Essex Research Reports in Linguistics

Volume 57

Number 5

4 August, 2008

Karen Roehr
University of Essex
kroehr@essex.ac.uk

Gabriela Adela Gánem-Gutiérrez
University of Essex
aganem@essex.ac.uk

Metalinguistic knowledge in instructed L2 learning: An individual difference variable?

Dept. of Language and Linguistics,
University of Essex,
Wivenhoe Park,
Colchester, Essex, UK,
CO4 3SQ

http://www.essex.ac.uk/linguistics/errl/

Essex Research Reports in Linguistics present ongoing research activities of the members of the Department of Language and Linguistics.

The main purpose of these reports is to provide a quick publication outlet. They have 'pre-publication status', and most will subsequently appear in revised form as research articles in professional journals or in edited books.

If you have technical problems downloading a paper, or for further information about these reports, please contact the editor:

Doug Arnold: doug@essex.ac.uk.

Citation Information:

Karen Roehr and Gabriela Adela Gánem-Gutiérrez. 'Metalinguistic knowledge in instructed L2 learning: An individual difference variable?', Essex Research Reports in Linguistics, Vol. 57.5. Dept. of Language and Linguistics, University of Essex, Colchester, UK, August, 2008.
http://www.essex.ac.uk/linguistics/errl/errl57-5.pdf

Metalinguistic knowledge in instructed L2 learning: An individual difference variable?

Karen Roehr

Department of Language & Linguistics
University of Essex
Wivenhoe Park
Colchester CO4 3SQ
UK
kroehr@essex.ac.uk

Gabriela Adela Gánem-Gutiérrez

Department of Language & Linguistics
University of Essex
Wivenhoe Park
Colchester CO4 3SQ
UK
aganem@essex.ac.uk

Abstract

This study investigated second language metalinguistic knowledge, or explicit knowledge about the second language, in English-speaking university-level learners of German and Spanish. The status of metalinguistic knowledge in relation to the individual difference variables of language learning aptitude, working memory for language, and participants' language learning history was identified. Language learning experience in formal settings was found to be the strongest predictor for levels of metalinguistic knowledge attained by the participants. Moreover, it was found that despite a significant relationship with language learning aptitude, metalinguistic knowledge is separable and distinct from both aptitude and working memory. In conclusion, it is suggested that metalinguistic knowledge may be an individual difference variable in its own right—a hypothesis which is compatible with the results of the present study as well as previous research in the field of instructed second language learning.

Introduction

The aim of this article is to contribute to our understanding of metalinguistic knowledge, or explicit knowledge about language, in instructed adult second language (L2) learning. University-level language learners' metalinguistic knowledge was investigated in connection with selected individual difference variables that have been shown to influence adult L2 learning more generally and that have been hypothesized to influence the development of metalinguistic knowledge more specifically. In particular, we report on the status of metalinguistic knowledge in relation to language learning aptitude, working memory for language, and participants' language learning history. We found that language learning experience in formal settings was the strongest predictor for levels of metalinguistic knowledge in our participants. Furthermore, our results indicate that metalinguistic knowledge is distinct from both language learning aptitude and working memory for language. Drawing on these results as well as existing findings in the field, we conclude by hypothesizing that metalinguistic knowledge may be an individual difference variable in its own right.

Metalinguistic knowledge in L2 learning

In the context of this article, the term L2 is used generically to refer to any language(s) other than the learner's first language (L1). Thus, L2 can quite literally refer to a second language, but it can also refer to L3, L4, Lx. Broadly defined, the construct of metalinguistic knowledge refers to a person's explicit knowledge about language (Alderson, Clapham, & Steel, 1997; Bialystok, 1979; Elder, Warren, Hajek, Manwaring, & Davies, 1999). Explicit knowledge is knowledge that can be brought into awareness, that is potentially available for verbal report, and that is represented declaratively. [1] It can be contrasted with implicit knowledge, which cannot be brought into awareness or articulated (Anderson, 2005; Hulstijn, 2005). In the context of the present study, the construct of L2 metalinguistic knowledge is 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 (see also R. Ellis, 2004; Hu, 2002; Roehr, 2008a).

The notion of L2 metalinguistic knowledge, or explicit knowledge about the L2, straddles the areas of language learning and language teaching; accordingly, a fairly large body of research concerned with the role of explicit knowledge in second language

acquisition (SLA) and foreign language instruction is available. In what follows, only research that is central to the present study will be reviewed; our presentation of the literature is therefore selective.

Researchers conducting psycholinguistically oriented studies with an experimental design have often been inspired by the noticing hypothesis (Schmidt, 1990, 2001) which argues that, in practice at least, attention at the level of awareness is required for L2 learning to take place. Researchers have thus primarily drawn on the notion of explicit learning and the potential benefits of such an approach. Accordingly, experimental work has investigated the role of attention in short-term learning experiments (e.g. Camps, 2003; Gass & Alvarez Torres, 2005; Robinson, 1997) as well as the facilitative influence of different levels of learner awareness (e.g. Leow, 2000; Rosa & O'Neill, 1999).

Whereas experimental studies tend to concentrate on a small range of carefully selected L2 features (e.g. Camps, 2003; Leow, 2000; Sanz & Morgan-Short, 2004), pedagogically oriented studies involving university students typically include a broad range of L2 features (e.g. Alderson et al., 1997; Green & Hecht, 1992; Klapper & Rees, 2003; Macaro & Masterman, 2006; Roehr, 2008b), often chosen on the basis of participants' course syllabuses. Pedagogically oriented research thus often works with learners who are exposed to prolonged periods of L2 teaching in classroom settings. Rather than administering specific treatments, such research primarily focuses on the relationship between metalinguistic knowledge and L2 proficiency resulting from classroom-based instruction. Two main findings arising from this type of research are relevant to the current study.

First, positive correlations between levels of written L2 proficiency and levels of metalinguistic knowledge have been identified (Alderson et al., 1997; Elder et al., 1999; Elder & Manwaring, 2004; Roehr, 2008b). Second, it has been suggested that variables such as length and type of learners' prior language study (Elder & Manwaring, 2004; Elder et al., 1999; Renou, 2000; Sorace, 1985), level of L2 proficiency (Butler, 2002; Klapper & Rees, 2003; Roehr, 2008b), typological distance of L1 and L2 (Elder & Manwaring, 2004), individual learner differences in cognitive or learning style (Collentine, 2000), and individual differences in short-term versus long-term uptake of explicit instruction (Macaro & Masterman, 2006) may have an impact on metalinguistic knowledge in instructed L2 learning.

Interestingly, two individual difference variables which can arguably be expected to differentially affect metalinguistic knowledge in adult L2 learning have as yet hardly been investigated in connection with this construct: Language learning aptitude has only rarely

been studied empirically together with metalinguistic knowledge, while, as far as we are aware, working memory for language has not been studied at all in conjunction with metalinguistic knowledge in adult L2 learning.

Language learning aptitude and working memory for language

In the context of SLA, language learning aptitude can be defined as a complex set of abilities that enable some learners to acquire new language material more quickly and with greater ease than others (Dörnyei, 2005). Thus, an individual's performance on a test of language learning aptitude is expected to predict how fast and with how much ease they will learn an L2 relative to other individuals. Tests of aptitude were originally designed for selection and placement purposes; in current SLA research, however, tests of aptitude are primarily used to investigate the construct in relation to L2 learning success under different instructional conditions (Erlam, 2005; Williams & Lovatt, 2003), L2 learners' age profiles (Harley & Hart, 1997; Robinson, 2001), and other individual difference variables such as general cognitive ability (Ehrman & Oxford, 1995; Sasaki, 1996).

The Modern Language Aptitude Test or MLAT (Carroll & Sapon, 2002) probably constitutes the most widely used measure both for practical placement purposes and to achieve research objectives. Although the MLAT has justly been critiqued (for recent reviews, see Dörnyei, 2005; Dörnyei & Skehan, 2003), it has also been endorsed as one of the best instruments available (Sparks & Ganschow, 2001) whose validity has been amply demonstrated (Carroll, 1990).

According to the classic model of language learning aptitude developed by John B. Carroll (Carroll, 1990; Carroll & Sapon, 2002), the construct comprises four components, that is, phonetic coding ability, grammatical sensitivity, inductive language learning ability, and associative memory. 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, 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 construct. Hence, the MLAT consists of five subtests (Carroll, 1990; Carroll & Sapon, 2002): Number learning (MLAT 1), phonetic script (MLAT 2), spelling clues (MLAT 3), words in sentences (MLAT 4), and paired associates (MLAT 5).

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 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 motivated by the correlation between these two components and further justified by the theoretical claim that the two components appear to differ only in their degree of emphasis, rather than in qualitative terms (Dörnyei, 2005; Skehan, 1998). Put differently, the notion of language-analytic ability potentially comprises the capacity to internally derive knowledge about language, e.g. through the discovery of patterns in the input, as well as the application of knowledge about language, whether derived internally or assimilated from external sources. 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 (Dörnyei & Skehan, 2003; Erlam, 2005; Ranta, 2002; Skehan, 2002).

There is as yet relatively little published research which has directly sought to link either language learning aptitude as a whole or specific components of this construct with metalinguistic knowledge (Jessner, 1999, 2006). This is somewhat surprising, since the notion of metalinguistic knowledge arguably shares many characteristics with language-analytic ability in particular. Both of these concepts would appear to require the explicit representation and use of linguistic categories and relations between such categories. Even though metalinguistic knowledge is typically operationalized via the L2 and language-analytic ability as represented in the MLAT draws on the learner's L1, one can reasonably hypothesize a relationship between metalinguistic skills and language-analytic skills, as work by Alderson et al. (1997), Ranta (2002), and Roehr (2008b) demonstrates.

Alderson et al. (1997) investigated the relationship between L2 proficiency, metalinguistic knowledge, and language-analytic ability in L1 English university learners of L2 French. The researchers found positive correlations ranging from .37 to .46 between MLAT 4 and the various parts of their metalinguistic test battery. The results of a factor analysis produced no clear evidence that performance on MLAT 4 and the metalinguistic test battery were separate factors. It is worth noting that the metalinguistic test battery included both L1-based and L2-based measures.

Along similar lines, Ranta (2002) proposed that language-analytic ability and metalinguistic knowledge may be overlapping concepts (see also Jessner, 2006: 68).

Accordingly, Ranta (2002: 162) described MLAT 4 as a de facto metalinguistic task, i.e. a task drawing on explicit knowledge about language. In a study involving adolescent L1 French learners of L2 English, Ranta (2002) found a moderate correlation between the L1 measure of language-analytic ability and the L2 metalinguistic task. In a principal components analysis, the two measures loaded on separate components.

Following a study involving L1 English university learners of L2 German, Roehr (2008b) argued that metalinguistic knowledge is a complex construct consisting of at least two components, that is, the ability to describe and explain aspects of the L2 and L2 language-analytic ability. This proposal arose from the result of a principal components analysis which indicated that the two abilities as operationalized in the study constituted a single factor. Unlike previous research, however, this investigation operationalized the construct of language-analytic ability by means of an L2-based measure.

In sum, existing empirical evidence suggests that, first, metalinguistic knowledge is a multi-componential construct comprising the ability to describe and explain aspects of the L2 as well as L2 language-analytic ability; and, second, that there is, at the very least, a relationship between metalinguistic knowledge and (components of) language learning aptitude. The latter circumstance in particular deserves further investigation because pertinent findings will enhance our understanding of the nature of metalinguistic knowledge itself as well as the role of such knowledge in L2 learning.

The notion of working memory has only fairly recently begun to play a more prominent role in SLA research. Working memory refers to "the system or mechanism underlying the maintenance of task-relevant information during the performance of a cognitive task" (Shah & Miyake, 1999: 1). In other words, working memory allows for the temporary storage and manipulation of information which is being used during online cognitive operations such as language comprehension, learning, and reasoning (Baddeley, 2000). Researchers agree that working memory is limited in capacity; moreover, individuals differ in the maximum amount of activation available to them, i.e. individuals differ in terms of their working memory resources (Just & Carpenter, 1992; Miyake & Friedman, 1998).

In some empirical research, working memory for language has been conceptualized as phonological loop capacity and operationalized by means of digit span or non-word repetition tests (e.g. N. Ellis & Sinclair, 1996; Mackey, Philp, Egi, & Fujii, 2002). Alternatively, it has been conceptualized as involving simultaneous storage and processing of information and operationalized by means of reading or listening span tests (e.g. Harrington & Sawyer, 1992).

Individual learner differences in working memory for language appear to influence L2 development more generally (Erlam, 2005; Mackey et al., 2002) and the acquisition of vocabulary and multi-word units in particular (N. Ellis & Sinclair, 1996; Gathercole & Thorn, 1998). Crucial to the study presented here, some researchers have further suggested that working memory can be regarded as a component of language learning aptitude (McLaughlin, 1995; Robinson, 2005; Sawyer & Ranta, 2001), with aptitude and working memory apparently affecting L2 learning in combination. Even more radically, Miyake & Friedman (1998) have argued that the classic components of language learning aptitude could be seen as components of working memory, thus elevating the latter to the status of the more central construct.

In view of these arguments, it is plausible to hypothesize a relationship not only between working memory for language and language learning aptitude, but also between working memory and metalinguistic knowledge. Importantly, working memory is—metaphorically speaking—the locus of conscious processing (Baddeley, 2000; Baddeley & Logie, 1999). If metalinguistic knowledge can be brought into conscious awareness and articulated, an individual's level of metalinguistic knowledge and their ability to put this knowledge to use can be expected to depend on their working memory resources.

Research issues

The preceding review has highlighted several points which are relevant to our understanding of metalinguistic knowledge in L2 learning. Some of these issues have been investigated, but would benefit from further substantiation; other issues have not yet been addressed empirically.

First, existing empirical and theoretical research suggests that metalinguistic knowledge may be related to and possibly be affected by (components of) language learning aptitude and working memory for language. While there is evidence indicative of a relationship between metalinguistic knowledge and language-analytic ability as operationalized in the MLAT, the other components of aptitude have not yet been considered in any detail in this context. Moreover, no study to date has directly addressed the plausible hypothesis of an association between metalinguistic knowledge and working memory for language in adult L2 learners.

Second, existing research suggests that a range of variables is likely to influence the development of metalinguistic knowledge in instructed L2 learning. These variables include

cognitively-based learner-internal factors as well as length and type of prior language study. Yet, few published studies concerned with metalinguistic knowledge have incorporated the measurement of several such variables into a single research design.

The present study thus addressed two research questions:

RQ1 What is the relationship between university learners' metalinguistic knowledge, language learning aptitude, L1 working memory, and L2 working memory?

RQ2 Which variables predict university learners' level of metalinguistic knowledge?

Participants

A total of 39 L1 English university-level L2 learners participated in the present study; 19 were students of L2 German, 20 were students of L2 Spanish at the same British university.

The L2 German group consisted of 5 males and 14 females and ranged in age from 18 to 65 years (mean = 25.6; median = 19.0). A total of 18 learners were students, with 17 undergraduates and one occasional student. One learner was also a member of university staff. On average, the participants had studied the L2 for 4.4 years at school and/or college and for 0.8 years at university. Moreover, the learners had studied up to three other languages apart from the L2 under investigation (mean = 1.3). These languages included French, Italian, Spanish, Russian, Latin, and British Sign Language.

The L2 Spanish group consisted of 6 males and 14 females and ranged in age from 18 to 46 years (mean = 22.1; median = 20.0). A total of 19 learners were students, all of them undergraduates. One learner was also a member of university staff. On average, the participants had studied the L2 for 2.8 years at school and/or college and for 1.1 years at university. Moreover, the learners had studied up to three other languages apart from the L2 under investigation (mean = 1.7). These languages included French, Italian, Portuguese, German, Dutch, Irish, Latin, and Ancient Greek.

Length of L2 immersion varied considerably between learners, ranging from 0 to 192 weeks in the L2 German group (mean = 5.4) and from 0 to 80 weeks in the L2 Spanish group (mean = 11.9). In either sample, length of L2 immersion did not correlate significantly with the learners' metalinguistic knowledge as measured for the purpose of the study (see below for details of measurement), reflecting results obtained in previous studies (Alderson et al., 1997; Roehr, 2006).

Instrumentation

The participants completed a five-part instrument consisting of a biodata questionnaire, a test of metalinguistic knowledge, the MLAT, a test of L1 reading span, and a test of L2 reading span.

Biodata questionnaire

The biodata questionnaire contained a total of 12 questions about demographic variables (age, gender), the participants' current status at the university where the study was conducted (e.g. undergraduate student, postgraduate student, year of study, etc.), and their language learning history.

Test of metalinguistic knowledge (henceforth: MLK test)

The tests of German and Spanish metalinguistic knowledge were designed by the researchers, following a template used by Roehr (2005; 2008b), originally inspired by Alderson et al. (1997). Each MLK test consisted of two sections. The first section measured learners' ability to explicitly describe and explain aspects of the L2, operationalized as the ability to correct, describe, and explain highlighted sentence-level errors involving selected L2 features. The second section tested learners' L2 language-analytic ability, operationalized as the ability to identify the grammatical role of parts of speech in L2 sentences.

The first test section consisted of 20 L2 sentences, each of which contained one highlighted error. Participants were required to correct, describe, and explain these highlighted errors. [2] The description/explanation task effectively tested learners' ability to implement pedagogical grammar rules, since each targeted error could be described and 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 categories as well as explain the relations between these categories.

Items targeting syntactic, morphological, lexico-semantic, and pragmatic features of the L2 were included. Targeted L2 features were selected on the basis of the participants' language course syllabuses and included both item-based and generalizable aspects of the L2. Examples of targeted L2 German features are the use of case, negation with *nicht* vs. *kein*, formal vs. informal forms of address, and word order in subordinate clauses. Examples of

targeted L2 Spanish features are radical changing verbs, the use of reflexive pronouns, *ser* vs. *estar*, formal vs. informal forms of address, and the use of the conditional.

The second test section consisted of 25 items requiring learners to identify the grammatical role of highlighted parts of L2 sentences. When completing this test section, learners again needed to employ their metalinguistic knowledge about grammatical categories and relations between categories occurring in their language course syllabuses. Examples of targeted L2 German features are indirect objects in the dative case, predicatively used adjectives, and adverbs of manner. Examples of targeted L2 Spanish features are subjects of the main clause, *que* before a verb in the subjunctive, and possessive adjectives.

In this second test section, no description, explanation, explicit labelling, or use of technical terminology was required, since participants were presented with a sentence in which one part of speech had been highlighted. In a four-way multiple-choice task, they were then asked to indicate in a second sentence the appropriate part of speech which they regarded as playing an analogous grammatical role. This second section of the MLK test was modelled on MLAT 4.

The MLK tests were pretested and revised in accordance with pretesters' feedback. The tests as used in the current study showed good reliability after poorly performing items had been excluded. The final reliability indices for the German test sections were as follows: Correction (19 items) $\alpha = .78$; Description (19 items) $\alpha = .83$; Explanation (18 items) $\alpha = .87$; Language analysis (23 items) $\alpha = .80$. The final reliability indices for the Spanish test sections were as follows: Correction (20 items) $\alpha = .83$; Description (20 items) $\alpha = .82$; Explanation (17 items) $\alpha = .88$; Language analysis (22 items) $\alpha = .71$. The full set of L2 features targeted by the MLK tests can be found in Appendix A (German) and Appendix B (Spanish).

MLAT

The MLAT (Carroll & Sapon, 2002) consists of five sections, i.e. number learning (MLAT 1 – 43 items), phonetic script (MLAT 2 – 30 items), spelling clues (MLAT 3 – 50 items), words in sentences (MLAT 4 – 45 items), and paired associates (MLAT 5 – 24 items).

Test of L1 reading span (henceforth: L1 span test)

The English reading span test devised by Daneman & Carpenter (1980) was used to measure the storage and processing components of participants' working memory for their L1. In this test, the informant is presented with sentences which are shown one by one printed across a

card. On being shown a card, the informant is required to read the sentence aloud. As soon as they finish, the card is turned over, and a new sentence is presented. During the test, the informant is required to memorize the last word of each sentence they have read out.

The sentences are organized into sets, which in turn are organized into levels. Each level consists of five sets, but the number of sentences per set increases as the informant progresses through the test. Thus, in the first level, each set contains only two sentences; in the second level, each set contains three sentences, and so forth, up to six sentences in the final level. When they reach the end of a set of sentences, the informant is presented with a blank card. They are then required to write down from memory the last word of each sentence in that set. The entire test comprises 88 stimulus sentences, each containing between 11 and 16 words.

Test of L2 reading span (henceforth: L2 span test)

The L2 reading span tests were designed by the researchers following the template of the L1 span test. In terms of content and length, the L2 stimulus sentences were modelled on the instrument used by Harrington & Sawyer (1992). The L2 sentences ranged from 11-13 words in length, and each level comprised three sets of increasing size. There were four levels in total, with the entire test comprising 42 stimulus sentences. Following Harrington & Sawyer (1992), all Spanish sentences ended in common nouns. Half of the German sentences ended in common nouns and the other half in common verbs. Default German word order in declarative sentences does not readily allow for constructing a large number of idiomatic and natural-sounding sentences ending in nouns; hence sentences ending in verbs were introduced.

The L2 span test was conducted in the same way as the L1 span test. In addition, however, participants were required to judge the correctness of each sentence after reading it aloud. Following Harrington & Sawyer (1992), this component was introduced to ensure that participants would attempt to process each sentence rather than simply try to memorize the last word. As it is possible to read out L2 sentences without processing their content, and as such a strategy is even more likely to be adopted in the case of languages with regular grapheme-phoneme mappings such as Spanish and German, the chosen test design was intended to ensure that both the storage and the processing component of working memory for L2 would be tapped. In each L2 span test, half of the stimulus sentences were correct and half were incorrect. Items were made ungrammatical by violating constituent order in the middle

part of the sentence; these violations were meant to be as obvious as possible. Grammatical and ungrammatical sentences were presented in random order.

Data collection and scoring procedures

Data were collected in two separate sessions. During the first session, the MLK test was administered under supervised conditions. The test was untimed, and papers were collected when all participants had finished. None of the participants took more than one hour to complete the test. Participants then filled in the biodata questionnaire.

During the second session, the MLAT was administered under supervised conditions. The commercially available CD was used to time participants; use of the CD also ensured consistent instructions and controlled delivery of the aural components of the test. The test is timed and, accordingly, was completed by all participants in one hour. After a break, participants took the L1 and L2 span tests. These tests were administered one-to-one, and learners had a further break between the two tests. The span tests were stopped when a participant failed an entire level. Depending on learners' performance, each span test took between 10 and 20 minutes.

The MLK tests were scored dichotomously in accordance with prepared answer keys. For each fully appropriate error correction one point was awarded. For each adequate description and for each adequate explanation one point was awarded, respectively. Adequate descriptions and explanations were defined as any descriptions and explanations that were not incorrect and that showed at least some evidence of meaningful generalization beyond the instances provided in the test items themselves. Sample items and targeted answers as well as sample participant answers and scores awarded can be found in Appendix C (German) and Appendix D (Spanish).

As scoring learners' descriptions and explanations involved qualitative judgements, answers were first scored by one of the researchers and then scored blind by a second marker. In the case of the German MLK test, interrater agreement was 95.6%; in the case of the Spanish MLK test, interrater agreement was 92.7%. Disagreements were resolved through discussion between the two markers.

For each correct answer on the language-analytic section, one point was awarded in accordance with the prepared answer key.

The MLAT was scored dichotomously in accordance with the answer key stencil provided in the commercially available test kit. For each correct answer, one point was awarded.

The L1 span test was scored following Daneman & Carpenter's (1980) scheme. A participant's reading span is calculated by counting the number of last words correctly recalled. If a participant successfully recalled the last words of all sentences in three or more of the five sets per level, they were awarded one point for that level. If they successfully recalled the last words of all sentences in two of the five sets per level, they were awarded half a point for that level. L1 reading span scores can thus range from 0 to a maximum of 6.0.

The L2 span test was scored in a similar manner, but the smaller number of sets per level was taken into account. If a participant successfully recalled the last words of all sentences in two or more of the three sets per level, they were awarded one point for that level. If they successfully recalled the last words of all sentences in one of the three sets per level, they were awarded half a point for that level. L2 reading span scores can thus range from 0 to a maximum of 4.0.

All statistical analyses were carried out with SPSS version 14.0. Scores on MLK test sections were converted into percentages whenever the slightly different numbers of items included in the German and Spanish versions had to be taken into account.

Results

RQ1 What is the relationship between university learners' metalinguistic knowledge, language learning aptitude, L1 working memory, and L2 working memory?

As a first step towards answering RQ1, correlations (Pearson's r) between participants' performance on the MLK test, the MLAT, and the span tests were calculated. The results are shown in Table 1.

Table 1: *Correlations between MLK test scores, MLAT scores, and span test scores*

	Descr./ expl.	Lang. analysis	MLAT total	MLAT 1	MLAT 2	MLAT 3	MLAT 4	MLAT 5	L1 span	L2 span
MLK test total	.98**	.89**	.34*	.15	.11	.29	.42**	.10	.19	.13
Descr./ expl.	1	.82**	.36*	.12	.11	.29	.45**	.06	.16	.16
Lang. analysis		1	.42**	.21	.13	.36*	.41*	.26	.28	.08

** significant at the .01 level (two-tailed); * significant at the .05 level (two-tailed)

With regard to the relationship between performance on the MLK test and the MLAT as a whole, a weak to moderate correlation is in evidence (r = .34). Moreover, both description/explanation ability (r = .36) and language-analytic ability (r = .42) are moderately correlated with overall MLAT performance, if inspected separately. The relationship between performance on the MLK test and the various MLAT subsections resulted in an unsurprising pattern: Both description/explanation ability (r = .45) and language-analytic ability (r = .41) are moderately correlated with performance on MLAT 4 (words in sentences), i.e. the test section on which the language-analytic element of the MLK test was modelled. There are no significant relationships in evidence between working memory for language and (components of) metalinguistic knowledge.

Two further significant relationships were uncovered (not shown in Table 1): Language learning aptitude as measured by overall MLAT performance and L1 reading span are moderately correlated (r = .40*), and the two measures of working memory for language are moderately correlated with each other (r = .41**).

As a second step towards answering RQ1, a principal components analysis was conducted. Table 1 shows that a reasonable number of significant correlations was in evidence. The suitability of the data set for a principal components analysis was further confirmed by calculating the Kaiser-Meyer-Olkin measure of sampling adequacy which, at .690, exceeded the recommended value of .6. (Pallant, 2005). Finally, Bartlett's test of sphericity clearly reached significance at $p < .001$. The principal components analysis was conducted in an exploratory mode, since the number of participants in the present study was relatively small.

The analysis included 11 variables, i.e. the subcomponents of metalinguistic knowledge as operationalized in the MLK test (correction, description, explanation, language analysis), the two measures of working memory for language (L1 reading span, L2 reading span), and the subcomponents of language learning aptitude as operationalized in MLAT 1, MLAT 2, MLAT 3, MLAT 4, and MLAT 5. The principal components analysis (varimax rotation) led to the extraction of four components with an eigenvalue > 1. Inspection of the screeplot confirmed that a four-component solution was indeed appropriate, with a clear break after the fourth component. The rotated component matrix is shown in Table 2.

Table 2: *Rotated four-component matrix for L2 metalinguistic knowledge, language learning aptitude, and working memory for language*

	Component 1	Component 2	Component 3	Component 4
Eigenvalue	4.137	1.791	1.196	1.144
Correction	**.912**	-.026	.003	-.034
Description	**.908**	.170	.223	.058
Explanation	**.949**	.133	.023	.041
Language analysis	**.842**	.132	.116	.295
MLAT 1	.049	**.557**	.365	-.011
MLAT 2	.003	**.839**	-.126	.134
MLAT 3	.204	.122	.237	**.747**
MLAT 4	.327	**.756**	.236	.108
MLAT 5	-.008	.067	-.025	**.885**
L1 span	.184	.070	**.783**	.214
L2 span	.018	.114	**.825**	-.004

Taken together, the four components explain 75% of the variance. Loadings that clearly cluster together on a specific component are highlighted in bold. Two components are easily interpretable: The subcomponents of metalinguistic knowledge clearly load on Component 1, which in itself explains nearly 38% of the variance. With equal clarity, the working memory measures load on Component 3, which explains nearly 11% of the variance. Somewhat more surprisingly, the subcomponents of aptitude as operationalized in the MLAT load on two separate components, rather than a single component or, alternatively, five separate components. MLAT 1 (number learning), MLAT 2 (phonetic script), and MLAT 4 (words in sentences) load on Component 2, which explains about 16% of the variance. MLAT

3 (spelling clues) and MLAT 5 (paired associates) load on Component 4, which explains about 10% of the variance.

RQ2 Which variables predict university learners' level of metalinguistic knowledge?

In order to address RQ2, a stepwise multiple regression analysis was conducted. Summed performance on the description/explanation section and the language-analytic section of the MLK test constituted the dependent variable. A total of 11 independent variables were entered, that is, performance on the MLAT as a whole, performance on each of the five MLAT subsections, performance on the two reading span tests, and three biodata variables (years of formal L2 study, weeks of L2 immersion, and cumulative years of study of other L2s). The results of the regression analysis are shown in Table 3.

Table 3: *Stepwise multiple regression coefficients, model summary, and significance (ANOVA), with L2 metalinguistic knowledge as the dependent variable*

Model	Independent variables	Standardized β coefficients	R	R^2	F	Sig.
1	Cumulative years of study of other L2s	.512	.512	.262	12.443	.001
2	Cumulative years of study of other L2s	.533				
	Years of formal L2 study	.431	.669	.447	13.769	< .001
3	Cumulative years of study of other L2s	.457				
	Years of formal L2 study	.407				
	MLAT 4	.316	.736	.541	12.976	< .001
4	Cumulative years of study of other L2s	.518				
	Years of formal L2 study	.426				
	MLAT 4	.280				
	MLAT 5	.241	.771	.595	11.758	< .001

The results show that, taken together, four of the independent variables entered (Model 4) account for a fairly impressive 60% of the variance in learners' level of metalinguistic knowledge (R^2 = .595). The significant predictor variables are cumulative years of study of other L2s, years of formal L2 study, performance on MLAT 4 (words in sentences), and performance on MLAT 5 (paired associates). Cumulative years of study of other L2s is the strongest individual predictor, accounting for 26% of the variance, followed by years of formal L2 study, which accounts for a further 19%. Performance on MLAT 4 explains a

further 9% of the variance, with performance on MLAT 5 accounting for the final 6%. In summary, two background variables and two measures of aptitude subcomponents strongly predict participants' level of metalinguistic knowledge.

Discussion

The results reported in the previous section provide insights into two dimensions of interest, that is, first, variables impacting on the development of metalinguistic knowledge in instructed university-level L2 learners of German and Spanish, and second, the status of metalinguistic knowledge in relation to the individual difference variables of language learning aptitude and working memory for language. Results are discussed in terms of these two dimensions.

Variables impacting on the development of metalinguistic knowledge

The multiple regression analysis (see Table 3) revealed that cumulative years of study of other L2s and years of formal study of the L2s under investigation jointly accounted for 45% of the variance in the participants' level of metalinguistic knowledge. In other words, length of exposure to form-focused language instruction in itself predicts to a considerable extent the quality and quantity of metalinguistic knowledge the instructed L2 learners participating in the current study developed.

All participants were university students enrolled in a language programme; they were all educated adults who had had ample exposure to form-focused L2 instruction, both in the context of the L2s under study and in the context of other foreign languages. The participants were thus experienced language learners who had progressed through the educational system successfully; given their interest in foreign languages, they had taken up a course of study that emphasized, valued, and indeed often necessitated the development of metalinguistic knowledge. Hence, the participant sample consisted of a selected group of learners who had proved themselves able to acquire a certain command of explicit L2 knowledge alongside their developing L2 proficiency (see also Roehr, 2008b).

In sum, our result seems to support the claim that the development of metalinguistic knowledge is influenced not only by learner-internal individual difference variables, as hypothesized previously (Collentine, 2000; DeKeyser, 2003, 2005), but also by learner-external variables such as prolonged exposure to formal L2 instruction in itself (see also Elder

& Manwaring, 2004; Elder et al., 1999; Renou, 2000; Sorace, 1985). Of course, more likely than not, these two sets of variables interact with one another (see below).

The status of metalinguistic knowledge in relation to language learning aptitude and working memory for language

The outcome of the principal components analysis (see Table 2) indicates that metalinguistic knowledge is a construct which is separate and distinguishable from both language learning aptitude and working memory for language. The independence of metalinguistic knowledge and working memory for language is particularly clear. Neither L1 nor L2 reading span correlated significantly with any of the metalinguistic measures (see Table 1). Moreover, in the principal components analysis, the metalinguistic measures and the measures of working memory for language loaded on two separate, clearly identifiable components. Whilst it is worth bearing in mind that the principal components analysis was conducted in an exploratory mode with data from a relatively small sample, the results obtained seem coherent and highly interpretable.

A possible reason for the lack of a significant relationship between metalinguistic knowledge and working memory for language in the present study is the type of measurement employed. The MLK test was not timed, whereas the L1 and L2 span tests obviously required online storage and processing of language, i.e. participants performed under time pressure. It is thus plausible to hypothesize that whilst the span tests taxed the participants' working memory resources, performance on the MLK test remained mostly unaffected by individual differences in working memory capacity. Future research using a timed measure of metalinguistic knowledge would therefore be desirable.

Our results further yield several suggestive insights into the status of metalinguistic knowledge in relation to language learning aptitude. The correlations between (components of) metalinguistic knowledge and language learning aptitude as measured by the MLAT as a whole were weak to moderate, with coefficients ranging from .34 to .42. It appears that the relationship can be attributed to a specific subtest, MLAT 4, which showed moderate significant correlations throughout ($r = .41$ to $r = .45$). This finding is perfectly consistent with the expectation that a subtest tapping the analytically based component of aptitude will be related to metalinguistic knowledge, operationalized as L2 description/explanation ability and L2 language-analytic ability.

While the correlational patterns obtained in the current study thus corroborate previous assumptions that (components of) language learning aptitude as measured by the MLAT are related to L2 metalinguistic knowledge (Alderson et al., 1997; Elder et al., 1999; Ranta, 2002; Roehr, 2008b), the factor-analytic results suggest that—at least in the case of the learners participating in the present study—metalinguistic knowledge can be regarded as a variable in its own right. The subcomponents of the complex construct of metalinguistic knowledge—correction ability, description ability, explanation ability, and language-analytic ability—clearly loaded on a single, separate component, thus setting these abilities apart from the various subcomponents of language learning aptitude.

Performance on MLAT 4 is the subcomponent of language learning aptitude that came closest to loading on the metalinguistic factor, as might be expected, but even this loading was very weak (see Table 2). Instead, performance on MLAT 4, which served as the template for the language-analytic section of the MLK test, loaded much more strongly on a different component than the analogous MLK test section. This result indicates that language-analytic ability for L1 and language-analytic ability for L2 can be distinguished.

Taken together, the results arising from the correlational and principal components analyses are compatible with the assumption that, to a certain extent, MLAT 4 and the MLK test draw on similar mental processes, that is, the explicit identification of categories and relations between categories. Moreover, both tests clearly deal with the same cognitive domain, namely, language. However, each test requires knowledge about a different language, since MLAT 4 is based on English (the participants' L1), whereas the MLK tests were based on either German or Spanish (the participants' L2s). Hence, it appears that L1 language-analytic ability and L2 metalinguistic knowledge are related, yet clearly distinguishable.

By contrast, no such distinction was in evidence for working memory for language; in the principal components analysis, L1 and L2 reading span clearly loaded on the same component. A possible interpretation of this finding is that a higher-level mental faculty such as analytic reasoning about language is more domain-specific than a lower-level and thus more generic mental faculty like online storage and processing of linguistic information, since the more generic faculty ultimately subserves the higher-level faculty. This explanation would be consistent with the proposal that skilled L2 users may draw on the same working memory resources during both L1 and L2 processing (Miyake & Friedman, 1998). As the participant sample of the present study did not include beginners, performance patterns averaged across the entire group may well have shown this convergence of L1 and L2 working memory

processes. As it would be inaccurate to describe every single participating learner as an advanced L2 user, however, correlations between L1 and L2 reading span were found to be moderate, thus indicating that the two abilities are not entirely convergent. A further possible factor which may have contributed to this lack of total convergence may be found in the differing task conditions: The L2 span test required participants to additionally make acceptability judgements.

The multiple regression analysis showed that in addition to the strongest predictors of metalinguistic knowledge—cumulative years of study of other L2s and years of formal study of the L2s under investigation—MLAT 4 and MLAT 5 were significant predictors as well, respectively accounting for 9% and 6% of the variance. Hence, L1 language-analytic ability (tapped by MLAT 4) and associative memory (tapped by MLAT 5), i.e. an analytic subcomponent of aptitude and a memory-based subcomponent of aptitude, had predictive value for the level of metalinguistic knowledge learners achieved. Indeed, MLAT 4 and MLAT 5 appear to cover the two key skills involved in attaining (aspects of) written L2 proficiency, that is, the modality strongly associated with metalinguistic knowledge (Alderson et al., 1997; Elder et al., 1999; Elder & Manwaring, 2004; Roehr, 2008b). Unlike the other three MLAT test sections, MLAT 4 and MLAT 5 incorporate no phonetic elements, respectively requiring the identification of the grammatical role of parts of speech in written English sentences and the learning of L1-L2 vocabulary pairings presented in written format.

A further result emerging from the principal components analysis deserves consideration, since it is of indirect relevance to the status of metalinguistic knowledge in relation to language learning aptitude and working memory for language. The two reading span measures and the various subcomponents of aptitude clearly loaded on distinct components, indicating not only that working memory for language and language learning aptitude do not overlap with metalinguistic knowledge, but also that they each constitute a separate construct. This finding appears to undermine the theoretical argument put forward by some researchers that working memory should be treated as an aspect of language learning aptitude (McLaughlin, 1995; Robinson, 2005; Sawyer & Ranta, 2001) or, conversely and more radically, that aptitude may be seen as a component of working memory (Miyake & Friedman, 1998). The results of the present study at least are not compatible with either account.

Conclusion

Taking into account the findings reported and discussed above as well as proposals put forward in prior research, we would like to conclude by hypothesizing that metalinguistic knowledge may be an individual difference variable in its own right, ranking among more established learner-internal factors such as general cognitive ability, language learning aptitude, cognitive and learning style, learner beliefs and attitudes, and language learning strategies (Dörnyei, 2005; Macaro, 2006). We acknowledge that this proposal is not new (see Kemp, 2001: 150), although, as far as we are aware, it has not yet found its way into the published literature.

The findings arising from the present study indicate that metalinguistic knowledge is separate and distinguishable from two neighbouring individual difference variables, that is, language learning aptitude and working memory for language. Moreover, previous empirical work has demonstrated that different individuals show different levels of L2 metalinguistic knowledge (Alderson et al., 1997; Butler, 2002; Elder & Manwaring, 2004; Elder et al., 1999; Roehr, 2006, 2008b). Existing research also suggests that increased language learning experience is associated with increased levels of metalinguistic knowledge (Jessner, 1999, 2006). Taken together with our finding that cumulative years of study of other L2s and years of formal study of the L2s under investigation strongly predicted participants' L2 metalinguistic knowledge, accounting for an impressive 45% of the variance, it appears that, to a certain extent at least, metalinguistic knowledge may be transferable across languages (Cummins, 1987).

L2 metalinguistic knowledge is acquired during an individual's lifetime. It is therefore different from learner variables that are believed to be enduring and thus relatively stable across the lifespan, such as language learning aptitude, general cognitive ability, or—to a lesser extent perhaps—cognitive and learning style (for a recent review, see Dörnyei, 2005). Hence, metalinguistic knowledge seems to be in the same league as language learning strategies, which are normally treated as an individual difference variable (Dörnyei, 2005; Macaro, 2006). Like strategies (Bruen, 2001; Purpura, 1999), metalinguistic knowledge is not only learnable, but also malleable and at least partly task-dependent (Alderson et al., 1997; Clapham, 2001). Both metalinguistic knowledge and language learning strategies can potentially be brought into awareness and articulated, with processes involving these kinds of knowledge drawing on the higher-level mental faculties of reasoning and analysis.

Both metalinguistic knowledge and strategies are problem-oriented, in that they tend to be used, respectively, to enhance L2 performance in particular (N. Ellis, 2005; R. Ellis, 2004) and to improve language learning more generally (Oxford, 1990, 2003). The acquisition of both metalinguistic knowledge and language learning strategies is partly predicted by related, more stable individual difference variables. Specifically, components of language learning aptitude may predict levels of metalinguistic knowledge, as the present study suggests (but see also DeKeyser, 2003, 2005), while cognitive and learning style have been shown to influence strategy use (Carson & Longhini, 2002; Littlemore, 2001). Despite these analogies, however, metalinguistic knowledge and language learning strategies are distinct constructs (Roehr, 2004).

The hypothesis that metalinguistic knowledge may be an individual difference variable in its own right should be put to the test in empirical research involving larger numbers of participants as well as different learner populations. Thus, it would be of interest to examine the status of metalinguistic knowledge in naturalistic L2 learners, i.e. learners who have not had extensive exposure to formal L2 instruction. Furthermore, it would be worth investigating metalinguistic knowledge in relation to individual difference variables other than language learning aptitude and working memory for language, such as cognitive or learning style, for instance. This would allow for a more detailed assessment of the independent status of metalinguistic knowledge hypothesized here.

Acknowledgements

The study reported in this paper was supported by the University of Essex Research Promotion Fund. We are grateful to Phil Scholfield for statistical advice. We would also like to thank Charles Alderson and Phil Scholfield for their constructive comments on an earlier version of this paper.

Notes

[1] While this is probably the most common understanding of metalinguistic knowledge in the adult L2 learning literature, work primarily concerned with metalinguistic development in children and/or bilinguals may assume slightly different definitions. For instance, Bialystok (1994a; 1994b; 2001; Bialystok & Ryan, 1985) does not equate metalinguistic knowledge with conscious awareness, and Birdsong (1989) does not regard conscious awareness as a defining characteristic of knowledge about language.

[2] It is accepted that the correction task does not necessarily tap metalinguistic knowledge, since it is clearly possible to correct a highlighted error spontaneously and intuitively. However, previous research suggests that correction is a step that naturally precedes the description/explanation of an error (Roehr, 2005). For this reason, the correction task was retained.

References

Alderson, J. C., Clapham, C., & Steel, D. (1997). Metalinguistic knowledge, language aptitude and language proficiency. *Language Teaching Research, 1*, 93-121.

Anderson, J. R. (2005). *Cognitive psychology and its implications* (6th ed.). New York, NY: Worth Publishers.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-423.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). Cambridge: Cambridge University Press.

Bialystok, E. (1979). Explicit and implicit judgements of L2 grammaticality. *Language Learning, 29*(1), 81-103.

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

Bialystok, E. (1994b). Representation and ways of knowing: Three issues in second language acquisition. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 549-570). London: Academic Press.

Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge: Cambridge University Press.

Bialystok, E., & Ryan, E. B. (1985). Toward a definition of metalinguistic skill. *Merrill-Palmer Quarterly, 31*(3), 229-251.

Birdsong, D. (1989). *Metalinguistic performance and interlinguistic competence*. Berlin: Springer.

Bruen, J. (2001). Strategies for success: Profiling the effective learner of German. *Foreign Language Annals, 34*(3), 216-225.

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), 451-480.

Camps, J. (2003). Concurrent and retrospective verbal reports as tools to better understand the role of attention in second language tasks. *International Journal of Applied Linguistics, 13*(2), 201-221.

Carroll, J. B. (1990). Cognitive abilities in foreign language aptitude: Then and now. In T. S. Parry & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 11-29). Englewood Cliffs, NJ: Prentice Hall.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.

Carroll, J. B., & Sapon, S. (2002). *Modern language aptitude test: MLAT*. N. Bethesda, MD: Second Language Testing Inc.

Carson, J. G., & Longhini, A. (2002). Focusing on learning styles and strategies: A diary study in an immersion setting. *Language Learning, 52*(2), 401-438.

Clapham, C. (2001). The assessment of metalinguistic knowledge. In C. Elder & A. Brown & E. Grove & K. Hill & N. Iwashita & T. Lumley & T. Macnamara & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 31-41). Cambridge: Cambridge University Press.

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

Cummins, J. (1987). Bilingualism, language proficiency, and metalinguistic development. In P. Homel & M. Palij & D. Aaronson (Eds.), *Childhood bilingualism: Aspects of linguistic, cognitive, and social development* (pp. 57-73). Hillsdale, NJ: Erlbaum.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 313-348). Malden, MA: Blackwell.

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

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Erlbaum.

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589-630). Malden, MA: Blackwell.

Ehrman, M. E., & Oxford, R. L. (1995). Cognition plus: Correlates of language learning success. *Modern Language Journal, 79*(1), 67-89.

Elder, C., & Manwaring, D. (2004). The relationship between metalinguistic knowledge and learning outcomes among undergraduate students of Chinese. *Language Awareness, 13*(3), 145-162.

Elder, C., Warren, J., Hajek, J., Manwaring, D., & Davies, A. (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), 305-352.

Ellis, N. C., & Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology, 49A*(1), 134-250.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning, 54*(2), 227-275.

Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research, 9*(2), 147-171.

Gass, S. M., & Alvarez Torres, M. J. (2005). Attention when? An investigation of the ordering effect of input and interaction. *Studies in Second Language Acquisition, 27*(1), 1-31.

Gathercole, S. E., & Thorn, A. S. C. (1998). Phonological short-term memory and foreign language learning. In A. F. Healy & L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 141-158). Mahwah, NJ: Erlbaum.

Green, P. S., & Hecht, K. (1992). Implicit and explicit grammar: An empirical study. *Applied Linguistics, 13*(2), 168-184.

Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition, 19*, 379-400.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14*, 25-38.

Hu, G. (2002). Psychological constraints on the utility of metalinguistic knowledge in second language production. *Studies in Second Language Acquisition, 24*(3), 347-386.

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), 129-140.

Jessner, U. (1999). Metalinguistic awareness in multilinguals: Cognitive aspects of third language learning. *Language Awareness, 8*(3&4), 201-209.

Jessner, U. (2006). *Linguistic awareness in multilinguals: English as a third language*. Edinburgh: Edinburgh University Press.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122-149.

Kemp, C. (2001). *Metalinguistic awareness in multilinguals: Implicit and explicit grammatical awareness and its relationship with language experience and language attainment.* Unpublished PhD thesis, University of Edinburgh.

Klapper, J., & Rees, J. (2003). Reviewing the case for explicit grammar instruction in the university foreign language learning context. *Language Teaching Research, 7*(3), 285-314.

Leow, R. P. (2000). A study of the role of awareness in foreign language behavior: Aware versus unaware learners. *Studies in Second Language Acquisition, 22*(4), 557-584.

Littlemore, J. (2001). An empirical study of the relationship between cognitive style and the use of communication strategy. *Applied Linguistics, 22*(2), 241-265.

Macaro, E. (2006). Strategies for language learning and for language use: Revising the theoretical framework. *Modern Language Journal, 90*(3), 320-337.

Macaro, E., & Masterman, L. (2006). Does intensive explicit grammar instruction make all the difference? *Language Teaching Research, 10*(3), 297-327.

Mackey, A., Philp, J., Egi, T., & Fujii, A. (2002). Individual differences in working memory, noticing of interactional feedback and L2 development. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 181-209). Amsterdam: John Benjamins.

McLaughlin, B. (1995). Aptitude from an information-processing perspective. *Language Testing, 12*, 370-387.

Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy & L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339-364). Mahwah, NJ: Erlbaum.

Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. Boston, MA: Heinle & Heinle.

Oxford, R. L. (2003). Language learning styles and strategies: Concepts and relationships. *IRAL, 41*(4), 271-278.

Pallant, J. (2005). *SPSS survival manual* (2nd ed.). Maidenhead: Open University Press.

Purpura, J. E. (1999). *Learner strategy use and performance on language tests: A structural equation modeling approach*. Cambridge: Cambridge University Press.

Ranta, L. (2002). The role of learners' language analytic ability in the communicative classroom. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 159-180). Amsterdam: John Benjamins.

Renou, J. M. (2000). Learner accuracy and learner performance: The quest for a link. *Foreign Language Annals, 33*(2), 168-180.

Robinson, P. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition, 19*, 223-247.

Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research, 17*(4), 368-392.

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics, 25*, 46-73.

Roehr, K. (2004). Exploring the role of explicit knowledge in adult second language learning: Language proficiency, pedagogical grammar and language learning strategies. *Lancaster University CRILE Working Paper, 59*.

Roehr, K. (2005). *Metalinguistic knowledge in second language learning: An emergentist perspective.* Unpublished PhD thesis, Lancaster University.

Roehr, K. (2006). Metalinguistic knowledge in L2 task performance: A verbal protocol analysis. *Language Awareness, 15*(3), 180-198.

Roehr, K. (2008a). Linguistic and metalinguistic categories in second language learning. *Cognitive Linguistics, 19*(1), 67-106.

Roehr, K. (2008b). Metalinguistic knowledge and language ability in university-level L2 learners. *Applied Linguistics, 29*(2), 173-199.

Rosa, E., & O'Neill, M. D. (1999). Explicitness, intake, and the issue of awareness. *Studies in Second Language Acquisition, 21*, 511-556.

Sanz, C., & Morgan-Short, K. (2004). Positive evidence versus explicit rule presentation and explicit negative feedback: A computer-assisted study. *Language Learning, 54*(1), 35-78.

Sasaki, M. (1996). *Second language proficiency, foreign language aptitude, and intelligence*. New York: Peter Lang.

Sawyer, M., & Ranta, L. (2001). Aptitude, individual differences, and instructional design. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 319-353). Cambridge: Cambridge University Press.

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

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

Shah, P., & Miyake, A. (1999). Models of working memory: An introduction. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 1-27). Cambridge: Cambridge University Press.

Skehan, P. (1986). Cluster analysis and the identification of learner types. In V. Cook (Ed.), *Experimental approaches to second language learning* (pp. 81-94). Oxford: Pergamon.

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* (pp. 69-94). Amsterdam: John Benjamins.

Sorace, A. (1985). Metalinguistic knowledge and language use in acquisition-poor environments. *Applied Linguistics, 6*(3), 239-254.

Sparks, R., & Ganschow, L. (2001). Aptitude for learning a foreign language. *Annual Review of Applied Linguistics, 21*, 90-111.

Williams, J. N., & Lovatt, P. (2003). Phonological memory and rule learning. *Language Learning, 53*(1), 67-121.

Appendix A

L2 features targeted by the German MLK test

Test Section 1: Correction, Description, Explanation

Morphosyntactic features

Agreement: Present tense verb endings

Use of personal pronouns

Use of case: Direct object in the accusative

Negation: *nicht* vs. *kein*

Perfect tense with *haben* vs. *sein*

Separable verbs

Use of case: Prepositions with dative

Use of case: Prepositions with accusative or dative: *Wechselpräpositionen*

Word order/position of the verb in subordinate clauses/subordinating conjunctions

Use of case: Adjectival inflection

Use of case: Genitive (possession)

Use of tense/mood/voice: Past subjunctive (*Konjunktiv II*)

Adverbials of place and direction

Use of tense/mood/voice: *seit* and present tense

Use of modal verbs: Word order/position of the verbs in the main clause of declarative sentences

Lexico-semantic and pragmatic features

Collocations/idiomatic use of the L2: *mit dem Bus fahren* vs. *go by bus*

Idiomatic use of tense/mood/voice of modal verbs: *sollen* vs. *sollte*

Formal vs. informal form of address/register

Use of fixed phrases: Tipping in a restaurant

Politeness: Making excuses

Test Section 2: Language Analysis

Coordinating conjunction

Direct object in the accusative case

Indirect object in the dative case

Verb of the main clause in the simple past/*Präteritum*

Subject of the main clause

Um introducing infinitive construction with *zu*/*erweiterter Infinitiv mit zu*

Verb used as a noun/*substantiviertes Verb*

Present participle of the verb used as an adjective

Numeral used as pronoun/subject of the main clause

Subordinating conjunction

Non-finite verb in a compound tense: Passive construction with *werden*

Preposition

Predicatively used adjective

Adverb modifying an adjective

Dependent infinitive without *zu*

Adverb of manner

Finite verb in a compound tense: Auxiliary in the present perfect

Object in the nominative case

Relative pronoun in the accusative case

Non-finite verb in a compound tense: Past participle in a passive construction with *werden*

Indefinite article introducing a prepositional object in the accusative case

Particle of a separable verb

Question word used as a relative pronoun

Attributively used adjective

Genitive case

Appendix B

L2 features targeted by the Spanish MLK test

Test Section 1: Correction, Description, Explanation

Morphosyntactic features

Use of personal pronouns

Comparative of adjectives

Radical changing verbs

Use of reflexive pronouns

Agreement: Present tense verb endings

Imperative mood

Conditional tense

Preterite tense

Back-to-front verbs

Verbs followed by the infinitive

Ser and *estar*

Para and *por*

Personal *a*

Perfect tense to indicate length of an action: *llevar* and gerund

Subjunctive after statements of possibility/probability

Lexico-semantic and pragmatic features

Collocations/idiomatic use of the L2: *pedir la cuenta*

Collocations/idiomatic use of the L2: *caer bien*

Collocations/idiomatic use of the L2: *hacer frío/calor*

Use of fixed phrases: Christmas and New Year wishes

Formal vs. informal form of address/register

Test Section 2: Language Analysis

Coordinating conjunction

Direct object

Indirect object

Verb of the main clause in the simple past/preterite

Subject of the main clause

A introducing an infinitive

Verb used as a noun

Present participle of the verb used as an adjective

Prepositional object

Numeral used as an indefinite pronoun/subject of the main clause

Uses of the passive

Verb *to be*: Copulative *ser*

Preposition: *para*

Reflexive pronouns

Finite verb in a compound tense: Auxiliary in the present perfect

Modal auxiliaries

Que before a verb in the subjunctive mood

Use of fixed phrases: Saying good night

Conditional conjunctions

Adverb of manner

Definite article

Possessive adjectives

The infinitive to express warnings and instructions

Question word used as a relative pronoun

Comparative of adjectives

Appendix C

Sample items and targeted answers from the German MLK test (Section 1), sample participant answers, and scores

Item 3

Die kleine Martina hat gestern zum ersten Mal geflogen. Sie ist ganz begeistert und freut sich schon auf den Rückflug.

Gloss:	Yesterday little Martina travelled by plane for the first time. She loved it and she is already looking forward to the flight home.
Correction:	*ist*
Description/explanation:	As the verb *fliegen* expresses (directional) movement and is not accompanied by a direct object, it requires a form of *sein* in the perfect tense.

Participant 12

Correction:	ist
Description/explanation:	With the verb ending the sentence being 'geflogen', which is in the past tense, the appropriate verb haben or sein needs to be used. 'Hat' has been given, but it is the verb sein that should be used and the third person singular of this verb: ist
Scores awarded:	C = 1; D = 1; E = 0

Participant 18

Correction:	ist
Description/explanation:	the past participle geflogen requires sein as the auxiliary in the past tense as fliegen is a verb referring to motion and therefore cannot be used with haben to create the perfect tense
Scores awarded:	C = 1; D = 1; E = 1

Item 4

Ich habe leider nicht Geschwister.

Gloss:	Unfortunately I have no brothers or sisters.

Correction: *keine*

Description/explanation: As a noun (object) with (an indefinite or) no article is being negated, a form of *kein* is required rather than *nicht*.

Participant 7

Correction: nichts

Description/explanation: It's not negating an idea, so it should be 'nichts' instead of 'nicht'

Scores awarded: C = 0; D = 0; E = 0

Participant 17

Correction: keinen

Description/explanation: Use kein rather than nicht to negate a noun

Scores awarded: C = 0; D = 1; E = 1

Participant 18

Correction: keine

Description/explanation: Nicht can be used to negate verbs, but kein must be used to negate nouns

Scores awarded: C = 1; D = 1; E = 1

Item 5

Möchtest du der Apfelkuchen oder die Sahnetorte?

Gloss: Would you like the apple pie or the gateau?

Correction: *den*

Description/explanation: As the definite article is part of the direct object of the sentence, the accusative case needs to be used.

Participant 5

Correction: die or das

Description/explanation: der is the wrong case for that word

Scores awarded: C = 0; D = 1; E = 0

Participant 16

Correction:	den
Description/explanation:	In this sentence Apfelkuchen is a direct object of the verb, and as such it should be in the accusative case. The definite article of Apfelkuchen is 'der' (masculine) which in the accusative case changes to 'den'
Scores awarded:	C = 1; D = 1; E = 1

Appendix D

Sample items and targeted answers from the Spanish MLK test (Section 1), sample participant answers, and scores

Item 10

Mi hermano juga fútbol los domingos.

Gloss:	My brother plays soccer on Sundays.
Correction:	*juega*
Description/explanation:	As *jugar* is a radical-changing verb which contains the vowel *u*, the *u* changes to *ue* when stressed.

Participant 24

Correction:	juega
Description/explanation:	Jugar is an irregular verb
Scores awarded:	C = 1; D = 0; E = 1

Participant 34

Correction:	juega
Description/explanation:	As 'jugar' is a radical-changing verb in Spanish, it must be conjugated by using '-ue' in place of 'u' in the 1st, 2nd, 3rd person singular and the 3rd person plural in the present tense. 'Mi hermano' is referred to in the third person, therefore 'juega' is the correct conjugation
Scores awarded:	C = 1; D = 1; E = 1

Participant 36

Correction:	juega
Description/explanation:	Inserting "ue" makes the 'j' in jugar 'hard'
Scores awarded:	C = 1; D = 1; E = 0

Item 12

La posibilidad de que hubo un cambio político en México provocó la intervención Estadounidense.

Gloss:	The possibility of a political change in Mexico provoked the American intervention.
Correction:	*hubiera/hubiese*
Description/explanation:	As the subordinate clause expresses possibility, the pattern possibility + *que* + subordinate verb in the subjunctive (imperfect in this case) is required.

Participant 22

Correction:	habría
Description/explanation:	Conditional is required as it is a possibility eg 'could'
Scores awarded:	C = 0; D = 0; E = 1

Participant 28

Correction:	? [The participant actually put a question mark.]
Description/explanation:	Subjunctive may be used here?
Scores awarded:	C = 0; D = 1; E = 0

Participant 30

Correction:	hubiera
Description/explanation:	As this sentence is stating possibility/uncertainty it requires the subjunctive tense rather than the indicative tense. As the action states is in the past i.e. provocó, you must use the imperfect subjunctive to agree with this
Scores awarded:	C = 1; D = 1; E = 1