=.001, *d*=3.20). Although a significant decrease was observed post-delayed in both groups (TE-FM *z*=2.429, *p*=.015, *d*=-.69; TE-F *z*=2.429, *p*=.015, *d*=-.74), indicating a decline in verbalisable knowledge, this remained well above pre-test levels, with large effect sizes.

Comparison of the interventions

Across all measures, no differences were found between the TE-FM and TE-F groups at post- or delayed post-test (Table 6).

Table 6: Mann Whitney U-test comparing TE-FM and TE-F scores

Modality	Task	Time	*U*	z	*p*	*d*
Written	Sentence Matching (OVS)	Post	856.5	0.014	0.989	-0.089
		Delayed	924.0	0.640	0.522	-0.062
	Gap-Fill (SVO)	Post	859.5	0.043	0.966	0.037
		Delayed	847.0	-0.075	0.940	0.033
	Gap-Fill (OVS)	Post	941.5	0.813	0.416	-0.204
		Delayed	894.5	0.372	0.710	-0.059
Oral	Act-Out Comprehension (OVS)	Post	897.5	-0.220	0.826	0.028
		Delayed	807.5	-1.011	0.312	0.152
	Act-Out Production (Object)	Post	855.0	-0.601	0.548	0.160
		Delayed	859.0	-0.056	0.573	0.079
	Sentence Repetition (SVO)	Post	746.0	-1.580	0.114	0.273
		Delayed	865.0	-0.517	0.605	0.082
	Sentence Repetition (OVS)	Post	779.0	-1.323	0.186	0.261
		Delayed	849.5	-0.745	0.456	0.095
Metalinguistic	Sentence Reconstruction	Post	817.5	-0.956	0.339	0.262
		Delayed	748.0	-1.527	0.127	0.299

DISCUSSION

Effectiveness of the interventions

The performance of the groups is broadly represented as follows:

Pre-test: TE-FM = TE-F = Control

Post-test: TE-FM = TE-F > Control

Delayed post-test: TE-FM = TE-F

The TE-FM and TE-F scores improved on all measures at post-test. Scores were maintained at delayed post-test (nine weeks later), with the exception of the metalinguistic test scores, which decreased though were still well above baseline.

These results indicated that both the TE-FM and TE-F interventions had large positive effects on the learners' comprehension and production of the feature. Crucially, no gains were observed in the Control group, indicating no test effect.

These gains are in line with previous research demonstrating beneficial effects of form-focussed instruction for young learners albeit in immersion contexts (e.g. Harley 1998; Bouffard and Sarkar 2008; White 2008). In contrast, the present study has provided substantial evidence that even in a low-input environment, like many Anglophone FL classrooms, a relatively short, input-based intervention (five 50 minute sessions) can result in substantial, durable gains in both comprehension and production of morphosyntax. As noted above, FL instruction tends to be limited to one hour per week, focusing on lexicon and formulaic phrases (Cable et al. 2012). Our study suggests that it can be feasible and effective to implement short bursts of grammar instruction within such environments.

Equivalence of the interventions

The TE-FM group's improvement is in line with previous studies, which have consistently demonstrated learning gains following task-essential FMC practice, both when practice is preceded with EI (e.g. VanPatten and Cadierno 1993; Marsden 2006; Agiasophiti 2013; DeKeyser and Prieto Botana 2015) and without EI (e.g. VanPatten and Oikkenon 1996; Marsden and Chen 2011; Stafford *et al.* 2012; DeKeyser and Prieto Botana 2015). Such findings have been interpreted as evidence that practice in attending to a grammatical form *and* its meaning (or function) is necessary for learning (regardless of how the underpinning learning mechanisms are conceived). However, our study has demonstrated that EI followed by practice that oriented attention only on a form (identifying der/den as communicating 'the') resulted in equivalent learning gains to FMC practice. We now consider why the two interventions were similarly effective.

Our study built on previous research comparing task-essential FMC practice to enriched input, both with (Marsden 2006) and without (Marsden 2011) pre-practice EI. Neither study observed gains following enriched input, suggesting that simply exposing learners to a grammatical form did not result in learning (see also Schmidt 1990; DeKeyser 1995; Svalberg 2012). Our data here indicated that incorporating a noticing task into the enriched input (TE-F) (clicking or circling instances of *der*/*den*) was effective in pushing the learners to attend to the form in a way that led to

learning. This corroborates Svalberg's (2012) proposal that engaging with the form may increase the otherwise minimal impact of enriched input. As noted earlier, although input enhancement is intended to increase awareness of the target feature (Sharwood Smith and Truscott 2014), there is no guarantee that learners will notice the feature in a way that promotes learning, even when EI is provided. The noticing task combined with EI in our TE-F intervention oriented learners' attention repeatedly and explicitly to the form, enabling them to benefit from the increased exposure afforded through the activities.

The role of EI must also be considered. As noted, some studies indicated that EI did not mediate the effectiveness of FMC practice (e.g. VanPatten and Oikkenon 1996; Stafford *et al.* 2012), whilst others demonstrated its facilitative role (DeKeyser and Prieto Botana 2015). In our study, the pre-practice EI, albeit brief, likely contributed to the success of the TE-F intervention by focussing the learners' attention on the FMC prior to practice, even though the FMC was not required for the practice itself. Our results are compatible with Prieto Botana's (2013) observation (cited in DeKeyser and Prieto Botana 2015) that EI is necessary for learning with non-task-essential practice.

Although the EI explained the feature's function, several observations suggest that it was the *interaction* of EI with task-essential noticing practice, rather than EI alone, that benefitted the TE-F group. First, the TE-FM group was not advantaged by the extra 92 instances of EI in the feedback. Secondly, Marsden (2006) found that providing EI with enriched input, without noticing practice, did not lead to learning. Thirdly, previous studies found no gains for learners given EI alone (e.g. VanPatten & Oikkenon 1996). Based on these observations, we suggest that the effectiveness of the TE-F intervention was due to the combination of pre-practice EI with noticing practice. For example, it is possible that the TE-F learners rehearsed the FMC information when completing their noticing activities.

Although we cannot conclude that the TE-F learners consistently made a FMC for every item, since FMCs were not task-essential, we speculate this may have happened to some extent,

intentionally or incidentally on the part of the learners[7]. We suggest that the noticing practice, at least, facilitated the tallying (explicitly and/or implicitly) of instances, once the initial FMC had been established as a representation by the pre-practice EI (Schmidt 1990; Ellis 2002). DeKeyser (2007), in his discussion of Skill Acquisition Theory, observed that the "whole sequence of proceduralization and automatization cannot get started if the right conditions for proceduralization are not present (the declarative knowledge required by the task at hand and a task set-up that allows for use of that declarative knowledge)" (p. 100). The TE-F (and TE-FM) instruction may have provided suitable conditions for skill acquisition: with the necessary declarative knowledge, learners were given repeated opportunities to proceduralise (and perhaps automatise) this knowledge.

Measuring learning

The written tests were untimed, constrained response, and sentence-level, allowing learners to reflect consciously on their comprehension and production (Doughty 2003; Ellis 2009). Improvement on these tests could, therefore, reflect explicit knowledge, corroborated by the learners' ability to articulate the target rule, often with appropriate metalanguage, on the metalinguistic post-test (see Hanan 2015). Our data extend previous findings that young learners can be taught explicit analysis of their L2 within an immersion context (e.g. Bouffard and Sarkar 2008), to analysis of a FL after very little exposure. Our findings contribute to the growing evidence (e.g. White, Spada, Lightbown and Ranta 1991; Harley 1998; White 2008; Ammar, Lightbown, and Spada 2010; Horst, White, and Bell 2010; Serrano 2011) that young learners (here age 9-11) are developing metalinguistic awareness, and are therefore able to benefit from explicit instruction (Philp, Mackey and Oliver 2008; Tellier and Roehr-Brackin 2013). Although our learners' ability to explain the function of the feature had declined by delayed post-test (concordant with the notion that explicit knowledge can decay over

[7] In some TE-F activities (e.g. 'sensible/silly') learners may have attended to the syntactic roles of the *nouns* illustrated in the picture, even though the activity did not force attention on subject-object roles. However, linking the *articles* to the syntactic roles in the picture would have been incidental, and is unlikely to explain the effectiveness of the TE-F, given a) the low numbers of these items and b) previous evidence (Marsden 2006; Marsden and Chen 2011) that simple attention to exemplars did not lead to learning. We suggest it was the noticing practice combined with EI that facilitated learning.

time (Ellis, 2009)), the pre-delayed post-test effect sizes were large, compared with other PI studies, Norris & Ortega (2000), and Plonsky & Oswald (2014). This suggests good retention of some verbalisable knowledge. We also note that declines in verbalisable knowledge do not necessarily reflect an equivalent decline in explicit knowledge.

A more contentious question is whether the learners had also developed a different, perhaps more implicit, knowledge type. The oral tests focussed primarily on meaning and exerted communicative pressure, reducing access to explicit knowledge (Ellis 2009). Also, comprehension and production gains were sustained at delayed post-test despite the decreased verbalisable knowledge, possibly indicating increased reliance on more durable, implicit knowledge (Ellis 2009). However, oral performance was not timed and required "constrained, constructed responses" (Norris and Ortega 2000: 440); nor did we document fluency (e.g. pauses, reformulations) or awareness (e.g. confidence ratings, source attributions) (Rebuschat 2013). We cannot therefore conclude that the learners had developed implicit knowledge. Nevertheless, our findings are compatible with the argument that learners had proceduralized some declarative knowledge that was accessible under time and communicative pressure in the oral modality tests, in line with skill acquisition and information processing models (DeKeyser 2007). Given the decline in articulated declarative knowledge, the other delayed post-test measures could have drawn at least partially on different knowledge types (see Hanan 2015 for a principle component analysis supporting this argument), and an explanation involving some automatization of learners' initial explicit knowledge is broadly compatible with our findings. In any case, identifying implicit from automatized explicit knowledge is not necessarily of highest importance for an applied study. For many practitioners, "the idea that with practice, learners are able to proceduralize their explicit knowledge, leading to greater automaticity over time" (Spada 2015: 78) is probably sufficiently informative, particularly for the FL classroom in which extensive input (useful for implicit knowledge development) is rarely feasible.

In sum, our evidence suggests that both practice in noticing forms and FMCs can lead young learners to develop knowledge that is accessible under some time and communicative pressure and after a two-month delay.

Future research

Understanding the extent to which the observed effectiveness can be attributed to the EI and/or practice could require further research isolating EI from practice. However, we are cautious about the utility of such research: First, Marsden (2006) found that slightly older learners did not benefit from pre-practice EI when forms were embedded in the input without instruction to 'notice' them. Secondly, and perhaps most importantly, two of the critical conditions for such a study (EI without practice, or practice without *any* EI - deductive or inductive) are unlikely to be ecologically valid for teachers intending to teach grammar. Such studies would, therefore, not respond to calls for pedagogically relevant applied linguistics research (Mitchell 2000; Spada 2015).

Future research is required to identify the nature of knowledge gains (e.g. implicit) and of the noticing practice (e.g. was this at the level of understanding the FMC or, more simply, at the level of awareness (Leow 2015) or even implicit tallying (Ellis, 2002)). Other areas include whether different (e.g. more complex) FMCs and even younger learners could benefit from TE-FM and TE-F to the same extent. Future research should also determine whether shorter treatments could be equally beneficial, given our relatively lengthy intervention (4 hours 10 minutes). Finally, as noted by Larsen-Freeman (2006: 598) "group averages can conceal a great deal of variability", and further analysis would reveal whether the treatments benefitted all individuals similarly.

CONCLUSION

The improvement made by the TE-FM and TE-F learners, across a battery of tests, suggested that both task-essential FMC practice and noticing practice were equally useful. We demonstrated that explicit, input-based grammar instruction can be effective, supporting Bouffard and Sarkar's (2008) claim that young learners are "mature enough to attend to form if they are taught how to" (p. 22), extending this to the low-exposure FL classroom. Our learners' performance on the metalinguistic

test corroborated the metalinguistic skills observed amongst children aged seven to 11 (Philp *et al.*, 2008), though after nine weeks we found language use was better sustained than metalinguistic knowledge. Finally, as our tests reflected the skills-based priorities (reading, writing, listening, speaking) of many curricula (e.g. DfE 2013b), we hope that this increases the likelihood that our findings strike a chord with those who face 'real-world problems'.

REFERENCES

Agiasophiti, Z. 2013. 'Exploring possible effects of gender and enhanced versus unenhanced Processing Instruction on the acquisition of case marking in L2 german' in Lee, J. F. & A. Benati (eds.): *Individual Differences and Processing Instruction*. Equinox Publishing.

Ammar, A., Lightbown, P. M., and **Spada, N.** 2010. 'Awareness of L1/L2 differences: does it matter?' *Language Awareness* 19/2, 129-146.

Barton, A., Bragg, J., and **Serratice, L.** 2009. ''Discovering language' in primary school: an evaluation of a language awareness programme.' *Langauge Learning Journal* 37/2, 145-164.

Bouffard, L. A., and **Sarkar, M.** 2008. 'Training 8-year-old French immersion students in metalinguistic analysis: an innovation in form-focussed pedagogy.' *Language Awareness* 17/1: 3-24.

Bryant, P., Devine, M., Ledward, A., and **Nunes, T.** 1997. 'Spelling with apostrophes and understanding possession.' *British Journal of Educational Psychology* 67/1: 91-110.

Bryant, P. Nunes, T., and **Barros, R.** 2014. 'The connection between children's knowledge and use of grapho-phonic and morphemic units in written text and their learning at school.' *British Journal of Educational Psychology* 84/2, 211-225.

Cable, C., Driscoll, P., Mitchell, R., Sing, S., Cremin, T., Earl, J., . . . Heins, B. 2012. 'Language learning at key stage 2: findings from a longitudinal study.' *Education 3-13: International Journal of Primary, Elementary and Early Years Education* 40/4: 363-378.

Chan, A., Meints, K., Lieven, E., and **Tomasello, M.** 2010. 'Young children's comprehension of English SVO word order revisited: testing the same children in act-out and intermodal preferential looking tasks.' *Cognitive Development* 25/1: 30-45.

Culman, H., Henry, N., and **VanPatten, B.** 2009. 'The role of explicit information in instructed SLA: an online study with Processing Instruction and German accusative case inflections.' *Die Unterrichtspraxis* 42/1: 19-31.

DeKeyser, R. 1995. 'Learning second language grammar rules: an experiment with a miniature linguistic system.' *Studies in Second Language Acquisition* 17/3: 379-410.

DeKeyser, R. 2007. 'Skill acquisition theory' in VanPatten, B. (ed.), *Theories in Second Language Acquisition: An Introduction*. Lawrence Erlbaum Associates.

DeKeyser, R., and **Larson-Hall, J.** 2005. 'What does the critical period really mean?' in Kroll, J. & de Groot, A. (eds.): *The Handbook of Bilingualism: Psycholinguistic Approaches*. Oxford University Press.

DeKeyser, R., and **Prieto Botana, G.** 2015. 'The effectiveness of Processing Instruction in L2 grammar acquisition: a narrative review.' *Applied Linguistics* 36/3: 290-305.

DeKeyser, R., Salaberry, R., Robinson, P., and **Harrington, M.** 2002. 'What gets processed in Processing Instruction? A commentary on Bill Vanpatten's ''Processing Instruction: an update''.' *Language Learning* 52/4: 805-823.

DfE. 2013a. *National Curriculum in England: English Programmes of Study (Key Stages 1 and 2).* London: Department for Education.

DfE. 2013b. *National Curriculum in England: Languages Programmes of Study (Key Stage 2)* London: Department for Education.

Doughty, C. J. 2003. 'Instructed SLA: constraints, compensation, and enhancement.' in Doughty and Long (eds.).

Doughty, C. J. and **M. H. Long.** 2003. *The Handbook of Second Language Acquisition*. Blackwell Publishing Ltd.

Ellis, N. 2002. 'Frequency effects in language processing.' *Studies in Second Language Acquisition* 24/2: 143-188.

Ellis, N. 2006. 'Selective attention and transfer phenomena in L2 acquisition: contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning.' *Applied Linguistics* 27/2: 164-194.

Ellis, R. 2009. 'Implicit and explicit learning, knowledge and instruction' in Ellis, Loewen, Elder, Erlam, Philp, and Reinders (eds.).

Ellis, R., S. Loewen, C. Elder, R. Erlam, J. Philp, and **H. Reinders.** 2009. *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching*. Multilingual Matters.

Erlam, R. 2006. 'Elicited imitation as a measure of L2 implicit knowledge: an empirical validation study.' *Applied Linguistics* 27/3: 464-491.

Farley, A. P. 2004. 'Processing Instruction and the Spanish subjunctive: is explicit information needed?' in VanPatten (ed.). *Processing Instruction: Theory, Research, and Commentary*. Lawrence Erlbaum Associates, Inc.

Gass, S. M. and A. Mackey. 2012. *The Routledge Handbook of Second Language Acquisition*. Routledge.

Goldschneider, J. M., and **DeKeyser, R.** 2001. 'Explaining the "natural order of L2 morpheme acquisition" in English: a meta-analysis of multiple determinants.' *Language Learning* 51/1: 1-50.

Hanan, R.E. 2015. The effectiveness of explicit grammar instruction for the young foreign language learner: A classroom-based experimental study. Unpublished Ph.D thesis, University of York, UK.

Harley, B. 1998. 'The role of focus-on-form tasks in promoting child L2 acquisition.' in Doughty, C. J. & Williams, J. (eds.): *Focus on Form in Classroom Second Language Acquisition.*: Cambridge University Press.

Harley, B., and **Hart, D.** 2002. 'Age, aptitude and second language learning on a bilingual exchange' in Robinson, P. (ed.), *Individual Differences and Instructed Language Learning*. John Benjamins Publishing Company.

Hawkins, E. 2005. 'Out of this nettle, drop-out, we pluck this flower, opportunity: re-thinking the school foreign language apprenticeship. *Language Learning Journal* 32/1, 4-17.

Henry, N., Culman, H., and **VanPatten, B.** 2009. 'More on the effects of explicit information in instructed SLA: a partial replication and a response to Fernández (2008).' *Studies in Second Language Acquisition* 31/4: 559-575.

Horst, M., White, J., and **Bell, P.** 2010. 'First and second language knowledge in the language classroom.' *International Journal of Bilingualism* 14/3, 331-349.

Ingold, C.W. and **Wang, S.C.** (2010). *The teachers we need: transforming world language education in the United States.* College Park, MD: National Foreign Language Center at the University of Maryland.

Jackson, C. 2007. 'The use and non-use of semantic information, word order and case-markings during comprehension by L2 learners of German.' *The Modern Language Journal* 91/3: 418-432.

Jones, N., Barnes, A., and **Hunt, M.** 2005. 'Thinking through langauges: a multi-lingual approach to primary school languages.' *Language Learning Journal* 32/1, 63-67.

Kim, T.J., Kuo, L-J., Ramírez, G., Wu, S., Ku, Y-M., de Marin, S., Ball, A., and **Eslami, Z.** 2015. 'The relationship between bilingual experience and the development of morphological and morphosyntactic awareness.' *Language Awareness* 24/4, 332-354.

Lanzer, H., and **Wardle, M.** 2009. *Edexcel Gcse German*. Pearson Education Limited.

Lenneberg, E. H. 1967. *Biological Foundations of Language*. John Wiley and Sons.

Leow, R. P. 2007. 'Input in the L2 classroom: an attentional perspective on receptive practice' in DeKeyser, R. (ed.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology*. Cambridge University Press.

Leow, R. P. 2015. *Explicit Learning and the L2 Classroom: A Student Centred Approach*. Routledge.

Leung, J. H. C., and **Williams, J.** 2011. 'The implicit learning of mappings between forms and contextually derived meanings.' *Studies in Second Language Acquisition* 33/1: 33-55.

Lo Bianco, J. (2009). *Australian education review: Second languages and Australian schooling.* Victoria: ACER Press.

Loewen, S., Erlam, R., and **Ellis, R.** 2009. 'The incidental acquisition of third person -s as implicit and explicit knowledge' in Ellis, Loewen, Elder, Erlam, Philp, and Reinders (eds.).

Loschky, L., and **Bley-Vroman, R.** 1993. 'Grammar and task-based methodology' in Crookes, G. & Gass, S. (eds.): *Tasks and Language Learning*. Multilingual Matters.

Lyster, R., Saito, R. and **Sato, M.** 2013. 'Oral corrective feedback in second language classrooms.' *Language Teaching* 46/1: 140.

MacWhinney, B. 2012. 'The logic of the unified model' in Gass and Mackey (eds.).

Marsden, E. 2005. Input-based grammar pedagogies: A comparison of two possibilities. *Language Learning Journal* 31/1, 9-20.

Marsden, E. 2006. 'Exploring input processing in the classroom: An experimental comparison of processing instruction and enriched input.' *Language Learning* 56/3, 507-566.

Marsden, E. and **Chen, H-Y.** 2011. 'The roles of structured input activities in processing instruction and the kinds of knowledge they promote.' *Language Learning* 61/4, 1058-1098.

Marsden, E., Altmann, G., and **St Claire, M.** 2013. 'Priming of verb inflections in L1 and L2 French: A comparison of 'redundant' and 'non-redundant' training conditions.' *International Review of Applied Linguistics* 51/3, 271-298.

Mitchell, R. 2000. 'Applied linguistics and evidence-based classroom practice: the case of foreign language grammar pedagogy.' *Applied Linguistics* 21/3: 281-303.

Muñoz, C. 2008. 'Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning.' *Applied Linguistics* 29/4: 578-596.

Norris, J. M., and **Ortega, L.** 2000. 'Effectiveness of L2 instruction: a research synthesis and quantitative meta-analysis.' *Language Learning* 50/3: 417-528.

Norris, J. M., Plonsky, L., Ross, S. J., and **Schoonen, R.** 2015. 'Guidelines for reporting quantitative methods and results in primary research.' *Language Learning* 65/2: 470-476.

Philp, J., Mackey, A., and **Oliver, R.** 2008. 'Child's play? Second language acquisition and the younger learner in context' in Philp, Oliver, and Mackey (eds.).

Philp, J., R. Oliver and **A. Mackey**. 2008. *Second Language Acquisition and the Younger Learner: Child's Play?* John Benjamins Publishing Company.

Plonsky, L. and **Oswald, F.L.** (2014). 'How big is "big"? Interpreting effect sizes in L2 research.' *Language Learning* 64/4, 879-912.

Rebuschat, P. 2013. 'Measuring implicit and explicit knowledge in second language research.' *Language Learning* 63/3: 595-626.

Reinders, H., and **Ellis, R.** 2009. 'The effects of two types of input on intake and the acquisition of implicit and explicit knowledge' in Ellis, Loewen, Elder, Erlam, Philp, and Reinders (eds.).

Rhodes, N.C. and **Pufahl, I.** (2009). *Foreign language teaching in US schools: Results of a national survey.* Washington: Center for Applied Linguistics.

Robinson, P. 2003. 'Attention and memory during SLA' in Doughty and Long (eds.).

Roehr, K. 2008. 'Metalinguistic knowledge and language analytic ability in university-level L2 learners.' *Applied Linguistics* 29/2: 173-199.

Schicker, C., M. Waltl, and **C. Malz.** 2011. *Zoom Deutsch 1*. Oxford University Press.

Schmidt, R. 1990. 'The role of consciousness in second language learning.' *Applied Linguistics* 11/2: 129-158.

Schmidt, R. (2010). 'Attention, awareness, and individual differences in language learning.' In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J.W. Sew, T. Suthiwan, and I. Walker, *Proceedings of CLaSIC 2010*, Singapore, December 2-4 (pp. 721-737).

Sealey, A., and **Thompson, P.** 2009. 'Corpus, concordance, classification: young learners in the L1 classroom.' *Language Awareness* 16/3, 208-223.

Serrano, R. 2011. 'From metalinguistic instruction to metalinguistic knowledge, and from metalinguistic knowledge to performance in error correction and oral production tasks.' *Language Awareness* 20/1, 1-16.

Sharwood Smith, M., and **Truscott, J.** 2014. 'Explaining input enhancement: a MOGUL perspective.' *International Review of Applied Linguistics* 52/3: 253-281.

Shintani, N. 2015. 'The effectiveness of Processing Instruction and production-based instruction on L2 grammar acquisition: a meta-analysis.' *Applied Linguistics* 36/3: 306-325.

Spada, N. 2015. 'SLA research and L2 pedagogy: misapplications and questions of relevance.' *Language Teaching* 48/1: 69-81.

Spada, N., and Tomita, Y. 2010. 'Interactions between type of instruction and type of language feature: a meta-analysis.' *Language Learning* 60/2: 263-308.

Stafford, C. A., Bowden, H. W., and **Sanz, C.** 2012. 'Optimizing language instruction: matters of explicitness, practice and cue learning.' *Language Learning* 62/3: 741-768.

Svalberg, A. M.-L. 2012. 'Language awareness in language learning and teaching: a research agenda.' *Language Teaching* 45/3: 376-388.

Tellier, A., and **Roehr-Brackin, K.** 2013. 'The development of language learning aptitude and metalinguistic awareness in primary-school children.' *Essex Research Reports in Linguistics* 62/1: 1-28.

VanPatten, B. 2012. 'Input processing' in Gass and Mackey (eds.).

VanPatten, B. 2015. 'Foundations of Processing Instruction.' *International Review of Applied Linguistics* 53/2: 91-109.

VanPatten, B., and **Borst, S.** 2012. 'The role of explicit information and grammatical sensitivity in Processing Instruction: nominative-accusative case marking and word order in German L2.' *Foreign Language Annals* 45/1: 92-109.

VanPatten, B., and **Cadierno, T.** 1993. 'Explicit instruction and input processing.' *Studies in Second Language Acquisition* 15/2: 225-243.

VanPatten, B., and **Oikkenon, S.** 1996. 'Explanation versus structured input in Processing Instruction.' *Studies in Second Language Acquisition* 18/4: 495-510.

White, J. 2008. 'Speeding up acquisition of *his* and *her*: explicit L1/L2 contrasts help.' in Philp, Oliver, and Mackey (eds.).

White, J., Muñoz, C. and **Collins, L.** 2007. 'The his/her challenge: making progress in a 'regular' L2 programme.' *Language Awareness* 16/4, 278-299.

White, L., Spada, N., Lightbown, P. M., and **Ranta, L.** 1991. 'Input enhancement and question formation.' *Applied Linguistics* 12/4: 416-432.

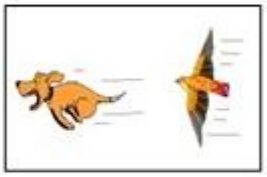
Williams, J. 2005. 'Learning without awareness'. *Studies in Second Language Acquisition* 27/2: 269-304.

Supplementary Figure 1: Vocabulary list

Gender	Nouns: People	Nouns: Animals	Nouns: Objects
Masculine	Mann	Elefant	Brief
	Junge	Hund	Ball
	Vater	Affe	Stuhl
	Sohn	Panda	Tisch
	Onkel	Fisch	Computer
	Opa	Vogel	Bus
	Enkelsohn	Hahn	Traktor
	Schüler	Bulle	Kuchen
	Lehrer	Tiger	Apfel
	Professor	Löwe	Teddybär
	Student	Hamster	
	Farmer	Papagei	
	Clown	Flamingo	
		Bär	
Feminine	Mutter	Kuh	Tür
	Oma	Katze	Banane
	Tante	Maus	Hausaufgabe
	Cousine	Spinne	Uhr
	Tochter	Schlange	Pizza
	Enkeltochter	Biene	
	Frau		
	Lehrerin		
	Fußballspielerin		
	Tennisspielerin		
Neuter	Mädchen	Kaninchen	Fenster
	Baby	Schwein	Buch
	Kind	Schaf	Eis
		Pferd	Frisbee
		Meer-schweinchen	Fahrrad

Verbs	
begrüßen	to greet
benutzen	to use
besuchen	to visit
erschrecken	to scare
essen	to eat
fotografieren	to photograph
fragen	to ask
füttern	to feed
haben	to have
hören	to hear
kaufen	to buy
kicken	to kick
küssen	to kiss
lesen	to read
lieben	to love
machen	to do / make
milken	to milk (a cow)
öffnen	to open
putzen	to clean
rufen	to call
schlagen	to hit (so-sth)
schreiben	to write
sehen	to see
streicheln	to stroke
treffen	to hit (sth-so)
umarmen	to hug
wäschen	to wash
verfolgen	to chase

Supplementary Figure 2: Item from metalinguistic Sentence Reconstruction test



verfolgt | **Vogel** | **der** | **Hund** | **den**

Supplementary Figure 3: Items from written Sentence Matching and Gap-fill tests

A. Das Baby umarmt den Opa. A B

(the-NOM/ACC baby hugs the-ACC grandfather)

B. Der Opa umarmt das Baby.

(the-NOM grandfather hugs the-NOM/ACC baby)

SVO sentence, missing object (SV<u>O</u>):

Die Katze verfolgt …….…..…..…… (the hamster).

the-NOM/ACC cat chases ……..

Supplementary Figure 4: Descriptive statistics for comprehension of SVO items and production of der

Modality	Task	Group	N	Pre		Post		Delayed	
				M	*SD*	*M*	*SD*	*M*	*SD*
Written	Sentence Matching (SVO) (k = 12)	TE-FM	45	11.48	1.37	11.16	1.49	10.89	1.32
		TE-F	38	11.76	1.32	11.13	1.51	11.16	1.60
		Control	50	10.70	1.93	11.04	1.62	-	-
	Gap-fill (SVO) (k = 6)	TE-FM	45	5.77	0.67	5.29	1.16	5.33	1.28
		TE-F	38	5.34	1.34	5.18	1.49	5.42	1.24
		Control	50	5.14	1.83	5.42	1.60	-	-
	Gap-fill (OVS) (k = 6)	TE-FM	45	5.60	0.72	4.56	1.95	5.13	1.39
		TE-F	38	5.50	1.22	5.00	1.64	5.29	1.33
		Control	50	5.01	1.76	5.18	1.77	-	-
Oral	Act-Out Comprehension (SVO) (k = 9)	TE-FM	45	8.80	0.40	8.73	0.58	8.67	0.64
		TE-F	41	8.54	0.81	8.63	0.62	8.63	0.66
		Control	46	8.35	1.06	8.78	0.42	-	-
	Act-Out Production (Subject) (k = 12)	TE-FM	45	9.76	3.54	11.04	2.46	11.47	1.32
		TE-F	41	10.29	3.27	10.93	2.69	10.66	3.28
		Control	46	9.87	3.40	10.02	3.24	-	-
	Sentence Repetition (SVO) (k = 3)	TE-FM	45	2.60	0.65	2.47	0.73	2.44	0.91
		TE-F	41	2.76	0.49	2.54	0.90	2.10	1.16
		Control	46	2.30	0.94	2.54	0.75	-	-
	Sentence Repetition (OVS) (k = 3)	TE-FM	45	1.98	1.10	2.47	0.87	2.69	0.56
		TE-F	41	2.29	0.90	2.32	0.99	2.39	0.86
		Control	46	2.01	1.05	2.28	0.96	-	-

Supplementary Figure 5: Results of Friedman's ANOVA analysing change over time in TE-FM and TE-F groups' scores for comprehension of SVO items and production of der

Modality	Task	Group	χ^2	*p*
Written	Sentence Matching (SVO)	TE-FM	4.105	0.128
		TE-F	3.694	0.158
	Gap-fill (SVO)	TE-FM	3.205	0.201
		TE-F	0.775	0.679
	Gap-fill (OVS)	TE-FM	6.743	0.034
		TE-F	3.840	0.147
Oral	Act-Out Comprehension (SVO)	TE-FM	1.069	0.586
		TE-F	0.644	0.725
	Act-Out Production (Subject)	TE-FM	11.341	0.003
		TE-F	1.740	0.419
	Sentence Repetition (SVO)	TE-FM	0.804	0.669
		TE-F	10.603	0.005
	Sentence Repetition (OVS)	TE-FM	14.000	0.001
		TE-F	0.022	0.989