Applied Linguistics 29/2: 173–199
doi:10.1093/applin/amm037

Metalinguistic Knowledge and Language Ability in University-Level L2 Learners

KAREN ROEHR

University of Essex

Existing research indicates that instructed learners' L2 proficiency and their metalinguistic knowledge are moderately correlated. However, the operationalization of the construct of metalinguistic knowledge has varied somewhat across studies. Metalinguistic knowledge has typically been operationalized as learners' ability to correct, describe, and explain L2 errors. More recently, this operationalization has been extended to additionally include learners' L1 language-analytic ability as measured by tests traditionally used to assess components of language learning aptitude. This article reports on a study which employed a narrowly focused measure of L2 proficiency and incorporated L2 language-analytic ability into a measure of metalinguistic knowledge. It was found that the linguistic and metalinguistic knowledge of advanced university-level L1 English learners of L2 German correlated strongly. Moreover, the outcome of a principal components analysis suggests that learners' ability to correct, describe, and explain highlighted L2 errors and their L2 language-analytic ability may constitute components of the same construct. The theoretical implications of these findings for the concept of metalinguistic knowledge in L2 learning are considered.

INTRODUCTION

University-level second language (L2) instruction aimed at advanced language learners often utilizes grammar books, either to structure a specific focus-on-forms strand of the language course as a whole, or as supplementary material in a focus-on-form course. Pedagogical grammar books target a comprehensive set of morphological, syntactic, semantic, and pragmatic aspects of the L2 (e.g. Durrell 1992, 1996; Dreyer and Schmitt 2001 for L1 English learners of L2 German). Hence, learners are often exposed to explicit teaching and learning of aspects of the L2 that permit systematic description. In view of the assumption that such teaching and learning will be of benefit, it is of interest to establish the nature of the relationship between learners' L2 proficiency and their L2 metalinguistic knowledge, or explicit knowledge about the L2.

Over the past two decades, several studies have addressed this issue (e.g. Sorace 1985; Green and Hecht 1992; Alderson *et al.* 1997; Elder *et al.* 1999; Renou 2000; Elder and Manwaring 2004), and a fuller picture is beginning to emerge. At the same time, however, the definition and

operationalization of the notion of metalinguistic knowledge has varied somewhat across studies. Thus, whilst the practical relevance of gaining an understanding of the role of metalinguistic knowledge in instructed L2 learning is all but undisputed, the theoretical basis of research concerned with the construct of metalinguistic knowledge is arguably not yet fully established.

Accordingly, the present study had two aims, namely (1) to provide further insight into the relationship of university-level learners' L2 proficiency and their L2 metalinguistic knowledge, and (2) to investigate the hypothesized components of metalinguistic knowledge itself. These issues were addressed in the context of a correlational research design incorporating a narrowly focused measure of L2 proficiency and a two-part measure of L2 metalinguistic knowledge that reflected both the more traditional operationalization of the construct as learners' ability to correct, describe, and explain faulty sentences, and a more recently hypothesized component of the construct, that is, learners' language-analytic ability.

L2 PROFICIENCY, METALINGUISTIC KNOWLEDGE, AND LANGUAGE-ANALYTIC ABILITY

Existing empirical research includes studies with longitudinal (e.g. Klapper and Rees 2003) and cross-sectional designs (e.g. Bialystok 1979; Sorace 1985; Green and Hecht 1992; Alderson *et al*. 1997; Elder *et al*. 1999; Renou 2000). Overall, four main findings have arisen from such research. First, when comparing learners' ability to correct L2 errors and to state the violated grammar rules, it was found that students did not necessarily acquire the rules they had been taught (Sorace 1985; Green and Hecht 1992). However, being unable to state the pedagogical grammar rule did not mean that learners were consequently less able to correct L2 items instantiating the rule in question (Sorace 1985; Green and Hecht 1992; Elder *et al*. 1999). Second, researchers report that some rules and categories of pedagogical grammar had been acquired and were applied more successfully than others (Bialystok 1979; Green and Hecht 1992; Renou 2000). Third, larger-scale correlational studies have revealed the inter-learner variability of metalinguistic knowledge as well as the variable application of such knowledge across tasks (Alderson *et al*. 1997; Elder *et al*. 1999; see also Clapham 2001). Fourth, positive correlations between L2 proficiency and metalinguistic knowledge have been identified. However, these correlations were often only moderate in strength, typically ranging from around 0.3 to around 0.5 (Alderson *et al*. 1997; Elder *et al*. 1999), although a recent study has yielded a more mixed pattern which included stronger coefficients ranging from around 0.6 to around 0.7, as well as altogether non-significant results (Elder and Manwaring 2004). Overall, it appears that the relationship between L2 proficiency and metalinguistic knowledge is less substantial than one might

expect, especially given the widespread use of pedagogical grammar in university classrooms. Moreover, significant positive correlations were mainly obtained on the basis of written measures of L2 proficiency.

Several explanations for the often moderate, yet somewhat differing levels of correlational strength seem feasible: mediating variables such as the distance of the L1–L2 combination under investigation (Elder and Manwaring 2004), participants' L2 proficiency levels (Butler 2002; Roehr 2005), length and type of prior language study (Alderson *et al*. 1997; Elder *et al*. 1999), and individual learner differences in cognitive or learning style (Collentine 2000) may have had an impact. Furthermore, the tests that are used to measure L2 proficiency and metalinguistic knowledge, that is, the operationalization of the constructs under investigation, may be a mediating factor as well.

The larger-scale correlational studies cited here employed comprehensive L2 proficiency test batteries which included grammar, cloze, and C-tests, reading comprehension and writing tests, as well as listening comprehension tests (Alderson *et al*. 1997), or a subset of these measures used in conjunction with university-internal achievement tests covering the 'four skills' (Elder *et al*. 1999; Elder and Manwaring 2004). Scrutiny of the tests employed to measure learners' metalinguistic knowledge reveals some noticeable differences across studies. Typically, metalinguistic knowledge is operationalized as learners' ability to correct, describe, and explain errors (e.g. Green and Hecht 1992; Renou 2000).[1] Furthermore, some researchers included tests of learners' ability to label parts of speech (Alderson *et al*. 1997; Elder *et al*. 1999; Elder and Manwaring 2004), a task which, broadly-speaking, likewise draws on metalinguistic description ability. Also, some studies measured both L1 and L2 metalinguistic knowledge (Green and Hecht 1992; Alderson *et al*. 1997), while others exclusively focused on L2 metalinguistic knowledge (Bialystok 1979; Elder and Manwaring 2004).

Finally, and most interesting to the present discussion, two recent studies (Alderson *et al*. 1997; Elder *et al*. 1999) additionally employed tests of learners' L1 language-analytic ability, which can be defined as a learner's 'capacity to infer rules of language and make linguistic generalizations or extrapolations' (Ranta 2002: 161, referring to Skehan 1998). In the two studies reviewed here, language-analytic ability was either treated as a separate construct to begin with (Alderson *et al*. 1997), or as an integrated component of metalinguistic knowledge (Elder *et al*. 1999). More specifically, both Alderson *et al*. (1997) and Elder *et al*. (1999) used a dedicated test of inductive language learning ability as well as a measure of grammatical sensitivity, that is, the words-in-sentences subtest (Part IV) of the Modern Languages Aptitude Test (MLAT) (Carroll and Sapon 2002).

According to the classic model of language learning aptitude developed by John B. Carroll (Carroll 1990; Carroll and Sapon 2002), inductive language

learning ability and grammatical sensitivity are two of the four constituent abilities of aptitude (Nagata *et al*. 1999; de Bot *et al*. 2005; Dörnyei 2005):

1 phonetic coding ability, i.e. the ability to identify and remember sounds in the L2;
2 grammatical sensitivity, i.e. the ability to recognize how words function grammatically in sentences;
3 inductive language learning ability, i.e. the ability to infer grammatical rules from language examples;
4 rote-learning ability, i.e. the ability to form and remember associations between sounds and meaning.

Whilst the MLAT is intended to measure these four components of language learning aptitude, its subtests are not necessarily direct operationalizations. In accordance with psychometric tradition (Carroll 1981, 1993), the MLAT was developed on the basis of empirical data gleaned from large-scale factor-analytic studies, so the test itself preceded the more detailed theoretical conceptualization of the underlying construct. Thus, the MLAT consists of five subtests (Carroll and Sapon 2002).

The words-in-sentences subtest (Part IV) of the MLAT can be seen as a direct measure of grammatical sensitivity. It requires participants to identify the grammatical role of parts of speech in English sentences. Test takers are presented with a key sentence in which one part of speech is underlined. This is followed by a second sentence in which five parts of speech are underlined. Participants are asked to select the option which they believe plays the same grammatical role as the underlined word(s) in the key sentence. Conversely, none of the MLAT subtests directly measures inductive language learning ability, even though it has been suggested that the number-learning subtest (Part I) may tap this ability to a limited extent (Carroll 1981, 1990).

Carroll's four-component model of language learning aptitude was updated in the wake of empirical studies conducted in the 1980s (Skehan 1986, 1989), which, based on the identification of mainly analytically-oriented and mainly memory-oriented learner types, led to the proposal that the components of grammatical sensitivity and inductive language learning ability be subsumed under a single label, that is, language-analytic ability. This reconceptualization was further justified by the theoretically motivated claim that the two components appear to differ only in their degree of emphasis, rather than in qualitative terms (Skehan 1998; Dörnyei 2005). More specifically, both grammatical sensitivity and inductive language learning ability are believed to play a part in the same L2 processing stages, that is, the identification and generalization of linguistic patterns (Skehan 1998, 2002). In several recent discussions of the construct of aptitude, the notion of language-analytic ability in the sense of a learner's ability to identify and extrapolate linguistic patterns has been adopted (Ranta 2002; Dörnyei and Skehan 2003; Erlam 2005).

Hence, at a conceptual level, a primarily analytic component of aptitude comprising grammatical sensitivity and inductive language learning ability—as subsumed under the label of language-analytic ability—may be distinguished from the primarily memory-based components of phonetic coding ability and rote-learning ability. It is noteworthy, however, that while reference to the theoretical notion of language-analytic ability is relatively widespread, the operationalization of the construct has varied somewhat.

Research directly investigating the relationship between components of language learning aptitude, metalinguistic knowledge, and the role of these notions with respect to L2 proficiency is as yet scarce. Closely related to this, the question of how the cognitive abilities measured by aptitude tests facilitate learning under different instructional conditions has been raised (Robinson 2001, 2005; Sawyer and Ranta 2001), and answers are beginning to be forthcoming. As the use of metalinguistic knowledge in the L2 classroom could be viewed as a particular instructional condition, it is likewise worth asking how components of aptitude relate to this construct.

Although not immediately concerned with the notion of metalinguistic knowledge, Erlam (2005) found that, in adolescent L1 English learners of L2 French, deductive instruction involving explicit rule explanation, form-focused activities, output practice, and corrective feedback seemed to minimize the effects of individual learner differences in phonetic coding ability and language-analytic ability, operationalized by means of the words-in-sentences subtest of the MLAT. By contrast, Ranta (2002) concluded that in adolescent L1 French learners of L2 English, a communicative classroom environment apparently could not counteract the effects of individual differences in language-analytic ability. Put differently, language-analytic ability seemed to impact on learner performance regardless of instructional condition. In Ranta's study, language-analytic ability was operationalized by means of a written L1 error detection and correction task.

In her theoretical discussion, Ranta additionally proposes that language-analytic and metalinguistic ability are overlapping concepts. Accordingly, the words-in-sentences subtest of the MLAT is described as a '*de facto* metalinguistic task' (2002: 162). It is argued that while aptitude may be viewed as a stable trait, metalinguistic ability refers to a range of skills which differentially emerge over the course of a learner's development. Hence, aptitude, and in particular language-analytic ability, may be seen as affecting the development of metalinguistic skill, so that language-analytic ability and metalinguistic skill can be viewed as two sides of the same coin.

As indicated above, a further study incorporating the notion of language-analytic ability was conducted by Alderson *et al.* (1997), who investigated the relationship between L2 proficiency, L1 and L2 metalinguistic knowledge, and L1 language-analytic ability in L1 English university-level learners of L2 French. Unlike most of their colleagues, the researchers directly operationalized both of the original notions subsumed under the label of

language-analytic ability, that is, grammatical sensitivity and inductive language learning ability. Grammatical sensitivity was assessed by means of the words-in-sentences subtest of the MLAT. The test of inductive language learning ability presented learners with a short passage in Swahili, a language they were unfamiliar with. An English translation of the first few sentences was provided, and participants were then required to derive the English equivalent of subsequent sentences. Elder *et al.* (1999) used the same metalinguistic test battery and measures of language-analytic ability with a group of L1 English learners of advanced L2 French at an Australian university.

The test of inductive language learning ability did not correlate significantly with any other part of the instruments used in the two studies.[2] However, Alderson *et al.* (1997) did find positive correlations ranging from 0.37 to 0.46 between the words-in-sentences subtest and the various parts of their metalinguistic test battery. The results of a factor analysis produced no clear evidence that performance on the words-in-sentences subtest and metalinguistic knowledge as measured by the metalinguistic test battery were separate factors.

The main issues arising from previous studies concerned with the relationship of L2 proficiency and metalinguistic knowledge can be summarized as follows. Firstly, existing empirical research has uncovered a positive, but mostly moderate relationship between learners' L2 metalinguistic knowledge in the sense of correction, description, and explanation ability, and their L2 proficiency as measured by means of various written tests. Secondly, existing empirical research as well as recent theoretical argumentation suggest that L2 metalinguistic knowledge in the sense of correction, description, and explanation ability and language-analytic ability might be parts of the same underlying construct.

Accordingly, the present study had two main aims. The first aim was to investigate the relationship that would be obtained on the basis of a more narrowly focused written measure of L2 proficiency and a measure of L2 metalinguistic knowledge in the sense of correction, description, and explanation ability. The second aim was to address the hypothesis that L2 metalinguistic knowledge in the sense of correction, description, and explanation ability and language-analytic ability may be components of the same construct.

RESEARCH QUESTIONS

RQ1 What is the relationship between advanced university-level learners' L2 proficiency and their L2 metalinguistic knowledge?

RQ2 What is the relationship between advanced university-level learners' ability to correct, describe, and explain L2 errors and their language-analytic ability, operationalized as the ability to identify the grammatical role of parts of speech in L2 sentences?

CONSTRUCT DEFINITIONS

The construct of L2 proficiency was defined in a narrow sense as learners' knowledge of L2 grammar and vocabulary, that is, a subcomponent of general language ability (Bachman and Palmer 1996). The rationale for this approach was the hypothesis that a more focused operationalization of L2 proficiency concentrating on L2 structures and lexis might lead to a stronger relationship with metalinguistic knowledge, especially if the mostly moderate correlations obtained in previous research were primarily attributable to the operationalization of L2 proficiency via the 'four skills'.

In the most general terms, metalinguistic knowledge can be defined as learners' explicit knowledge about language (e.g. Bialystok 1979; Alderson *et al*. 1997; Elder *et al*. 1999). While implicit knowledge is knowledge that cannot be brought into awareness or articulated, explicit knowledge is declarative knowledge that can be brought into awareness and that is potentially available for verbal report (e.g. Anderson 2005; Hulstijn 2005). More specifically, drawing on Hu (2002) and R. Ellis (2004), the construct of L2 metalinguistic knowledge was defined as a learner's explicit knowledge about the syntactic, morphological, lexical, phonological, and pragmatic features of the L2. It includes explicit knowledge about categories as well as explicit knowledge about relations between categories.

INSTRUMENTATION AND PARTICIPANTS

The L1–L2 combination under investigation was L1 English–L2 German. Learners' L2 proficiency, operationalized as knowledge of L2 grammar and vocabulary in the present study, was assessed by means of a 45-item test (henceforth, 'language test'). Learners were required to produce 22 constrained constructed responses in gap-fill format and respond to 23 multiple-choice items. The language test had been pretested and revised following an item analysis; the amended version was piloted before being employed in the present study (for details, see Roehr 2005). Following item trimming, which reduced the total number of items and thus the maximum number of points that could be scored to 42, the test was highly reliable ($\alpha = 0.913$).

The language test included a range of L2 features which were broadly representative of aspects addressed in tertiary-level foreign language instruction aimed at L1 English-speaking learners of L2 German. Hence, targeted features were based on notions of pedagogical grammar (Swan 1994; Westney 1994; McDonough 2002), rather than a specific linguistic theory. In accordance with this rationale, the language test items covered:

- features of the L2 constituting either real cognates, in the sense that direct English translation equivalents exist (e.g. modal particles), or false cognates, in the sense that apparent analogies between the L1 and L2 mask formal or functional differences (e.g. German *seit* typically

combining with the present tense as opposed to English *since* typically combining with the present perfect tense);
- functional features of the L2 that exist in English but differ in terms of their formal realizations (e.g. word order in subordinate clauses; passive constructions); and
- formal features of the L2 that have no direct equivalents in English (e.g. separable verbs; grammatical gender).

The construct of L2 metalinguistic knowledge was operationalized by means of a two-section test (henceforth, 'metalanguage test'). The first section was aimed at measuring learners' ability to correct, describe, and explain selected L2 features. The second section was aimed at measuring learners' language-analytic ability.

Each test section included 15 items. The description/explanation section consisted of twelve L2 sentences (items 1–12), each of which contained one highlighted error. Learners were required to correct, describe, and explain the highlighted mistakes. A maximum of 12 points could be obtained for successful correction. The description/explanation section further contained three short L2 passages which had been paraphrased in an inappropriate manner (items 13–15). Learners were required to describe and explain why the given paraphrases were unacceptable. This task type was used to take into account L2 features depending more strongly on pragmatic and discursive context, that is, features which could not easily be described and explained on the basis of an isolated sentence.

The description/explanation section tested learners' ability to implement pedagogical grammar rules, since each targeted error or inappropriate paraphrase could be described/explained by means of a statement of the type 'As form X occurs / function X is being expressed, form Y needs to be used'. Essentially, the targeted description answered the question 'What form?', while the targeted explanation answered the question 'Why this form?'. Put differently, learners were required to describe metalinguistic categories as well as explain the relations between these categories. Items targeting syntactic, morphological, and lexical features of the L2 were included. As each of the 15 items was scored separately for description and explanation, this test section yielded a maximum of 30 points.[3]

The items in the description/explanation section of the metalanguage test were designed to match, as far as possible, the items on the language test. The rationale for this approach was that if, as previous research suggests, metalinguistic knowledge is positively correlated with (aspects of) L2 proficiency, the relationship will be revealed best under optimal conditions. If participants' performances on two closely matched tests do not correlate strongly, correlations obtained on the basis of other measures can only be expected to be weaker.

The L2 features included in the language test and the description/explanation section of the metalanguage test are summarized in Table 1.[4]

Table 1: L2 features included in the language test and the description/ explanation section of the metalanguage test

Metalanguage test: Item no.	L2 features	Language test: Item no.
1	Separable verbs	24
2	Prepositions and cases (accusative/ dative)	3, 4, 15, 16, 41, 43, 44
3	Attributively used adjectives/ adjectival inflection	23, (27), 28, 29, 30, 31
4	Lexically expressed directional movement	5
5	*Seit* and present tense	22
6	Subordinating conjunctions/word order in subordinate clauses	10, 32, 36
7	Past subjunctive (*Konjunktiv II*)	1, 2, 25, 26, 34
8	Genitive case	33, 35
9	Collocations: Idiomatic use of the L2	17, 21
10	Attributively used adjectives/ adjectival inflection	23, (27), 28, 29, 30, 31
11	Past subjunctive (*Konjunktiv II*)	1, 2, 25, 26, 34
12	Negation (*nicht* versus *kein*)	13, 14
13	Passive and alternatives to the passive	7
14	Past participle	40
15	*Lassen* as an alternative to the passive/infinitive constructions without *zu*	6, 8, 9, 39
—	Grammatical gender	(37), 38
—	Modal particles: Idiomatic use of the L2	11, (12), 18, 19, 20, 42, 45

The language-analytic section of the metalanguage test consisted of 15 items requiring learners to identify the grammatical role of highlighted parts of L2 sentences. This section was modelled on the words-in-sentences subtest of the MLAT; unlike previous research, however, the current study operationalized language-analytic ability in terms of the L2. This decision was informed by the construct definition of metalinguistic knowledge given above. When completing the language-analytic section of the metalanguage test, learners were again required to employ their knowledge about grammatical categories and relations between grammatical categories typically occurring in L2 German pedagogical grammar. Examples include 'subject', 'relative pronoun', 'object in the dative case', etc. No metalinguistic

labelling or use of technical terminology was needed in this section, since learners were presented with a sentence in which one part of speech had been highlighted. In a four-way multiple-choice task, they were then required to indicate in a second sentence the appropriate part of speech playing an analogous grammatical role. Sample items from the language test and the metalanguage test can be found in the Appendix (http://applij.oxfordjournals.org/cgi/content/full/amm037/DC1).

The metalanguage test had been pretested, amended, and piloted (for details, see Roehr 2005). Item trimming reduced the final number of items and thus the maximum number of points that could be achieved on the language-analytic section to 12. Therefore, the total number of points that could maximally be attained on the metalanguage test was 54 (12 for correction, 15 for description, 15 for explanation, 12 for language analysis). The revised version as used in the current study resulted in somewhat mixed reliability indices ($\alpha = 0.640$ for correction; $\alpha = 0.818$ for description/explanation; $\alpha = 0.624$ for language analysis). The relatively low reliability of the correction and language-analytic sections needs to be borne in mind when interpreting the results of the present study. Given the small number of items in these sections, however, decreased reliability was not unexpected.

The original pool of participants consisted of 60 mostly L1 English-speaking learners (43 females, 17 males; mean age 20.1 years) enrolled as full-time undergraduate students of Advanced German at a British university. A total of 34 participants were in their first year of undergraduate study; the remaining 26 participants were in their fourth and final year of study. All participants were exposed to L2 instruction on a regular basis. Classes were skills-oriented and included weekly sessions of grammar practice with a focus-on-forms orientation, oral L2 practice, and written L2 practice. As the learners studied the L2 in an academic setting, the participant sample is representative of a specific subpopulation of language learners, rather than L2 learners more generally.

The language test and the metalanguage test were administered during the learners' regular class time. Five intact groups of students—three first-year and two fourth-year classes—were tested under the supervision of the researcher. In each class, the tests were administered in two separate sessions, with the language test preceding the metalanguage test by one week. The tests were in familiar paper-and-pencil format. Other than the constraints of the lesson, learners were under no time pressure, and all participants completed each of the tests in 50 minutes or less. Due to several learners missing individual test sessions, the final data pool consisted of 52 completed language tests and 54 completed metalanguage tests.

RESULTS

The descriptive statistics for the language test, the metalanguage test, and the sub-sections of the metalanguage test are shown in Table 2.[5]

Table 2: Descriptive statistics (all learners)

	Language test	Metalanguage test	Correction	Description/ explanation	Language analysis
No. of valid protocols	52	54	54	54	54
No. of items	42	54	12	30	12
Mean % correct	58	49	60	39	63
Mean score	24.27	26.46	7.15	11.8	7.52
Standard deviation	8.993	8.878	2.269	5.041	2.353
Minimum	6	12	3	4	3
Maximum	40	45	11	22	12

Table 2 shows that, overall, the metalanguage test was more challenging for the participants than the language test, with the description/explanation section proving most difficult. Moreover, there is greater variation in learners' L2 grammar and vocabulary competence than in their correction, description/explanation, and language-analytic ability.

Descriptive statistics for the language test, the metalanguage test, and the sub-sections of the metalanguage test by year group are shown in Tables 3 and 4. These two tables demonstrate that the language test proved considerably harder for the first-year learners than for the fourth-year learners. Indeed, this is not a surprising result, since the first-year learners are expected to be at a lower level of L2 proficiency. In addition, the language test scores show a larger standard deviation as well as a wider range of scores, indicating that the group of first-year learners may accommodate a greater mix of proficiency levels than the fourth-year group.

Tables 3 and 4 further reveal that the metalanguage test was more challenging for the first-year learners than for the fourth-year learners. Interestingly, the difference between the two groups is not as pronounced as in the case of the language test. Moreover, the first-year group is more homogeneous than the fourth-year group in their performance on the metalanguage test.

Scatterplots of the test scores achieved by the two learner groups are shown in Figures 1–4. In general, the distribution of scores suggests a linear and positive relationship between performance on the language test and the other four variables, that is, performance on the metalanguage test as a whole (Figure 1), the correction task (Figure 2), the description/explanation task (Figure 3), and the language-analytic task (Figure 4).

In order to address RQ1, bivariate correlations (Pearson's r) for the various parts of the instrument were calculated. The suitability of the data set for the use of parametric statistics had been ascertained through Kolomogorov–Smirnov tests, which resulted in non-significant p-values for all parts of the

Table 3: Descriptive statistics (first-year learners)

	Language test	Metalanguage test	Correction	Description/ explanation	Language analysis
No. of valid protocols	26	31	31	31	31
No. of items	42	54	12	30	12
Mean % correct	45	43	52	33	56
Mean score	18.92	22.97	6.29	9.97	6.71
Standard deviation	7.818	7.282	1.953	4.278	1.970
Minimum	6	12	3	4	3
Maximum	36	42	11	20	11

Table 4: Descriptive statistics (fourth-year learners)

	Language test	Metalanguage test	Correction	Description/ explanation	Language analysis
No. of valid protocols	26	23	23	23	23
No. of items	42	54	12	30	12
Mean % correct	70	58	69	48	72
Mean score	29.62	31.17	8.30	14.26	8.61
Standard deviation	6.664	8.784	2.183	5.020	2.426
Minimum	16	12	4	4	4
Maximum	40	45	11	22	12

instrument (Hatch and Lazaraton 1991; Field 2000). The correlation coefficients obtained for the entire sample of learners are shown in Table 5.[6]

Table 5 shows that, in general, all parts of the instrument correlate strongly and at a high level of significance. The only coefficients that do not reach the 0.7 level are the correlation between the language-analytic section of the metalanguage test and the language test, and the correlation between the language-analytic section and the correction section of the metalanguage test. Nonetheless, these correlations are still of medium strength.

In order to probe whether the focused design of the two tests was responsible for these generally strong intercorrelations, coefficients were calculated separately for twelve categories of pedagogical grammar underlying the matched items of the language test and the description/explanation section of the metalanguage test. The results are summarized in Table 6.

The number of items aimed at testing each L2 feature was necessarily only small, with maximum scores ranging from just one to seven in the case of

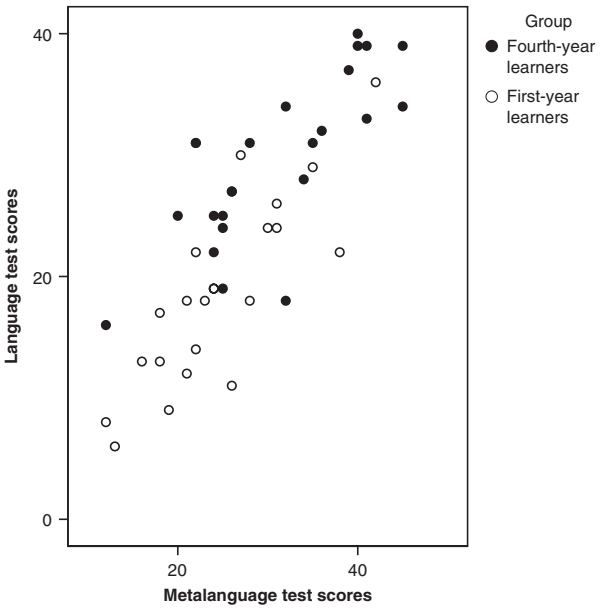

Figure 1: Language test and metalanguage test scores

the language test, and from two to four in the case of the metalanguage test.[7] It is interesting to note that, nonetheless, eight of the twelve individual correlations are significant. Moreover, the correlations are not only positive, but also of reasonable strength, ranging from 0.31 to 0.61. The non-significant results are exclusively based on L2 features represented by only one or two items, which may help to explain the absence of significant correlations in these cases.

Independent samples *t*-tests based on the respective scores achieved by the first-year and the fourth-year learners showed that the two groups of participants differed significantly in their performance on all parts of the instrument, that is on the language test ($t(50) = 5.308$, $p < 0.001$), the metalanguage test as a whole ($t(52) = 3.750$, $p < 0.001$), the correction section of the metalanguage test ($t(52) = 3.564$, $p = 0.001$), the description/explanation section of the metalanguage test ($t(52) = 3.387$, $p = 0.001$), and the language-analytic section of the metalanguage test ($t(52) = 3.173$, $p = 0.003$). Accordingly, separate correlations were calculated for the first-year and fourth-year learners. The results are shown in Tables 7 and 8, respectively.

One noticeable difference between the first-year and fourth-year learners lies in the respective strengths of the relationship between the language and metalanguage test scores. While the two measures correlate strongly in the case of the fourth-year learners, the correlation is somewhat less strong in the case of the first-year learners. For both participant groupings, similar

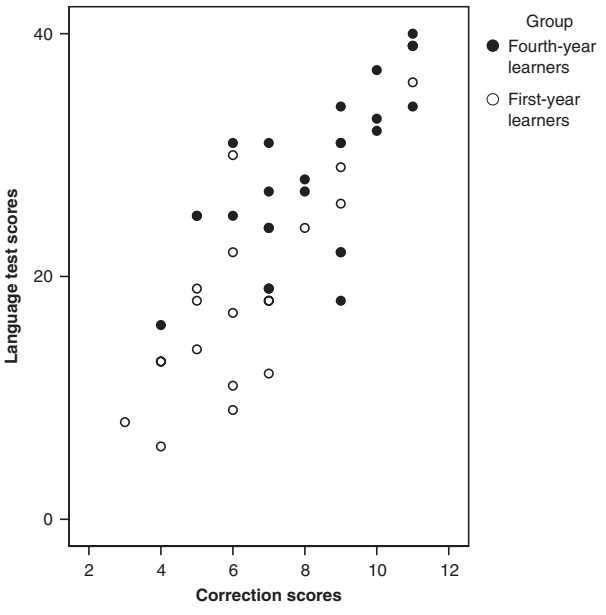

Figure 2: Language test and correction scores

patterns can be observed for the correction between the description/explanation scores and the language test scores, while the opposite pattern obtains for the correction scores. Indeed, the correlation between language test performance and performance on the correction section of the metalanguage test is the only coefficient that is noticeably stronger in the case of the first-year learners than in the case of the fourth-year learners.

The relationship between language test performance and performance on the language-analytic section of the metalanguage test constitutes a second rather striking difference between the two groups. The two measures correlate at 0.64 in the case of the fourth-year learners, but, at 0.47, the correlation is lower in the case of the first-year learners.

In order to address RQ2, a principal components analysis was carried out. As the coefficients in Table 5 show, all parts of the metalanguage test used in the current study intercorrelated strongly and significantly. The suitability of the data set for a principal components analysis was confirmed by calculating the Kaiser–Meyer–Olkin value, which, at 0.719, exceeded the recommended value of 0.6 (Pallant 2005), and by conducting Bartlett's test of sphericity, which, at <0.001, clearly reached statistical significance.

The principal components analysis included three variables, that is, learners' performance on the correction section, the description/explanation section, and the language-analytic section of the metalanguage test. Not unexpectedly in view of the strong intercorrelations, the analysis revealed the presence of a single factor with an eigenvalue above 1

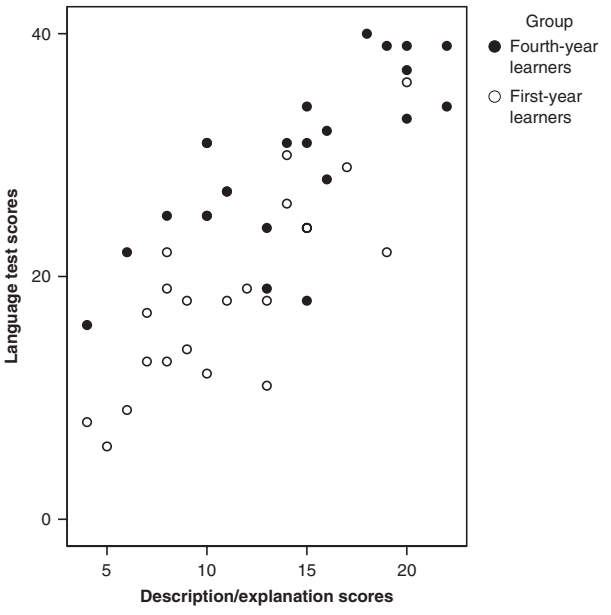

Figure 3: Language test and description/explanation scores

(eigenvalue = 2.467), which explained 82 per cent of the variance.[8] An inspection of the screeplot confirmed that a one-factor solution was indeed appropriate, since a clear break after the first component was in evidence.

Given that medium to strong positive correlations were identified across the instrument, a second principal components analysis was conducted, which included all parts of the instrument as variables. The Kaiser–Meyer–Olkin value (0.838) and Bartlett's test of sphericity (<0.001) had been employed to confirm the suitability of the data set. The analysis again resulted in a single factor with an eigenvalue above 1 (eigenvalue = 3.194), explaining nearly 80 per cent of the variance. Scrutiny of the screeplot supported the appropriateness of a one-factor solution, once more showing a clear break after the first component.

DISCUSSION

With regard to RQ1, two main findings resulted from the correlational analysis. First, taking into account the entire sample of learners, all parts of the instrument were correlated strongly, with the exception of the language-analytic section, which resulted in correlations of medium strength with the language test and the correction section of the metalanguage test. Overall, these results are more substantial than the correlation coefficients obtained in previous research, which mostly found moderate relationships (Alderson *et al*. 1997; Elder *et al*. 1999). Only the correlations obtained in

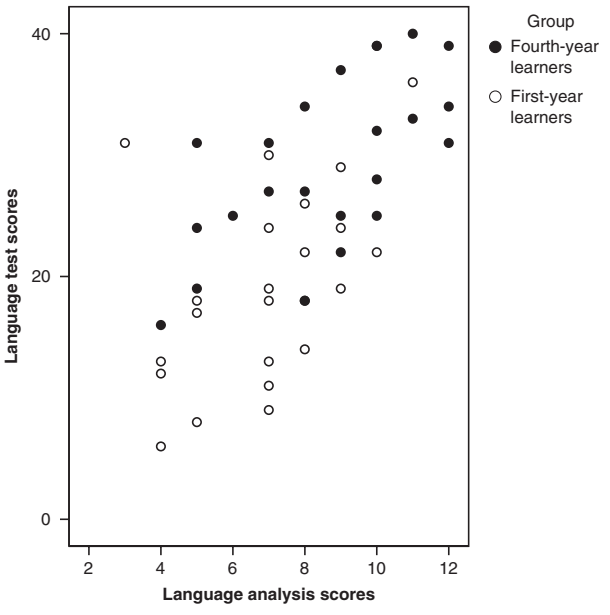

Figure 4: Language test and language analysis scores

one recent study (Elder and Manwaring 2004) approach the strength of the current results.

A plausible explanation for the strong relationship between L2 proficiency as operationalized in the present study and (the first section of) the metalanguage test lies in the design of the instrument. The mostly significant positive correlations obtained on the basis of individual L2 features represented in the two tests provide support for this interpretation. In other words, it appears that the strong relationship between L2 proficiency and L2 metalinguistic knowledge as identified in the current study is indeed at least partly attributable to the narrow focus of the language test on L2 structures and lexis, as well as the matched nature of items across the language test and the metalinguistic description/explanation subtest. Thus, learners who have knowledge of a specific linguistic feature as measured by a structures- and lexis-focused proficiency test often also seem to have explicit knowledge about the feature in question, even though it is not clear if their proficiency arose from their explicit knowledge, or vice versa. By contrast, the language-analytic subtest did not directly reflect the L2 features targeted by the language test and the first section of the metalanguage test. Accordingly, correlations were comparatively weaker.

The L2 features which correlated most strongly across the language and metalanguage tests were 'subordinating conjunctions/word order in subordinate clauses' ($r = 0.61$) and 'attributively used adjectives/adjectival

Table 5: Correlations between language and metalanguage test scores (all learners)

	Language test	Metalanguage test	Correction	Description/ explanation	Language analysis
Language test	1	0.810**	0.800**	0.773**	0.624**
Metalanguage test		1	0.902**	0.966**	0.835**
Correction			1	0.828**	0.667**
Description/explanation				1	0.703**
Language analysis					1

**Significant at the 0.01 level (one-tailed).

inflection' ($r = 0.56$). It is not immediately obvious why this should be the case. Subordination is a syntactic feature, while adjectival inflection is an aspect of morphology. Moreover, subordination can be regarded as a phonologically salient feature, since it is reflected in the word order of a clause, while adjectival inflection lacks phonological salience, invariably involving monosyllabic, unstressed endings. Finally, German adjectival inflection is somewhat opaque in DeKeyser's (2003, 2005) sense, with the same endings potentially signalling different combinations of number, gender, and case. By comparison, subordination is less affected by opacity; indeed, a relatively reliable pedagogical grammar rule can be formulated to describe the phenomenon.

At first glance, then, these contrasting characteristics might lead to the prediction that items testing subordination and adjectival inflection should yield contrasting correlational patterns. However, the two L2 features also have several characteristics in common. Drawing on criteria that have been formulated to account for the differential difficulty of the same L2 features in terms of either implicit or explicit knowledge (R. Ellis 2006), it is possible to identify at least three traits shared by subordination and adjectival inflection. First, even without consulting relevant corpora, it is probably safe to assume that both of these L2 features are comparatively frequent in written and spoken language, contrary to other, less widely used aspects of the L2 included in the instrument, such as past subjunctive or *lassen* as an alternative to the passive. Second, both subordination and adjectival inflection can be described and explained by means of conceptually simple pedagogical grammar rules.

Third, both features are arguably communicatively redundant, constituting formal aspects whose functions, if identifiable to the learner at all, are much more obviously marked on lexical items occurring in constructions involving subordination and adjectival inflection. To exemplify, the communicative value of the subordinating conjunction *weil* (because) is located in its lexical semantics, rather than in the fact that it forces a clustering of verbs at the

Table 6: Correlations between language and metalanguage test scores for individual L2 features (all learners)

L2 feature	Max. score language test	Max. score metalanguage test (description/ explanation section)	Pearson's *r*
Separable verbs	1	2	0.323*
Prepositions and cases (accusative/dative)	7	2	0.318*
Attributively used adjectives/ adjectival inflection	5	4	0.562*
Lexically expressed directional movement	1	2	NS
Seit and present tense	1	2	NS
Subordinating conjunctions/ word order in subordinate clauses	3	2	0.610**
Past subjunctive (*Konjunktiv II*)	5	4	0.352**
Genitive case	2	2	0.433**
Collocations: Idiomatic use of the L2	2	2	0.309*
Negation (*nicht* versus *kein*)	2	2	NS
Passive and alternatives to the passive/*lassen* as an alternative to the passive/ infinitive constructions without *zu*	5	4	0.477**
Past participle	1	2	NS

**Significant at the 0.01 level (one-tailed); *significant at the 0.05 level (one-tailed).

end of the clause. By the same token, in the construction *das Mädchen mit dem roten Mantel* (the girl with the red coat), the lexical semantics of *mit* (with), *rot* (red), and *Mantel* (coat) are of greater communicative value than the inflections of the determiner (*dem*) and the adjective (*roten*). In sum, it is possible to conjecture that the similarities of the two L2 features ultimately carry more weight than their differences—a circumstance which would allow for similar correlational patterns across the language and metalanguage tests in the case of subordination and adjectival inflection.

At this point, it is also worth noting that even though the language test and indeed the correction section of the metalanguage test could be resolved on the basis of implicit knowledge alone, it is by no means certain that learners did not deploy any explicit knowledge when completing these tests.

Table 7: Correlations between language and metalanguage test scores (first-year learners)

	Language test	Metalanguage test	Correction	Description/ explanation	Language analysis
Language test	1	0.768**	0.791**	0.745**	0.466*
Metalanguage test		1	0.854**	0.959**	0.769**

**Significant at the 0.01 level (one-tailed); *significant at the 0.05 level (one-tailed).

Table 8: Correlations between language and metalanguage test scores (fourth-year learners)

	Language test	Metalanguage test	Correction	Description/ explanation	Language analysis
Language test	1	0.804**	0.737**	0.778**	0.638**
Metalanguage test		1	0.903**	0.958**	0.827**

**Significant at the 0.01 level (one-tailed).

While it is generally accepted that explicit and implicit knowledge can be regarded as distinguishable constructs (Paradis 2004; N. Ellis 2005), designing measures which exclusively tap either one or the other type of knowledge in the context of L2 learning and performance is a different matter (see R. Ellis 2005 for a full discussion).

On the one hand, time pressure in combination with certain task types, for example tasks that focus learners' attention on meaning and require oral production, are likely to encourage the use of implicit knowledge. On the other hand, neither task design nor conditions of test administration can guarantee that learners will exclusively draw on either one or the other type of knowledge. In the context of the current study, it is therefore possible that participants used both implicit and explicit knowledge to complete the language test and the correction section of the metalanguage test, especially as the entire instrument was administered in a non-speeded condition.

This circumstance would be compatible with several of the findings obtained in the present study, such as the strong intercorrelation between the language test and the correction section of the metalanguage test in particular, as well as the result of the principal components analysis based on all parts of the instrument, which led to a single-factor solution. Finally, this circumstance might help to explain the finding that the correlation between performance on the language test and the correction section of the

metalanguage test was slightly stronger in the case of the first-year group than in the case of the fourth-year group. The first-year learners may have relied primarily on the same type of knowledge to resolve both tasks, and, possibly, this type of knowledge was implicit; crucially, they appear to have relied on this type of knowledge to a somewhat greater extent than the fourth-year learners.

This point leads to the second main finding arising from the correlational analysis, which was obtained on the basis of a separate treatment of scores attained by the first-year and fourth-year learners. It was found that the language test and metalanguage test scores correlated strongly in the case of the fourth-year learners and somewhat less strongly in the case of the first-year learners. Language test performance and performance on the language-analytic section of the metalanguage test correlated at 0.64 in the case of the fourth-year learners, but only at 0.47 in the case of the first-year learners.

This is an interesting and, arguably, counter-intuitive outcome. As university-level learners are exposed to metalinguistic knowledge in the form of pedagogical grammar throughout their language learning career in various educational settings, they will expect metalinguistic knowledge to help them acquire the L2. Likewise, materials designers and instructors drawing on metalinguistic knowledge for both textbook content and classroom activities will be guided by the assumption that such an approach will enhance the effectiveness of L2 learning and teaching. Finally, existing research as well as the present study confirm that there is indeed a positive relationship between university-level learners' L2 proficiency and their metalinguistic knowledge. Accordingly, one might have hypothesized a stronger correlation for the overall less proficient first-year learners, who can be expected to be more dependent on metalinguistic knowledge than their more advanced fourth-year colleagues. Instead, the opposite was found.

In light of this result, it is possible to speculate that knowledge of grammar and vocabulary as evident in proficient L2 performance may not only be built up on the basis of explicitly acquired metalinguistic knowledge, but may also help a learner develop their metalinguistic knowledge in the first place (see also R. Ellis 2004 for a similar suggestion). Naturally, correlation coefficients merely depict covariance and cannot reveal the direction of any cause–effect relationship, so no firm conclusions about the contribution of metalinguistic knowledge to L2 proficiency or vice versa can legitimately be drawn on the basis of the available statistics. Nonetheless, the results are not inconsonant with the hypothesis that knowledge about language may arise from language competence, rather than (or in addition to) the other way round. Even though it is conceded that the observable differences between the first-year and fourth-year groups are relatively small, the results are compatible with the argument that metalinguistic description/explanation ability, and even more so language-analytic ability, may have different roles

to play at different levels of L2 proficiency. Clearly, however, this conjecture requires further substantiation.

Another possible explanation for the overall stronger correlations obtained for the fourth-year participants may be found in the cumulative effect of individual learner variables over time. Put differently, the combined impact of factors which are likely to foster both successful L2 acquisition and the construction of metalinguistic knowledge may be stronger in the fourth-year learners than in the altogether less experienced first-year learners. Relevant individual learner variables may include general cognitive ability, motivation, and attitudes towards formal language study, for instance. Thus, high levels of general cognitive ability and motivation as well as positive attitudes would mutually reinforce one another over time, ultimately resulting in both higher language test scores and higher metalanguage test scores (see also N. Ellis and Larsen-Freeman 2006 for the interaction of multiple variables in a complex, dynamic system).

With regard to RQ2, a principal components analysis based on the correction, description/explanation, and language-analytic sections of the metalanguage test indicated the presence of a single factor, which explained 82 per cent of the variance. This result is consistent with the hypothesis that the ability to correct, describe, and explain highlighted L2 errors, and the ability to identify the grammatical role of parts of speech in L2 sentences may in fact be components of the same complex construct. In this context, it is important to remember that, unlike previous research that assessed language-analytic ability, the present study operationalized the construct by means of an L2-based measure. This circumstance may help account for both the stronger intercorrelations and the unambiguous result of the principal components analysis.

The proposal that L2 metalinguistic knowledge may be characterized as a complex construct consisting of at least two components, description/explanation ability and language-analytic ability, is further supported by theoretical argumentation. The construct definition used in the present study assumes that metalinguistic knowledge is equivalent to explicit knowledge about L2 categories and relations between categories. The key processes constituting competent use of metalinguistic knowledge defined in this way appear to be analysis (of language) and creative construction (of language) (see N. Ellis 2005). Hence, the description/explanation of an error requires the labelling of a linguistic unit as well as the linking of this unit with a stored pedagogical grammar rule that explains (an aspect of) its use in the context of a phrase, a sentence, or connected discourse. The pedagogical grammar rules targeted in the current study followed the standard pattern 'As form X occurs / function X is being expressed, form Y needs to be used', thus bearing close resemblance to the IF–THEN productions posited in general skill acquisition theory (Anderson 1996, 2005) and applied in conceptualizations of learning strategy use (e.g. O'Malley and Chamot 1990; Macaro 2006).

Identifying the grammatical role of parts of speech requires the identification of a linguistic unit in relation to other linguistic units in a set of sentences. While labelling the unit in question was not part of the operationalization of language-analytic ability in the current study, the explicit identification of a concept is typically closely linked to the naming of the concept (Taylor 2003; Dirven and Verspoor 2004). Once the targeted linguistic unit has been identified in the key sentence, it has to be compared with the linguistic units in the second sentence in order to identify a match.

In summary, then, the two components of metalinguistic knowledge posited here appear to involve the following processes: in the case of description/explanation, labelling a linguistic unit and linking this unit with a previously stored pedagogical grammar rule is required. In the case of language analysis, identification of a linguistic unit in relation to other linguistic units and comparison across sentences is required. At the most general level, the processes of identification, labelling, linking, and comparative matching all involve the deployment of knowledge about categories and relations between categories.

The proposal that L2 metalinguistic knowledge may be characterized as a complex construct consisting of at least two components—description/explanation ability and language-analytic ability—is not necessarily in opposition to Ranta's (2002) suggestion that (L1) language-analytic ability and metalinguistic skill may be two sides of the same coin, with the former notion representing a largely stable and possibly inborn trait (see also R. Ellis 2004), and the latter notion constituting a developmental outcome that is a function of this trait. Instead, the current proposal adds a further dimension, arguing that both L2 language-analytic ability and L2 description/explanation ability are developmental phenomena that constitute components of the complex construct of metalinguistic knowledge. In other words, both L2 language-analytic ability and L2 description/explanation ability are honed in the course of a learner's development: Both abilities are based on the L2, which is being acquired at a relatively mature stage of cognitive development. This argument is further compatible with the arguably more controversial hypothesis put forward above, that is, that L2 metalinguistic knowledge may not only help learners construct knowledge evident in proficient L2 performance, but may have arisen from such knowledge in the first place.

CONCLUSION

The present study addressed two research questions. With respect to RQ1, it was found that in university-level learners of L2 German, knowledge of L2 grammar and vocabulary and L2 metalinguistic knowledge were correlated strongly and significantly. This finding represents an update on previous research. The greater strength of the correlation coefficients obtained in the current study appears to be at least partially attributable to the design of the

instrument used. Put differently, instructed advanced learners' knowledge of L2 structures and lexis and their explicit knowledge about these L2 features co-vary strongly and significantly when matched tests are employed.

When the results obtained from first-year and fourth-year learners, whose performance differed significantly, were considered separately, it was found that the language test and the metalanguage test scores were correlated strongly in the case of the fourth-year learners and somewhat less strongly in the case of the first-year learners. As one possible explanation for this somewhat counterintuitive finding, it was suggested that, contrary to learners' and teachers' expectations, metalinguistic knowledge may be constructed on the basis of increased L2 competence, rather than, or in addition to, being instrumental in building up L2 proficiency. It was acknowledged, however, that this conjecture requires further substantiation, as existing evidence is as yet only indirect.

As another possible explanation, it was proposed that the fourth-year learners' performance reflects the combined effect of a range of cognitive and affective individual difference variables which, over time, have resulted in a stronger association between L2 proficiency and metalinguistic knowledge. When considering the merits of either explanation, however, it is worth bearing in mind that the findings presented here were obtained from a specific sub-population of L2 learners, that is, highly educated university-level language students with considerable exposure to form-focused instruction. Therefore, the current findings may not generalize beyond such learners.

With respect to RQ2, the results of a principal components analysis indicated that the ability to correct, describe, and explain highlighted L2 errors and the ability to identify the grammatical role of parts of speech in L2 sentences appear to be parts of the same multi-componential construct. This finding led to the proposal that L2 metalinguistic knowledge may have to be reconceptualized as a complex notion incorporating, at the very least, L2 description/explanation ability as well as L2 language-analytic ability. It was further noted that the constituent abilities of L2 metalinguistic knowledge in use can be regarded as developmental phenomena that are being built up in the course of an individual's language learning career.

Needless to say, these proposals would benefit from further investigation. In particular, a larger-scale study which makes use of a full range of tests including measures of language learning aptitude as operationalized in the MLAT, measures of L1 metalinguistic knowledge, and measures of L2 metalinguistic knowledge including L2 language-analytic ability would be needed to probe in greater depth the claims that have been put forward here.

Moreover, a longitudinal study, or a study drawing comparisons across several proficiency levels would be needed to provide more informative insights into cause–effect relationships. In other words, the interesting question of whether metalinguistic knowledge about specific L2 features is constructed on the basis of L2 knowledge as measured by proficiency tests,

whether it helps learners acquire such L2 knowledge, or whether both types of knowledge mutually reinforce one another is still waiting to be addressed.

Final version received August 2007

ACKNOWLEDGEMENTS

I would like to thank Charles Alderson, Adela Gánem, Phil Scholfield, and the anonymous *Applied Linguistics* reviewers for their helpful and constructive comments.

NOTES

1 It is worth noting that early studies (e.g. Bialystok 1979; Sorace 1985) operationalized metalinguistic knowledge more broadly by including learners' ability to judge the acceptability of L2 sentences (for a detailed review of measures of explicit knowledge, see also R. Ellis 2004). While some more recent studies also employed acceptability judgements (e.g. Renou 2000), and while many metalinguistic tests appear to include the (identification and) correction of errors as a pre-task to stating rules, researchers would generally agree that acceptability judgements, error identification, and error correction do not necessarily involve the use of metalinguistic knowledge. Thus, scores achieved on such tasks may be treated separately from scores achieved on the unequivocally metalinguistic tasks of explicit description and explanation, as exemplified by labelling parts of speech with appropriate terminology, stating pedagogical grammar rules, etc.

2 It should be added that the test did correlate significantly, although very weakly ($r = 0.23$), with measures of L2 proficiency in beginning learners of L2 Italian assessed by Elder *et al.* (1999). However, these learners only completed the test of inductive language learning ability and the words-in-sentences subtest of the MLAT. As they did not complete the metalinguistic test battery, this result is not included in the present discussion.

3 Scores were awarded to all descriptions/explanations that were relevant and not incorrect with regard to the targeted L2 feature; thus, the scoring criterion was minimal acceptability. A description was considered minimally acceptable if it mentioned the targeted category (e.g. 'accusative'), a superordinate of the targeted category (e.g. 'case'), or the concrete instantiation of the targeted category as it appeared in the task sentence (e.g. *kein* vs. *nicht*). An explanation was considered minimally acceptable if it linked the targeted category with the appropriate function or form by mentioning this function or form either in general terms (e.g. 'possessive'), in concrete terms as it appeared in the task sentence (e.g. *da*), or as a concrete English paraphrase (e.g. 'the house of the writer').

4 Language test items shown in brackets were excluded following item trimming.

5 Statistics were calculated with SPSS for Windows version 12.0.

6 In view of the findings of previous research, which consistently resulted in positive correlations between measures of L2 proficiency and measures of metalinguistic knowledge,

one-tailed tests of significance were chosen. Correlations were also calculated for biodata variables. Variables correlating significantly and positively with performance on the language test were the number of other languages studied apart from the L2 under investigation ($r = 0.304$, $p = 0.015$), the cumulative years of study of these languages ($r = 0.353$, $p = 0.006$), and the number of months of German immersion ($r = 0.321$, $p = 0.010$). Perhaps worryingly for language teachers, years of German study at school correlated significantly and negatively with language test performance ($r = -0.245$, $p = 0.040$). However, this correlation is clearly very weak. The only biodata variable correlating significantly and positively with performance on the metalanguage test was cumulative years of study of languages other than the L2 ($r = 0.315$, $p = 0.013$). The absence of a significant correlation between months of L2 German immersion and performance on the metalanguage test reflects the analogous finding reported in Alderson *et al*. (1997).

7 The reader is reminded that the ffiteen items in the first section of the metalanguage test were in fact scored twice, once for description and once for explanation, so each item yielded a maximum of two points.

8 In a principal components analysis, the eigenvalue refers to an estimate 'of the proportion of variance in each observed variable explained by each factor' (Colman and Pulford 2006: 142).

REFERENCES

Alderson, J. C., C. Clapham, and **D. Steel.** 1997. 'Metalinguistic knowledge, language aptitude and language proficiency,' *Language Teaching Research* 1: 93–121.

Anderson, J. R. 1996. *The Architecture of Cognition*. Mahwah, NJ: Erlbaum.

Anderson, J. R. 2005. *Cognitive Psychology and its Implications* (6th edn). New York, NY: Worth Publishers.

Bachman, L. F. and **A. S. Palmer.** 1996. *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.

Bialystok, E. 1979. 'Explicit and implicit judgements of L2 grammaticality,' *Language Learning* 29/1: 81–103.

Bialystok, E. 1994. 'Representation and ways of knowing: Three issues in second language acquisition' in N. C. Ellis (ed.): *Implicit and Explicit Learning of Languages*. London: Academic Press, pp. 549–70.

Butler, Y. G. 2002. 'Second language learners' theories on the use of English articles: An analysis of the metalinguistic knowledge used by Japanese students in acquiring the English article system,' *Studies in Second Language Acquisition* 24/3: 451–80.

Carroll, J. B. 1981. 'Twenty-five years of research on foreign language aptitude' in K. C. Diller (ed.): *Individual Differences and Universals in Language Learning Aptitude*. Rowley, MA: Newbury House, pp. 83–118.

Carroll, J. B. 1990. 'Cognitive abilities in foreign language aptitude: Then and now' in T. S. Parry and C. W. Stansfield (eds): *Language Aptitude Reconsidered*. Englewood Cliffs, NJ: Prentice Hall, pp. 11–29.

Carroll, J. B. 1993. *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge: Cambridge University Press.

Carroll, J. B. and **S. Sapon.** 2002. *Modern Language Aptitude Test: MLAT. N*. Bethesda, MD: Second Language Testing Inc.

Clapham, C. 2001. 'The assessment of metalinguistic knowledge' in C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. Macnamara, and K. O'Loughlin (eds): *Experimenting with Uncertainty: Essays in Honour of Alan Davies*. Cambridge: Cambridge University Press, pp. 31–41.

Collentine, J. 2000. 'Insights into the construction of grammatical knowledge provided by user-behavior tracking technologies,' *Language Learning and Technology* 3/2: 44–57.

Colman, A. and **B. Pulford.** 2006. *A Crash Course in SPSS for Windows* (3rd edn). Malden, MA: Blackwell.

de Bot, K., W. Lowie, and **M. Verspoor.** 2005. *Second Language Acquisition: An Advanced Resource Book*. London: Routledge.

DeKeyser, R. M. 2003. 'Implicit and explicit learning' in C. J. Doughty and M. H. Long (eds): *The Handbook of Second Language Acquisition*. Malden, MA: Blackwell, pp. 313–48.

DeKeyser, R. M. 2005. 'What makes learning second-language grammar difficult? A review of issues,' *Language Learning* 55/s1: 1–25.

Dirven, R. and **M. Verspoor.** 2004. *Cognitive Exploration of Language and Linguistics* (2nd edn). Amsterdam: John Benjamins.

Dörnyei, Z. 2005. *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*. Mahwah, NJ: Erlbaum.

Dörnyei, Z. and **P. Skehan.** 2003. 'Individual differences in second language learning' in C. J. Doughty and M. H. Long (eds): *The Handbook of Second Language Acquisition*. Malden, MA: Blackwell, pp. 589–630.

Dreyer, H. and **R. Schmitt.** 2001. *A Practice Grammar of German*. Ismaning: Hueber.

Durrell, M. 1992. *Using German: A Guide to Contemporary Usage*. Cambridge: Cambridge University Press.

Durrell, M. 1996. *Hammer's German Grammar and Usage* (3rd edn). London: Arnold.

Elder, C. and **D. Manwaring.** 2004. 'The relationship between metalinguistic knowledge and learning outcomes among undergraduate students of Chinese.' *Language Awareness* 13/3: 145–62.

Elder, C., J. Warren, J. Hajek, D. Manwaring, and **A. Davies.** 1999. 'Metalinguistic knowledge: How important is it in studying a language at university?' *Australian Review of Applied Linguistics* 22/1: 81–95.

Ellis, N. C. 2005. 'At the interface: Dynamic interactions of explicit and implicit language knowledge,' *Studies in Second Language Acquisition* 27/2: 305–52.

Ellis, N. C. and **D. Larsen-Freeman.** 2006. 'Language emergence: Implications for applied linguistics,' *Applied Linguistics* 27/4: 558–89.

Ellis, R. 2004. 'The definition and measurement of L2 explicit knowledge,' *Language Learning* 54/2: 227–75.

Ellis, R. 2005. 'Measuring implicit and explicit knowledge of a second language: A psychometric study,' *Studies in Second Language Acquisition* 27/2: 141–72.

Ellis, R. 2006. 'Modelling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge,' *Applied Linguistics* 27/3: 431–63.

Erlam, R. 2005. 'Language aptitude and its relationship to instructional effectiveness in second language acquisition,' *Language Teaching Research* 9/2: 147–71.

Field, A. 2000. *Discovering Statistics Using SPSS for Windows*. London: Sage.

Green, P. S. and **K. Hecht.** 1992. 'Implicit and explicit grammar: An empirical study,' *Applied Linguistics* 13/2: 168–84.

Hatch, E. and **A. Lazaraton.** 1991. *The Research Manual: Design and Statistics for Applied Linguistics*. Boston, MA: Heinle and Heinle.

Hu, G. 2002. 'Psychological constraints on the utility of metalinguistic knowledge in second language production,' *Studies in Second Language Acquisition* 24/3: 347–86.

Hulstijn, J. H. 2005. 'Theoretical and empirical issues in the study of implicit and explicit second-language learning: Introduction,' *Studies in Second Language Acquisition* 27/2: 129–40.

Klapper, J. and **J. Rees.** 2003. 'Reviewing the case for explicit grammar instruction in the university foreign language learning context,' *Language Teaching Research* 7/3: 285–314.

Macaro, E. 2006. 'Strategies for language learning and for language use: Revising the theoretical framework,' *Modern Language Journal* 90/3: 320–37.

McDonough, S. 2002. *Applied Linguistics in Language Education*. London: Arnold.

Nagata, H., D. Aline, and **R. Ellis.** 1999. 'Modified input, language aptitude and the acquisition of word meanings' in R. Ellis (ed.): *Learning a Second Language through Interaction*. Amsterdam: John Benjamins, pp. 133–49.

O'Malley, J. M. and **A. U. Chamot.** 1990. *Learning Strategies in Second Language Acquisition*. Cambridge: Cambridge University Press.

Pallant, J. 2005. *SPSS Survival Manual* (2nd edn). Maidenhead: Open University Press.

Paradis, M. 2004. *A Neurolinguistic Theory of Bilingualism*. Amsterdam: John Benjamins.

Ranta, L. 2002. 'The role of learners' language analytic ability in the communicative classroom' in P. Robinson (ed.): *Individual Differences and Instructed Language Learning*. Amsterdam: John Benjamins, pp. 159–80.

Renou, J. M. 2000. 'Learner accuracy and learner performance: The quest for a link,' *Foreign Language Annals* 33/2: 168–80.

Robinson, P. 2001. 'Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition,' *Second Language Research* 17/4: 368–92.

Robinson, P. (ed.). 2002. *Individual Differences and Instructed Language Learning*. Amsterdam: John Benjamins.

Robinson, P. 2005. 'Aptitude and second language acquisition,' *Annual Review of Applied Linguistics* 25: 46–73.

Roehr, K. 2005. Metalinguistic Knowledge in Second Language Learning: An Emergentist Perspective. Unpublished Ph.D. Thesis, Lancaster University.

Sawyer, M. and **L. Ranta.** 2001. 'Aptitude, individual differences, and instructional design' in P. Robinson (ed.): *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press, pp. 319–53.

Skehan, P. 1986. 'Cluster analysis and the identification of learner types' in V. Cook (ed.): *Experimental Approaches to Second Language Learning*. Oxford: Pergamon, pp. 81–94.

Skehan, P. 1989. *Individual Differences in Second-Language Learning*. London: Arnold.

Skehan, P. 1998. *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Skehan, P. 2002. 'Theorising and updating aptitude' in P. Robinson (ed.): *Individual Differences and Instructed Language Learning*. Amsterdam: John Benjamins, pp. 69–94.

Sorace, A. 1985. 'Metalinguistic knowledge and language use in acquisition-poor environments,' *Applied Linguistics* 6/3: 239–54.

Swan, M. 1994. 'Design criteria for pedagogic language rules' in M. Bygate, A. Tonkyn, and E. Williams (eds): *Grammar and the Language Teacher*. New York, NY: Prentice Hall, pp. 45–55.

Taylor, J. R. 2003. *Linguistic Categorization* (3rd edn). Oxford: Oxford University Press.

Westney, P. 1994. 'Rules and pedagogical grammar' in T. Odlin (ed.): *Perspectives on Pedagogical Grammar*. Cambridge: Cambridge University Press, pp. 72–96.