

# Surveys of Ph.D./Ed.D. Theses

## Review of doctoral research in second-language teaching and learning in England (2006)

**Emma Marsden** University of York, UK  
em502@york.ac.uk

**Suzanne Graham** University of Reading, UK  
s.j.graham@reading.ac.uk

*Using the British 'Index to Theses', we found forty-seven Ph.D.s relating to second and foreign language learning and/or teaching defended in English universities in 2006. Objective criteria led us to fourteen theses which had investigated both teaching and learning. Over half of these adopted a process-product research design with the aim of finding causal relationships between teaching and learning. Six theses focused on individual differences (motivation, strategies, attitudes), with three adopting an 'effectiveness-of-intervention' approach and three following more descriptive, exploratory designs. The designs of the 'effectiveness-of-intervention' studies varied greatly, ranging from naturalistic evaluations to highly controlled randomised control experiments. They covered a range of pedagogical concerns, including the use of computers, error correction, language portfolios, learner strategies and communicative-style activities. In addition to our own comments on the quality of the studies and reports, we present considerable methodological detail to enable the reader to evaluate the validity of the findings and claims made in each study. We argue that Ph.D. theses need to demonstrate fully that the implications drawn from the study are supported by the data collection and analyses described, which was not always the case in the theses reviewed. Finally, we make suggestions for future areas of investigation by postgraduate researchers.*

### 1. Introduction

For this review, we wanted to reduce selection bias as far as possible, to avoid, for example, favouring particular institutions or omitting relevant theses. This was our primary aim when developing selection procedures to reduce the number of theses to a feasible number for the purposes of this paper. The first section of the paper outlines how we tried to achieve this, and provides brief overviews of the broad areas which were studied at doctoral level but which we could not include in our more in-depth reviews. The second and third sections constitute more in-depth reviews of two (overlapping) groups of theses which investigated both second

or foreign (henceforth L2) language teaching AND learning: one group of ‘process–product’ studies, and one group of studies looking at ‘individual differences’ (some of which also took a ‘process–product’ approach). In the final section we briefly consider how theory has been used in the studies, and comment on our review process itself.

As our main aim was to be as objective and systematic as possible in our selection, we do not claim to have chosen the ‘most significant’ theses in terms of our own (subjective) judgement of the magnitude or quality of their potential contribution to the field of L2 teaching and learning. Rather, our selection focused on providing an overview of the nature and quality of doctoral studies which investigated relationships between L2 teaching and learning. The studies are all significant in that they give us insight into how these relationships were conceptualised and explored. Taken together, they illustrate some of the methodological, theoretical and ethical challenges faced when looking for useful causal relationships between teaching and learning.

## 2. Selection criteria and overview of the field

We carried out the first search using the British Library’s ‘Index to Theses’ (<http://www.theses.com>). The British Library is notified of the final submission of theses, and so this was an objective and systematic way of accessing an initial sample. We used just two key terms: ‘Language’ and ‘2006’.<sup>1</sup> These very broad terms were necessary, as more specific search criteria excluded some relevant theses.

This search revealed 442 records, including theses relating to L1 learning, literature, computational linguistics, historical linguistics, genetics, and philosophy. A manual search of these 442 abstracts was then carried out to retain all those related to L2 language learning and/or teaching. We excluded all Ed.D and MSc. theses, and all those not submitted in England (we found four Ph.D.s from Scotland that would have fitted our other criteria). This second search left a pool of 47 theses all related to L2 learning and teaching, though not necessarily both. These 47 theses seemed to fall into three categories: those broadly related to L2 and education (though not specifically related to learning an L2), those broadly related to L2 acquisition and use<sup>2</sup> (though not specifically linked to classroom instruction), and those which involved some study of both L2 teaching and language learning. Here we give a broad overview of the first two groups before providing a more in-depth review of the third group.

There were fifteen theses which were broadly related to language education in general. Language LEARNING was not a primary focus of the design, data collection or analysis. This group included six theses relating to language teacher education and teacher cognition, which space constraints do not allow us to reference. In addition, Thepsiri’s (2006) thesis documented teachers’ use of and attitudes towards scaffolding for different teaching functions during project work in a university in Taiwan, and Cheng’s (2006) study documented teachers’ views about and use of motivational techniques for L2 teaching. Liang (2006) described

<sup>1</sup> We found only four records from 2007. We did not include these as we could not be sure that this truly represented the area that year, or whether there was a delay in uploading records to the database.

<sup>2</sup> The terms acquisition and learning are used inter-changeably in this article.

vocabulary teaching practices in primary EFL (English as a Foreign Language) classes in Taiwan. Another thesis explored whether cultural background influenced use of Information and Communications Technology in L2 classrooms (Field 2006). Brereton (2006) examined whether using sign language to hearing, preschool children had any effects on the pupils' involvement in lessons. Manners (2006) investigated relationships between Specific Language Impairment and English as an Additional Language. One thesis (Ward 2006) was a textbook analysis, scrutinising the lexicon of undergraduate engineering textbooks. Another study addressed language planning (Coluzzi 2006). Finally, there was one study which looked at perceptions of a language course (Bin Yusot 2006).

A second set of fourteen studies related mainly to L2 learning and use. The L2 classroom or formal teaching was not an explicit design feature of the empirical study. Within this group, there were five studies which assumed a generativist (Universal Grammar) explanation for L2 acquisition phenomena. Four others looked at L2 acquisition from other linguistic perspectives: an analysis of native and non-native verb use in writing (Guo 2006), the use of prefabricated language in EAP articles (McKenny 2006), and the pragmatics of EFL refusals (Kasemsin 2006). One focussed on the effects of bilingualism on cognition (Athanasopoulos 2006). Five other theses were more sociolinguistically oriented, using ethnographic techniques within immigrant, migrant or minority communities: Goglia (2006) examined code-switching among immigrants in Italy, Souza (2006) looked at ethnic and social identity and language choice among a Brazilian speaking community in London, Carlson (2006) examined and evaluated the implementation of a home-based literacy programme for bilinguals, and Edwards (2006) looked at variables which influence use of L2 English and willingness to communicate among Korean learners of English outside the classroom. Mertin (2006) investigated differing expectations about teaching and learning among the social, personal and academic communities linked to a group of Japanese EFL learners at an international high school.

To reduce the number of theses to a manageable size for an in-depth review, two further selection criteria were used. One was to exclude three theses relating exclusively to language testing (Lonsdale 2006; Lu 2006; Tonkyn 2006). The other criterion was to exclude studies which had an individual action research approach. Ting (2006) fitted this criterion, looking at the teacher-researcher's own use of internet resources and their understanding and use of theories of multiple intelligences.

This left fourteen theses for more in-depth review. These included, as part of the design of their empirical study, an investigation of both L2 teaching and learning. We categorised them according to two broad criteria. One group contained eight 'process-product' studies which investigated causal relationships between teaching and learning to evaluate the effectiveness of interventions. One of these was unavailable (Lee 2006).<sup>3</sup> The other group brought together studies investigating individual differences, and included six theses relating to motivation, attitudes and learner strategies. The nature of these categorising principles was clearly different, one being methodological ('process-product') and one substantive ('individual differences'). Inevitably, these categories overlapped, as Wiliwan (2006) and Walters (2006) conducted intervention studies looking at learner strategy use, and Bavendiek (2006) looked

<sup>3</sup> We could not obtain the thesis via inter-library loans or from the author.

at the effect of a language portfolio programme on learner motivation. For the purposes of structuring this paper, these three were reviewed along with the individual differences theses.

### 3. Process–product studies

Seven ‘effectiveness of intervention’ theses are reviewed here. One was a naturalistic evaluation of current practice (Al-Ghamdi 2006). The other six developed interventions specifically for the research. Two of these incorporated an experimental intervention into ongoing practice (Hill 2006; Tuncel 2006); the remaining four set up experimental and comparison interventions for the purposes of the study. Two of these evaluated the effectiveness of specific interventions using computer technology (Ciobanu 2006; Niño 2006); the other two tested the effectiveness of specific classroom teaching techniques and materials (Sheen 2006; Skoufaki 2006).

Al-Ghamdi (2006) carried out an ambitious evaluation of an English for Specific Purposes (ESP) programme for 190 male first-year students on medicine and medical science degrees at a Saudi university. He evaluated whether the ESP course achieved its objectives, whether its content and methodology were appropriate to its objectives, and whether its stakeholders were satisfied with it. The course was taught and tested in six discrete components: grammar, reading, writing, lecture and note-taking, medical vocabulary, and communication skills.

Al-Ghamdi critically appraises approaches to evaluating language teaching, concluding that mixed methods are required to look at the product and the process, and his study focused on evaluating the process for formative purposes. Al-Ghamdi’s evaluation procedure used: (i) achievement data from purpose-designed pre- and post-English language tests, and in-house tests from the language centre and the medical faculties; (ii) real time observation notes from 35 hours of lessons, with data coded according to four, *a priori*, thematic categories: students’ motivation and participation, mother tongue use, students’ involvement with materials, and classroom management; (iii) four different questionnaires to four groups of stakeholders (fifteen subject (i.e. non-language) teachers, six language teachers, four administrators, and a midway and end-of-course questionnaire to 80 students) regarding attitudes, motivation, views and perceived needs regarding the course delivery, structure and resources; (iv) an analysis of the course textbook, syllabus specifications and curriculum objectives.

The pre–post-tests included items from IELTS (International English Language Testing System) academic tests (selected for their relevance to medicine), and a TOEFL (Test of English as a Foreign Language) structure and written expressions test. Two randomly selected intact classes (25% of the students) took the pre- and post-tests. Twelve students also took speaking tests, but these were then, unfortunately, excluded from analysis as Al-Ghamdi considered the sample too small. A medium positive correlation ( $r = .542$ ) was found between Al-Ghamdi’s tests and the in-house tests. Al Ghamdi (perhaps over-)confidently interpreted this as evidence that ‘the compiled proficiency test is a valid and appropriate means to evaluate the effectiveness of the programme’ (p. 114).

The pre- and post-tests showed statistically significant improvement from the start to the end of the ESP course in listening, reading, grammar and writing. Al-Ghamdi used data from the questionnaires and interviews to argue, reasonably, that the course was responsible for

these gains. A very high positive correlation ( $r^2 = .74$ ) was found between the in-house English achievement tests and the students' biochemistry and physiology scores. Al-Ghamdi suggests that this demonstrates the validity of the language centre's assessment procedures. However, this argument assumes that the medical scores are themselves valid. The possibilities that subject knowledge was causally related to language proficiency, or that both were related to another causative factor, were not discussed. We found the reporting of the statistics difficult to follow in places (e.g. sample size was equated with variance (p. 116), effect size was equated with the amount of variability explained (p. 119), and different tests were used on the same data without appropriate justification).

Most stakeholders said they would have preferred an intensive, course-prerequisite programme, and most, except the students, thought that greater priority should be given to reading than to lecture note taking and grammar. The dense document analysis and observation data did not result in a positive evaluation of the lecture note-taking classes. Al-Ghamdi used these, and other, findings to challenge the rationale behind component-based language teaching.

The subject teachers were divided equally between considering the course effective, ineffective and not knowing, whereas the other three groups of stakeholders all rated the effectiveness of the course more favourably. As this was a small case study, it is difficult to ascertain what this means for the language programme compared to stakeholders' opinions about other language programmes.

Al-Ghamdi's recommendations mainly focus on macro-delivery issues, such as the timing and duration of the course, integrating language skills to make two large components of 'literacy' and 'fluency', greater collaboration between medical and language staff about students' needs, improving staff workloads and training, and redesigning the syllabus specifications. He also suggests some changes in the types and use of textbooks, students' attendance and time management in the classroom. On the whole the recommendations are linked to the empirical findings, although a few are less clearly warranted by his data.

Four pages (pp. 222–226) discuss the evaluation methods themselves. Al-Ghamdi concludes that the programme is effective (p. 122), though his design does not investigate effectiveness as a relative construct. For us, the study raises questions as to whether such an approach to evaluation can isolate cause and effect relationships between perceived weaknesses and specific problems. As Al-Ghamdi himself points out 'the pre-experimental method [the one group pre- and post-test] cannot tell if the effect is necessarily caused by the programme itself' (p. 9). Nor can the pre-experimental model tell us whether the programme was as effective as another course. Al-Ghamdi sensibly recommends a follow-up experiment, using the current study as a baseline for comparison, and an experimental intervention incorporating the recommendations from his study. Standardised assessments specifically for medical English might also improve the generalisability of such case study evaluations. We were unable to contact Al-Ghamdi to see whether this case study has changed local practice.

The next two theses reviewed incorporated a specifically designed experimental intervention into on-going courses. In the first two-thirds of his thesis, Hill (2006) provides a (rather unwieldy) overview of many theories of grammar, language, cognition and learning. He ambitiously brings these different domains and perspectives to bear on a specific pedagogical

issue: how to teach English aspect and tense to Japanese learners. Hill claims to have evaluated various theories of grammar in terms of how well they describe aspect and tense, and how well they could account for its acquisition. His main thesis is that learners may benefit from being told, explicitly, about how tense and aspect are realised via lexical semantics, following the Aspect Hypothesis (Andersen & Shirai 1994). In addition, Hill incorporated the notions of using cognitive schemata, metaphor, metonymy, the Zone of Proximal Development and other regulation (Vygotsky 1978) in his conceptualisation of how tense and aspect might be taught.

The thesis then presents a wide range of research questions and a series of empirical studies which Hill claimed addressed them. The first empirical study was a questionnaire given to fourteen English teachers at a university in Japan, eliciting teachers' thinking about the teaching of grammar, and of tense and aspect in particular. One interesting finding to emerge from this was that a high proportion of teachers responded 'don't know' to questions about the use of metaphor for teaching grammar, suggesting that the approach suggested by Hill was indeed innovative.

The second and third empirical studies share some characteristics with classroom experiments, though methodological discussion was limited (at least in the electronic version Hill kindly provided). Hill reports that these studies were, essentially, action research, but with broadly defined 'control' groups. He used two intact groups at his university from two different faculties: one intact group for the treatment ( $n = 16$ ) and one for the control ( $n = 11$ ). The materials for the treatment group were a mix of written and structured oral pair-work exercises, lasting for two 90-minute sessions over about eight weeks. They include activities which operationalised a learner-friendly version of some features of the Aspect Hypothesis, alongside some rule and practice activities, such as gap filling, which were in fact very traditional, and their links to the literature reviewed was not always clear. One of the refreshing features of this study was that the intervention (though, unfortunately, not that of the control group) was fully documented.

Hill gave the subjects a pre- and post-test, in line with normal experimental practice. According to this, the control group was at a lower proficiency level before the treatment began. He also gave the learners a 'traditional' grammar test, but this was only administered as a post-test, so comparisons between groups were unreliable as it was not known how the groups would have performed prior to the intervention. Despite these two issues, Hill seemed confident in concluding that the specific nature of his teaching materials brought benefits to the experimental group, evidenced by learning gains on the 25-item post-test.

The third empirical study used a sample similar to the second, with a 90-minute intervention focussing on future tense and aspect, and then another 90-minute intervention on present perfect tense and aspect. Hill used semi-controlled sentence level oral tests to assess learning gains,<sup>4</sup> and ANOVAs and *t*-tests to compare some of his results, though more consideration might have been given to how appropriate these statistical procedures were with small non-randomised samples which were not matched at baseline.

<sup>4</sup> It was not clear from the thesis sent to us, and the author could not be contacted to clarify, whether there were also written tests.

Hill acknowledges some of the weaknesses of his study, such as the fact that the experimental group may have performed well because they received extra attention from Hill himself (a possible Hawthorne effect<sup>5</sup>), and towards the end of the thesis discusses some ethical concerns of his research design. His empirical studies illustrate some of the challenges of conducting rigorous classroom experiments, but, perhaps most usefully, his thesis demonstrates an attempt at bringing together socio-cultural and cognitive perspectives on language and learning, probably reflecting a recent trend in the applied linguistics community towards exploring these interfaces.

Tuncel (2006) aimed to search for ways to improve the effectiveness of English for Academic Purposes programmes run by private institutions who prepare students for study at English-medium universities in Northern Cyprus. He also aimed to address the lack of experiments evaluating the usefulness of incorporating communicative elements into a curriculum. His theoretical and empirical backdrop was the work done in the early 1980s on defining and discussing communicative language competence and teaching (CLT).

Tuncel first administered four questionnaires to preparation school teachers, test designers, university lecturers and students, totalling 971 respondents. Key findings that influenced the subsequent experimental study were: the lecturers and students thought that proficiency in speaking was the weakest skill area; the school teachers were not satisfied with student progression; most agreed that the students' proficiency varied between low-elementary and average pre-intermediate, and this was deemed inadequate by Tuncel for the purposes of studying at an English-medium university. The testing officers thought the tests mainly assessed the explicit grammar content of the course.

One preparation school was then selected to host a cluster-randomised experiment, with participants aged between 18 and 25 years. Two intact classes ( $n = 38$ ) were randomly selected to be the experimental groups, leaving five intact control classes ( $n = 114$ ). In the experimental group, the official syllabus was supplemented with teaching using CLT approaches, including 'communicative and authentic' (p. 17) materials, realistic contexts and language-discovery techniques. The sample activities provided suggest that the CLT activities focused on lexical features such as nouns, verb stems, and adjectives (including numbers) and sentential comprehension. The control groups followed the normal syllabus, focussing explicitly and proactively on grammatical structures through mechanical practice. Although Tuncel claimed that the GRAMMAR of the official syllabus had been incorporated into more communicative-style activities, it was not clear to us from the samples provided how this incorporation was achieved. The empirical work which has tried to address this major challenge (e.g. Doughty & Williams 1998) is not discussed in the thesis.

To accommodate the experimental teaching, the official material was completed 'more quickly' (p. 215), leaving 20% of the time for the experimental activities. The teaching ran for 25 hours a week, over nine months, amounting to about 700 hours with twenty staff teaching the seven control groups. Tuncel himself taught both experimental groups for ten out of the 25 hours per week. The remaining fifteen hours for each experimental group were taught by two teachers who also taught some of the control groups. Unlike some of the other theses reviewed, Tuncel does address the extent of disclosure to the participants; however,

<sup>5</sup> The effects that researchers themselves can have on the phenomena they are investigating.



somewhat controversially, the students were not told that they were in a research study until the end of the project.

Tuncel used the school's in-house tests in grammar, reading, listening, speaking and writing to monitor learning gains during the course finding, through *t*-tests and one-way ANOVAs, that there were no statistically significant differences at the pre-test between the seven classes in any of the five tests. At the post-test, statistically significant differences (of about 10%) were found between the control and the experimental groups in all the measures. These findings are broadly compatible with early CLT experimental work (e.g. Spada 1987). Tuncel also found that one year after they had left the school, the (former) experimental group had statistically significantly higher first-year scores on their undergraduate course than the (former) control group.

Tuncel notes potentially confounding factors, such as students' motivation or features of the teaching (although, unfortunately, no records of the control or experimental teaching were kept). Nevertheless, he argues that 'there is a direct relationship between ... the programme [and] the improvement of the proficiency level of the learners' (p. 307). However, we could not find an acknowledgement that Tuncel himself was a possible confounding factor, as he taught both experimental groups but not the control groups. Tuncel provides pedagogical guidance, including recommendations for more information gap activities, and more authentic materials and interaction. Some recommendations are more clearly warranted by his findings than others (such as the need to monitor teachers' work). The study demonstrates that there is a continuing need to investigate the relative effectiveness of different broad approaches to language teaching using experimental designs, though such questions also require careful consideration of the rationale behind the approaches being evaluated and the nature of the language competence to be tested, in addition to the methodological concerns raised above.

The next four theses reviewed designed experiments outside normal curriculum time. Sheen (2006) carried out a pre-test, post-test and delayed post-test experiment to compare the effectiveness of different error correction techniques for learning *the* (as an anaphoric referential definite marker, e.g. *Pass me **the** pen*), and *a* (as a first mention, referential indefinite marker, e.g. *I saw **a** strange man*).

The study was conducted in a US community college with 10 native-speaking American teachers and their 143 intermediate level learners, at 'level 2' from a program with four levels. Sheen trained the participating teachers to deliver the interventions. Twelve intact classes were randomly assigned to the following intervention conditions: two oral corrective feedback groups (from five intact classes), i.e., an oral recasts group (*n* = 26) and an oral metalinguistic group (*n* = 26); two written corrective feedback groups (from five intact classes), i.e., a written direct correction group (*n* = 31) and a written metalinguistic group (*n* = 32); and one control group (from two intact classes) (*n* = 28). Each group contained some learners with +article L1s (e.g. Spanish) and some with -articles L1s (e.g. Korean). Sheen argued that as every group had roughly the same proportion of L1 Spanish learners (between 20% and 35%), then any L1 effect would have been observed equally in each group.

Three tests were used to measure learning gains: a time-constrained dictation test, a writing test, and an error correction test (administered in that order during each session). The delayed post-tests were administered three to four weeks after the end of the intervention. In the first



intervention session, all learners read a story and discussed the moral, and then the different feedback conditions were administered. For the oral treatment types, students retold the story to the class in groups of three. Recasts (either with or without metalinguistic information) were provided on article errors to individual students, but the whole class also heard these errors and corrections. In the written treatments, students rewrote the story. In the next class students received their stories with corrected forms (either with or without metalinguistic information). The control group had no stimulus material between the tests.

Sheen found that oral metalinguistic corrective feedback facilitated learning whereas oral recasts alone did not. She argues that metalinguistic feedback made the forms more salient, and promoted understanding (following Schmidt's (1990) conceptualisations of awareness and noticing). Written correction with and without metalinguistic comments were both effective in improving the grammatical accuracy of English articles, but metalinguistic comments led to greater gains. Written correction was found to be a more effective strategy than oral recasts. Oral and written metalinguistic feedback was found to be equally effective. Sheen argues that this provides evidence that written grammar correction can develop interlanguage. An exit questionnaire indicated the following rank order of learners' reported awareness of the four corrective feedback techniques: written metalinguistic (52%), oral metalinguistic (35%), written direct (25%) and oral recasts (0%).

Sheen also looked at whether three individual different variables influenced the effectiveness of error correction techniques: attitudes towards error correction and grammatical accuracy, language analytic ability, and learner anxiety. Learners with stronger analytic abilities improved their grammatical accuracy with English articles to a greater extent than those with weaker analytic abilities when provided with all types of corrective feedback, except oral recasts. Less anxious learners benefited more from oral recasts with metalinguistic correction than more anxious learners. Oral recasts alone were effective only for learners with low levels of anxiety. Sheen suggests that anxiety interfered with the ability to notice oral error correction but not with either type of written feedback. Learners with strong preferences for error correction and linguistic accuracy achieved greater grammatical accuracy following all types of feedback except oral recasts, compared to those with negative attitudes towards corrective feedback. Sheen suggests that her data are compatible with previous studies which have found that focussing error correction on many forms is ineffective, as her data suggested that focussing on just one feature did result in learning gains.

Sheen takes care to base her implications on her actual data, and goes on to acknowledge appropriately almost every conceivable limitation to her study, with relatively precise suggestions for future research. She points out that the written format of the tests could have biased the results in favour of the metalinguistic feedback groups, though she could perhaps have gone further to address the issue of comparability between the oral and written feedback techniques. As Sheen notes, further studies are now required to tease apart the effects of providing the correct form and providing metalinguistic information, and to explore the role of other individual differences, such as working memory and phonemic coding ability, in the effectiveness of error correction. Critically, Sheen notes that as the control group did not have any stimuli, the study could not assess the role of simple exposure to the stimuli. Similarly, the brief nature of the treatment may have prevented the recasts (an implicit feedback technique) from leading to statistically significant learning gains. These last two

points are important in that they constitute an (albeit indirect) acknowledgement that the research design could not have found evidence AGAINST Schmidt's (1990) noticing hypothesis, though evidence compatible with it was found. However, such brief bursts of teaching on a language feature IS reflective of normal practice and this affords high ecological validity to the study. Another sensible argument put forward relating to ecological validity is that, whereas some SLA (second-language acquisition) research has investigated implicit corrective feedback which is tailored to INDIVIDUAL students' zone of proximal development (Vygotsky 1978), Sheen points out that this is unrealistic in a whole-class situation. Her study is therefore significant in that it suggests that metalinguistic feedback with a recast may be the best option under normal teaching conditions.

The relative effectiveness of different ways of presenting and practising idioms was the focus of Skoufaki (2006), who conducted two pairs of experiments with Greek students from various disciplines at the University of Athens, aged between 17 and 28 years, and who were attending classes preparing for the Cambridge Proficiency in English examination. Each experiment had a similar timed test phase: a cloze test, missing two or three words of each idiom, to assess retention of the idioms' form; and four questions to assess retention of the idioms' meanings. Skoufaki was the instructor throughout, thus addressing the 'teacher variable' more rigorously than in some of the other studies reviewed.

The first pair of intervention experiments compared the effectiveness of presenting L2 idioms in Conceptual-Metaphoric groups (e.g. grouping idioms which associate morality with cleanliness, as in 'a clean fight' or 'doing the dirty on someone') with their presentation in Functional groups (e.g. 'idioms about morality'). In the second experiment, these presentation conditions were claimed to be more incidental (participants were not asked to memorise the forms), a wider range of grammatical structures was included (noun phrases and phrasal and prepositional verbs), and there was a practice phase between the exposure and the test.

This first pair of experiments indicated that the Conceptual-Metaphoric categorisation of verb phrase (VP) idioms was more effective for remembering the FORM of idioms than the Functional grouping, provided that learning was intentional or that the idioms were highly transparent. Form retention did not differ significantly between conditions for different grammatical structures. Retention of the MEANING of the idioms was not affected by whether the idioms were in Conceptual Metaphoric or Functional groupings, regardless of whether there was practice or instructions to remember the idioms. Although somewhat counter-intuitive, these findings do seem to suggest, in line with Skoufaiki's argument, that the form-meaning link emphasised by the Conceptual-Metaphoric grouping was useful for learning form, but not meaning.

The second pair of experiments tested whether instructing learners to guess the meaning of idioms led to better memorisation of form and meaning than simply presenting the idioms. In the first experiment there were three exposure/test conditions: 1) presentation + guessing + comprehension practice + comprehension test; 2) presentation + comprehension practice + comprehension test; 3) presentation + guessing + comprehension practice + cloze test (to check for a task-test familiarity effect). In all conditions the idioms were presented in Conceptual-Metaphoric groups, and there were no instructions to memorise the idioms. Skoufaki found significantly better retention of the idioms' form following the '+guessing' conditions compared to the '-guessing' condition. She claims that this suggests that the former

approach is a more promising method for teaching the forms of L2 VP idioms. There was no difference between the two different test scores from conditions 1 and 3, suggesting that the nature of the achievement test did not matter and no significant differences in meaning retention between the conditions, which is compatible with existing literature claiming that the retention of the meaning of new words is not higher when students are encouraged to guess at the meaning of new vocabulary. The second in this pair of experiments replicated the first but without the practice phase. In terms of form retention, there were no differences between the conditions; in terms of meaning retention, the condition '+guessing' resulted in slightly higher scores, though Skoufaki argues this finding as a methodological artifact. Her findings, therefore, seem compatible with previous research which has found that different types of interventions do not lead to benefits in learning the meaning of idioms.

Skoufaki acknowledges which variables could not be addressed by her studies, and suggests future research should address the effects of L1/L2, age, proficiency and individual differences, among other issues. She also provides details about the shortcomings of her outcome measurements, and she checked the normality of the distribution of her test data to choose the appropriate statistical test, a step rarely reported in the theses we reviewed. Even though three out of the four experiments were carried out with two learners at a time (rather than in a more authentic whole-class situation) and sample sizes were small (with ten in each condition in each experiment), the implications Skoufaki drew for teaching are, on the whole, appropriately linked to the variables manipulated in her experiments. This inevitably resulted in rather constrained and highly specific pedagogical advice, an illustration of the tension between external and internal validity in classroom experiments.

A CALL (*computer-assisted language learning*) reading package was developed and evaluated by Ciobanu (2006), among Masters students training to become professional translators at the University of Leeds. He first presented a 'model' of L2 reading called 'the multilingual resource-rich reading model – M3RM'. This was specific to reading in an L3 with some knowledge of a typologically related L2 (in his study, learners of Romanian L3, with French as an L2). The model presented several perspectives on learning and on language acquisition, and was heavily influenced by Grabe & Stoller (2002). Ciobanu claimed that he then operationalised key aspects of this work in a multilingual, corpus-based computer environment called the TRilingual REAding Tutor (TREAT), though specific links between the research reviewed and features of the actual pedagogical tool were not always clear to us. Several instruments were used in the creation of the environment, including L3 semantic networks, POS (part of speech) and lemma taggers,<sup>6</sup> and lists of L1–L2 true cognates. TREAT contained a corpus of written on-line news items, categorised for readability, in English (131 articles), French (100 articles) and Romanian (182 articles), totalling over 238,000 tokens. Learners selected their own texts and then requested help from TREAT when they needed it, such as identifying semantically related texts, viewing vocabulary in different collocational patterns, accessing bilingual and trilingual definitions, synonyms and translations, obtaining

<sup>6</sup> Ciobanu found that the accuracy of the POS taggers decreased when dealing with closed class function words. He proposed that for future versions the parser should work only with words of a certain minimum length. However, how simply avoiding many function words would improve overall accuracy was not clear to us.

morphosyntactic and frequency information about lemmas and tokens or extracting the meaning of each sentence.

Two groups of students used TREAT, though no base line tests were used for their selection: eight in group one (G1), with no prior knowledge of Romanian, and seven in group two (G2), some of whom had previously attempted to learn Romanian. The intervention lasted for six 1.5-hour lessons. G1 used TREAT for all sessions, but G2 did not have access to it for the first two sessions and was then introduced to it in session three. Details regarding G2's 'traditional explicit instruction' (p. 15) during the first two sessions are unfortunately not provided in the thesis, though a personal communication from Ciobanu stated 'a couple of online bilingual dictionaries and some rather academic Romanian grammatical resources' were available to them. For both groups, the first two sessions consisted of 'a lot of grammatical and syntactic ... work on Romanian texts' (personal communication, 22/12/08) and the remaining four sessions consisted of four L3–L1 translation tasks, and summarising and skimming tasks. Before using TREAT, both groups were given a 15-minute introduction to it and a brief description of L3 morphology. The learners also had to translate a number of nouns which required morphological nominal affixes from L1 into L3, e.g. *right*, *the right*, *rights*. As far as we understand, the translation and morphological tasks also served as the outcome measures. The translations were rated by L3 native speakers for content and style.

Ciobanu acknowledges that his samples are small, but argues his data suggest that although G1 had less prior knowledge, they performed better than G2 in the tasks (e.g. achieving a mean score of 82.95% in the morphology tests, compared to 50%). Most of the comparisons presented suggest an improvement between G2's scores without access to TREAT and those once they had access to it.

Ciobanu also reports on a questionnaire eliciting attitudes and self-reported use of TREAT. He states that all participants used the L3 linguistic information provided. Although no direct tracking or think-aloud protocols were actually used to corroborate this, a user-tracking mechanism is proposed for future work. He also reports that participants thought their L2 had improved, though no measures of L2 knowledge were taken.

This study was not a formal experiment. Technical changes were made as the study progressed and task-familiarity (rather than learning *per se*) may have led to improvements, as learners became more familiar with the environment. As a result of these, and other potential alternative explanations, the effectiveness of TREAT compared to the (somewhat vague) comparison treatment was not entirely clear. Also, Ciobanu acknowledged that further work is required to investigate, and incorporate into CALL packages, lower and higher-level online reading processes, such as lexical access, parsing, and inferencing. Nevertheless, this work demonstrates how state-of-the-art technology can create rich sources of written input and metalinguistic knowledge for language learners. Also, as Ciobanu suggests, such online environments may offer one way of tracking reading processes, as they provide on-demand lexical and morphological assistance.

Niño (2006) also evaluated the use of computers in the language classroom. The study aimed to (i) describe and evaluate an intervention which required learners to edit machine translations (MT), (ii) categorise errors in the raw MT, in the MT once it had been edited by students, and in student-alone translations, (iii) compare the errors found in edited MT and student-alone translations, and (iv) make pedagogical suggestions for using MT. Niño

developed an error typology and an error tagset, and reliability measures were taken. The error tagging system aimed to satisfy Granger's (2003) four principles: Informative but manageable; Reusable; Flexible; Consistent. The unit of accuracy was 'words', and the system consisted of 50 error codes corresponding to 50 error categories, belonging to four error domains (lexical, grammatical, discursive, and spelling). She used a semi-automated error tagger (Markin<sup>7</sup>) to retrieve error statistics and also investigated the perceptions of teachers and learners on the use of MT in the language class.

In the 'experimental' group were sixteen adult learners of L2 Spanish, with no prior experience of MT, at Common European Framework 'Vantage' or B2 level, from three different institutions in Manchester. The 'control group' had sixteen students who were from one of these institutions, which perhaps weakened parity between the groups.

The students in the experimental group attended a ten-week course of two hours a week on editing MTs and writing practice for advanced-level Spanish. The first session consisted of general instruction in FL writing, including provision of some additional resources. They also had an introduction to MT, and guided practice at editing MTs.<sup>8</sup> In each of the following eight weeks they had one instructed session to practise editing MTs of different genres of writing, and then had one week in which to translate a text. They could also use other language resources (online dictionaries, glossaries, etc.). The students in the 'control' group had to translate the same texts but by hand. They were provided with the same course materials as the experimental group, except for the MTs, and they did not have the introduction to MT.

The results included details of error frequencies per error domain, per error category, and per text type. The MT output had 507 errors, 52% of which were grammatical, 38% lexical, 7% spelling and 2% discursive. It was found, unsurprisingly, that the students improved the MT texts. Medium effect sizes suggested that the edited MT had fewer errors (overall, and in grammatical, lexical and discursive domains) than the student-alone translations (3,282 compared to 4,250 errors). Critically, however, these differences were NOT statistically significant. The only error domain in which there were statistically significantly fewer errors in the MT group was in spelling. Although not discussed, this difference could partly be accounted for by typos, which computers are less prone to!

Niño quantified successful and unsuccessful editing. She found that a mean of 60% of all the MT errors were corrected by the learners, 74% of the correct MT was left as correct, 12% of the MT errors were left as incorrect, and 2% of the correct MT was changed unnecessarily. As far as we can calculate, this analysis accounted for 72% of all errors in the MT, and 76% of the correct MT. Niño argues that MT editing and translation activities are similar in terms of difficulty, and that this warrants the recommendation that MT editing should be used as a written production activity in the same way as translation. It was not entirely clear to us how comparable the experimental and control interventions were as it seemed, according to some parts of the thesis, that the MT group perhaps received more, and potentially better, instruction, in addition to the 'MT' texts. However, as Niño drew conclusions based on the

<sup>7</sup> We were unable to track any detailed reference entry for this work

<sup>8</sup> *Systran Professional Standard 3.0* was selected because, compared to three other systems, it had the lowest overall number of errors (79 in total) on the same set of 112 difficult sentences.

similarity of the results from different groups (not on differences between them), this issue may not threaten the validity of the findings too seriously in this case.

Attitudinal data were also collected from the MT group, via a questionnaire, and opinions about using MTs were generally positive. A survey sent to teachers (including some sent 'randomly to language teachers around the world' (p. 111)) on the use of translation and MT in the FL class revealed that 23% of the respondents had used MT in their lessons. The data collected therefore generally tapped teachers' perceptions, rather than their direct experience of how effective it might be for learning.

The author goes on to make several strong recommendations for using MT in lessons, including with learners at a different proficiency level. As acknowledged by Niño, future research will have to determine whether MT aids learning in the longer term, by using a pre-post-test, control group design. Niño also appropriately suggests that future research should investigate whether editing MT promotes the use of learner strategies. The study demonstrated that MT is a good source of human-like errors, useful if we consider explicit error correction beneficial. The findings should help raise awareness of the nature of MT output, particularly in conjunction with issues of plagiarism. Although unacknowledged, perhaps the similarity in type and number of errors of edited MT and student-only translations could be used to demonstrate that MT is UNLIKELY to help you get a better grade?

### 3.1 Summary and further commentary

This review suggests that process-product, effectiveness-of-intervention studies are receiving due attention in Ph.D. work, although all those reviewed were with learners in post-compulsory, higher education settings. The studies reviewed tended to use experimental or quasi-experimental designs, and quantitative data and analysis as their main measurement of effectiveness. Several studies also investigated what the teachers or learners thought of the intervention. In the studies which set up interventions (i.e. all except Al-Ghamdi's), most tried to justify, to some extent, the naturalness and appropriateness of these interventions, and therefore, their likely external validity. It was noted that none of the studies directly addressed the debate among educationalists about the usefulness, feasibility and methodology of educational experiments (e.g. Hammersley 1997; Torgerson & Torgerson 2003). Perhaps most surprisingly, Norris & Ortega's (2000) seminal paper, relating specifically to L2 teaching and learning experiments was not referred to. This body of literature discusses the roles of randomisation, control/comparison groups, and outcome measurements which have the potential to indicate reliable and valid comparisons. Perhaps because of their lack of engagement with this agenda, several studies reviewed here made pedagogical recommendations which were not warranted by their research design or findings. In most studies there was a tendency to imply greater generalisability than was warranted, often by using present rather than past tenses to report and discuss findings. Some researchers did not consider the role of extraneous variables in explaining their findings, such as the teacher effect, class or school (cluster) effect, test effects, maturational effects, or Hawthorne effects. Some did acknowledge such threats to validity, but their designs were not able to address them. Critically, there was often a lack of detail regarding the control or comparison groups' intervention



(except in Skoufaki and Sheen's theses). Descriptions of the nature and educational objectives of all the conditions (experimental, comparisons, and control) are useful if readers are to judge the relevance of the study to other contexts, and if researchers are to claim that the outcome measurements did not favour the learning objectives of one of the intervention types. Nevertheless, some studies reviewed, most notably Sheen's and Skoufaki's, illustrated that experiments with good internal validity can also contribute something externally valid to pedagogy.

#### 4. Individual differences

Three studies looked at aspects of motivation, although from different perspectives: learner autonomy (Bavendiek 2006); the relationship between attitudes, learning styles and experiences of language learning (Hemmings 2006); and attitudes to project-based learning (Stolyar 2006).

The questions underpinning Bavendiek's research are the extent to which a guided independent learning programme, based on portfolio learning and 'advisor' sessions, can increase undergraduate students' degree of reflectivity and hence their use of metacognitive language learner strategies, feeling of control, confidence and motivation, in both the short and long term. She also investigated the relationship between these variables. Bavendiek bases her study on a constructivist and experiential view of learning, following Kolb's (1984) experiential learning cycle and Thomas & Harri-Augstein's (1985) theory of self-organised learning in which awareness of the learning process is especially important. From this background, Bavendiek develops a process model of learner autonomy, which suggests that learners' engagement in portfolio work should improve their language learning skills and strategy use and therefore their sense of control over their learning, their confidence and hence their motivation. Thus, her initial assumption is that the PROMOTION of learner autonomy is possible, and she asserts that her study seeks to 'investigate the claimed effects identified from the model' (i).

The portfolio programme was implemented as a compulsory element in the first year of a degree course in German in an English university. Students completed four 'projects' as part of the portfolio, in which they recorded their reflections on work completed in a journal format. Each project was then discussed in an 'advisory session', with the researcher, in either the L1 or the L2. The impact of the portfolio was assessed by a variety of means, over a period of four years, giving an impressive array of data. Changes in learners' use of metacognitive strategies and their feeling of control were measured through two questionnaires administered before and after the portfolio intervention to an 'experimental' group of 55 learners, and to a 'control' group of 22 learners from other universities (the appropriateness of these terms is only briefly addressed by the author). Items to assess metacognitive strategy use were adapted from Oxford's (1990) Strategy Inventory for Language Learning, described as 'an established instrument in the field' with 'well documented' validity (Bavendiek 2006: 93), although the reliability and validity of the adapted version used in the study do not appear to have been assessed. Feeling of control was elicited through a set of attributional statements (explanations for perceived achievement in language learning) which participants ticked as they applied to



them. Contrary to expectations, the pre-test and post-test questionnaires showed a significant reduction in the mean metacognitive strategy use of the experimental group. The control group's strategy use declined even more (but not significantly more) than the experimental group's. Students' feeling of control hardly changed over the year regardless of their group, and no significant difference was found between the two groups in terms of changes. In addition, no significant relationship was established between metacognitive strategy use and feeling of control.

An ANOVA was used to explore the level of reflectivity shown in journals over time. Bavendiek reports a significant positive relationship ( $p = .041$ ), though it was weak ( $r^2 = .054$ ) and that 'the relationship shows a drop over time on this variable' (p. 129). She concludes that students who appreciated the project journals from the start continued to do so, and those who did not, retained their view of them, seeing them as burdensome and simply recording activities completed rather than their reflections on them.

An additional questionnaire completed at the end of the first year by participants in the experimental group sought to evaluate perceptions of the portfolio in terms of its impact on participants' language learning skills and strategies, target language proficiency, feeling of control over the learning process, confidence and motivation. Most learners felt that the programme had had a positive impact on these areas. Correlational analysis established a significant and moderately strong positive relationship ( $r$ ) between: motivation and perceived improved language proficiency (.636), motivation and acquisition of new language learning skills and strategies (.609), and perceived improved proficiency and confidence (.589). Weaker correlations were found between motivation and a sense of control (.424) and confidence and a sense of control (.450), leading the author to conclude that a sense of control is not always translated into increased confidence, as the process model of learner autonomy predicted it would be. A later questionnaire likewise produced mixed results, with 72% of respondents feeling that the programme had made them more conscious of the language learning process. While 94% of those who had not yet completed their year of study abroad generally felt the portfolio would help them, only 39% of those who had completed the year abroad felt the programme had assisted their language learning.

Finally, case studies were composed, using questionnaire, project data and semi-structured interviews with four learners, in order to explore whether any individual difference variables could be associated with different perceptions of the portfolio. Bavendiek concludes that a positive attitude towards a portfolio programme seems to be found among students with higher language proficiency at the end of the degree programme, an integrative orientation for language learning and those who already show a degree of independence in their language learning at the start of the course. In other words, the programme seemed to benefit most those who needed it least. In addition, Bavendiek claims that the study does not confirm 'the virtuous circle of learner autonomy' (p. 245) which suggests that reflection increases metacognitive strategy use and hence a sense of control (although she admits, perhaps with justification, that the validity of her instrument for measuring the latter may be open to question). In spite of these less than positive findings, however, Bavendiek argues that the portfolio programme seems to have at least lessened the decline in sense of control and metacognitive strategy use, as compared with what students in other universities appear to display during the transition to Higher Education. Consequently, the programme is still

being implemented at Bavendiek's university, but in a modified form, without the project journals.

The relationship between various factors that might contribute to learners' motivation is also explored by Hemmings (2006), one of only two theses reviewed in depth that investigated non-adult learners. The main focus of her study is the learning styles and attitudes towards language learning of pupils in the first year of secondary school in England (Year 7, aged between eleven and twelve), resulting in three main questions: (i) Is there a relationship between pupils' initial attitude towards language learning and their preferred learning styles? (ii) What is the influence of pupils' initial attitudes and preferred learning styles on their experience of language learning during Year 7? (iii) Is there a relationship between pupils' learning styles and the extent of their success in learning German by the end of Year 7?

The learning preferences and initial attitudes towards language learning of 162 learners at the start of Year 7 were assessed through a questionnaire, which included items on pupils' creativity, self-awareness, attitudes towards social interaction, experience of language learning at primary school, visits abroad, and expectations of language learning at secondary school. Language teachers at the project school also completed a questionnaire on their beliefs about language learning. A purposive sample of pupils learning German (80 pupils, three classes) and their teachers then completed three commonly-used learning style inventory questionnaires to establish the dominant learning style for pupils and teachers: a questionnaire based on Kolb's (1984) experiential learning model; Ginnis' (2002) sensory learning style inventory, based on the visual, auditory and kinaesthetic model of sensory preferences widely used in UK schools; and an inventory based on Sternberg's mental self-government model (1997), which was used at the school where the study took place. Acknowledging the shortcomings of the learning style inventories employed (as highlighted in the literature), Hemmings reviews the theoretical background of these, alongside theories of motivation and 'The Good Language Learner' study by Naiman et al. (1978). The attitudes and learning preferences of a smaller sample of pupils ( $n = 48$ ) were further explored through group interviews. Their teachers were also interviewed and German lessons observed. Pupils' success in German was ascertained by the ability group they were placed in for German in the following year, and scores on a test of cognitive and academic ability (CAT) scores (see Thorndike & Hagen 1986<sup>9</sup>) were taken as an indication of their general academic ability. Hemmings summarises her findings thus: some pupils displayed a negative attitude towards language learning even before they had started studying in Year 7; these pupils were more likely to be placed in a low set in Year 8 for German; more boys than girls showed an initial negative attitude; pupils with an initial negative attitude had a range of CAT scores (i.e. negative attitude was not linked to low academic ability); pupils with negative attitudes were more likely to be kinaesthetic and/or legislative and/or accommodative learners. However, from her pupil interviews, Hemming suggests that other factors (such as anxiety and fear of embarrassment in class) influence pupils with negative attitudes more their learning style does. She goes on to conclude that initial attitude seems to be the most important factor in determining success in language learning but adds (reasonably) that the high level of teacher control she observed in lessons made it difficult for learners to demonstrate a learning style preference in any case.

<sup>9</sup> This reference is suggested by the current authors, and was not cited by Hemmings.

In the other study that involved school-aged learners, Stolyar (2006) looked at attitudes towards one particular approach to language learning: project-based learning (PBL), within the context of Israeli learners of EFL in junior to high school (i.e. aged 11–18 years). She investigated both the nature of students' attitudes towards PBL and the extent to which certain factors (working in groups, level of interest in the project, prior experience of project-based work, age, place of residence and gender) contributed to positive attitudes towards PBL. The underlying premise of her study is that attitudes have a fundamental impact on attainment in foreign language learning. Project-based learning is described as a form of teaching whose goal is 'the satisfactory completion of a meaningful communicative task or enquiry' (Stolyar 2006: 30) with roots in the approaches of Dewey and Vygotsky.

Stolyar investigated learners' attitudes and factors contributing to them using mixed methods, beginning with a questionnaire to 531 learners from 13 schools. Participants were selected on the basis of their teacher's participation in a course on PBL led by the researcher (i.e., their teacher had knowledge of PBL). The students sampled differed in how much PBL experience they had: 53% were classed as 'experts', having done between two and four projects, and the rest were 'novices', on their first project. The questionnaire was based on published instruments, was piloted, and reliability assessed through factor analysis and Cronbach's alpha (steps that were not taken in the other theses reviewed in this section). Respondents had to indicate whether their project work was carried out in groups or individually. Two cycles of qualitative data collection were also completed with the researcher's class as they engaged in project-based learning, involving semi-structured interviews, lesson observations, field notes and video recorded lessons.

Factor analysis of the questionnaire revealed six constructs which accounted for 50% of the common variance: attitudes towards group work, general attitudes towards PBL, attitudes towards oral presentations, attitudes towards the perceived effect of PBL on students' learning, difficulties experienced in the PBL process, and attitudes towards the use of computers. Independent samples *t*-tests indicated that students who worked in groups had significantly more positive attitudes overall than those who worked individually. Correlational analyses showed a significant positive relationship between students' attitudes and the number of projects completed. In addition, ANOVAs followed by post-hoc tests (Scheffé) indicated that students on their fourth project were more positive towards PBL than those on their third project, who in turn were more positive than those on their first or second. Stolyar also found that level of 'interest' in the subject of the project correlated significantly with all six constructs. In order to investigate whether attitude toward PBL can be predicted by personal background variables (gender, junior/high school, working in groups/alone, age, level of interest in the project subject, school grade, project exposure), a least squares linear regression analysis was conducted. This revealed that 'level of interest' accounted for more than 35% of the variance on attitude towards PBL, with all other variables together adding only another 4%. The general trend was for Grade 7 learners to have the most positive attitudes towards PBL, and for Grade 9 learners to be the most negative.

Learners from whom qualitative data were collected were taught by the researcher and were working on two projects, each lasting nearly two months. At the time of the first project, the class consisted of 30 Grade 10 students, with a range of proficiency levels in English. Students chose their own groups to work in. For the second project, 14 additional students had joined the class, four of whom had had no previous PBL experience. Eight students from

the first project and 34 students from the second were interviewed about the PBL experience. Students interviewed perceived group work positively and most expressed a high level of interest in the topic of the project, leading Stolyar to conclude that giving students a choice of topic for the project is important. Generally, students who joined the class for the second project, with no previous PBL experience, were more negative than those who had been in the group at the time of the first project.

Stolyar draws a number of main conclusions: (i) that younger learners are more positive towards PBL, suggesting it should be introduced in early grades; (ii) that students' attitudes towards PBL are influenced positively by group work, by prior (positive) experience with PBL and by interest in the project of the topic; and (iii) that the end result is more positive attitudes towards 'learning in general' (p. 212). While the study provides a range of evidence for the first two of these conclusions, the last is based on more limited data from the qualitative interviews and is more debatable, with Stolyar arguing that 'once PBL creates a genuine interest in the subject matter, it can be ASSUMED that this results in better attitudes toward learning' (p. 199; emphasis added).

In the area of learner strategies, two studies were interventions (both concerned with reading strategies) and one a descriptive study of strategy use. Walters (2006) focussed on context-based forms of strategy instruction, and in her literature review she determines three main approaches: the teaching of specific context clues; cloze exercises as a way of enhancing awareness of the role of context; and the teaching of one general strategy for dealing with unknown words, i.e. looking for other words that might help work out the meaning. These three approaches then formed the basis of her study, which investigated whether instruction in the use of context to infer meaning leads to (i) 'increased ability to infer meaning from context', (ii) 'improved reading comprehension', (iii) long-term gains in these areas, (iv) 'greater vocabulary growth', and (v) whether the type of instruction in use of context affects 'the ability to infer meaning from context, or reading comprehension, or vocabulary growth' (Walters 2006: 91). Forty-four ESL university students on an intensive English language programme, from 12 different countries, took part in a pre-post-delayed post-test design, using the following instruments to assess each area highlighted in the research questions: a cloze test with rational deletion, developed by the researcher; three versions of the reading section of the IELTS ESOL examination; the 3K, 5K, Academic Word List and 10K subset of the Vocabulary Levels Test (Schmitt, Schmitt & Clapham 2001); the 3K, 5k and Academic Word list subsets of the productive Vocabulary Levels Test (Laufer & Nation 1999). To establish evidence of strategy use in practice after the training, Walters also conducted a retrospective verbalisation task with two to three subjects from each condition (not from the control group). Interactions between the training and the individual differences of proficiency, learning style and language aptitude were also explored.

Participants were allocated, not randomly, to one of four groups: control, general strategy, context clue, cloze practice. Students in the three treatment groups received three training sessions of two hours each, with the researcher delivering the training for the context clue and general strategy condition. The training was immediately followed by post-tests (between one and three days after the end of the intervention). In order to assess the effectiveness of each of the three types of instruction (Research Question 5), Walters compared all four groups with each other. She also combined the groups, and made two other sets of comparisons: a group labelled 'experimental' (consisting of the general strategy + context clue + cloze

practice groups) which was compared with the control group; and a group labelled 'training' (consisting of the context clue + general strategy groups) which was compared with a 'no-training' or 'practice' group (cloze practice) and the control group. This gave rise to a complex set of results, which are not always easy to follow in the thesis, and the most important of which are summarised below.

No significant differences were found in mean gain scores between any groups in any comparison in terms of ability to infer from context, although only the combined experimental group made significant pre to post-test gains. For the reading test, the mean gains of the combined experimental group and of the training group were significantly higher than those of the control group. When individual groups were compared (for Research Question 5), there were no significant differences between the groups for ability to infer (although the general strategy group alone showed significant pre to post-test improvement), but for the reading test, the context clue group made significantly greater gains than the control group (and both context clue and general strategy groups showed significant pre to post-test improvement). At delayed post-test, there were no significant differences between any groups for mean gains between the post and delayed post-test in ability to infer. For reading ability, while the experimental, training and practice groups maintained the gains made at post-test, the control group had regressed by the time of the delayed post-test. In terms of vocabulary development, the only significant findings reported are that the combined experimental group and the training group showed a significant improvement in receptive vocabulary at the 3K level, and that for productive vocabulary, the experimental and training groups showed significant or near-significant improvement (at the 3K and Academic Word List levels).

Walters comments that both context clue and general strategy instruction 'have a positive effect on the ability to infer from context and /or reading comprehension' (p. 151). We felt that her later appraisal of the findings, that there was only 'weak' evidence for the experimental conditions having an impact on learners' ability to infer (p. 234), is a more appropriate conclusion, as is her comment that it seems more likely for the effects of instruction in inferencing to manifest themselves in general reading ability, rather in the ability to infer as a separate skill.

In terms of individual differences, some interaction was found between training condition and language proficiency (advanced, intermediate or beginning based on TOEFL or TOEIC scores). Significant or near-significant differences were established in the inferencing test, for the context clue group, where advanced learners' gains were greater than the intermediate learners' ( $p < .06$ ), and for the strategy group (beginners making more progress than intermediate learners,  $p < .05$ ). From these findings Walters concludes that higher language proficiency may be needed to benefit from context clue instruction, while a general and simpler strategy for coping with unknown words may be best for beginners.

Although the study has limitations (as Walters acknowledges), including the small number of participants in each group, no assessment of strategy use before the training, unrandomised allocation to groups, as well as some doubts about the validity and reliability of the tests used for assessing ability to infer, it is noteworthy as one of the few studies to consider vocabulary development as a by-product of strategy instruction in reading and to investigate the interaction of strategy instruction with a range of individual differences.

One aspect of strategy instruction that is noticeable by its absence from the Walters study is a metacognitive element, which by contrast features prominently in Wiliwan (2006). Wiliwan also looked at strategy instruction as a means of improving reading comprehension, again in the context of adult EFL learners, but this time in Thailand and with a focus on improving main idea comprehension. She argues that finding the main idea is a crucial skill for effective reading and, reviewing models of reading comprehension, that successful reading requires an integration of bottom-up and top-down strategies, with metacognitive strategies regulating this integration and with skilled readers possessing a clear awareness of how they read. Linguistic theories of discourse processing (e.g. van Dijk & Kintsch 1983; Hoey 1991) are drawn on to argue that making students aware of lexical relations and lexical cohesion devices in a text can help them in establishing the main idea of the passage.

This theoretical background underpins the selection of the types of instruction that were assessed in the project: metacognitive strategy training alone (taught through a 'reciprocal teaching' method – the RT group); lexical cohesion plus metacognitive strategy instruction (LRT group); and 'traditional' skill-based instruction as a control (the ST group). The participants were 60 undergraduate students (aged 19–22 years), taking an English reading course at intermediate level. Stratified random sampling was used to assign students to the three conditions.

The four research questions addressed were: (1) whether the LRT and RT groups 'achieve significantly higher scores from the pre-test to post-test' than the ST group; (2) whether there are 'any differences between the mean values of the experimental and control groups' (presumably at post-test), and 'which of the three instructional methods is superior in producing higher levels of main idea comprehension improvement'; (3) whether the experimental and control groups differ in their attitudes towards their instructional methods; and (4) whether the LRT group uses different main idea processing strategies from the RT and ST groups (Wiliwan 2006: 65–66).

The instruction was carried out over a 15-hour period (the researcher taught the LRT group). For the ST group, this consisted of (i) pre-reading – e.g. looking at the passage title and making predictions; (ii) during reading – e.g. in small groups, the students discussed the passage content and translated each sentence into Thai; and (iii) post-reading – students discussed the passage content in groups and collectively came up with a main idea statement. For the RT group, the teacher modelled what Wiliwan calls four metacognitive strategies – summarising, questioning, clarifying and predicting (although whether all are really metacognitive strategies was unclear to us). Students took turns to lead the conversation about the text, after the teacher's initial modelling of how this should be done. Finally, the group came to a collective conclusion about the main idea of the text, which was then verified by the teacher. The instruction for the LRT group followed the same pattern as for the RT group, except they also learnt how such lexical cohesion devices as repetition and restatement link phrases and sentences together to convey the main idea of a passage. Practice activities were conducted involving lexical cohesion devices.

The impact of the three instructional approaches on students' main idea comprehension was assessed using a test designed by the researcher, including multiple-choice questions and production sentences, in which students had to write down details of the main idea expressed in texts. Each of these two sections was further sub-divided into 'stated' and 'unstated' main



idea sections. Students' strategy use was investigated through a think-aloud interview with a selection of participants (before and after training, an approach not always followed in strategy-based interventions), and then through strategy interviews with a sub-section of this group. Finally, students' perceptions of the instruction they had received were explored through a post-test questionnaire, given to all 60 participants.

For Research Question 1, paired sample *t*-tests compared pre and post-test means for the main idea test per group. Each group made a statistically significant improvement, with the LRT showing the biggest gain, followed by the ST group. Wiliwan (2006: 111) comments that 'it is obvious that the LRT treatment was the most effective instructional approach', although this claim is hard to justify, given that differences between the groups in mean gain scores were not explored through statistical tests. For Research Question 2, ANOVAs were conducted for the post-test scores (although group differences at pre-test were not explored), looking at the overall scores, and also each part of the test separately. No significant differences were found for any part or the whole of the test, although the LRT group had the highest score in each section, except for the stated main idea production section, in which the ST group had the highest score. In the questionnaire, all participants were generally positive about their instruction, although the means for individual items suggest that the LRT group was the most satisfied, and the RT group the least. Differences were, however, small and as no tests of significance were applied, firm conclusions are hard to draw from these data.

Quantitative and qualitative analyses of the transcripts from the think-aloud and strategy interviews indicate differences in strategy use between the three groups after instruction (Research Question 4), thus lending some support to the conclusions drawn from the test data: the LRT group used more bottom-up strategies than did the RT and ST group, and reported using a wider range (especially word recognition skills involving word relations between sentences). All three groups increased their use of top-down strategies from pre to post-test, especially RT students, while LRT students used the widest range. LRT and RT students showed an increase in metacognitive strategy use and used a wider range, while ST students used such strategies sparingly. The latter group made few comprehension checks and commented rarely on their reading or evaluated their understanding. In addition, Wiliwan reports that the LRT group combined bottom-up, top-down and metacognitive strategies at post-test more than the other two groups did, and showed a much greater sense of awareness of how they read, commenting more frequently on their own behaviour. She attributes the greater progress made by this group in the main idea reading comprehension test to this combination of a wider range of strategies. In addition it is argued that the metacognitive component of this group's training 'helped students become consciously aware of what they were doing and be able to manage the use of appropriate strategies for deciding the important information of the text' (p. 185).

In terms of a descriptive examination of learner strategy use, Garcia (2006) is also concerned with lexis, but within the context of L1 and L2 writing. Defining a strategy as 'the solution that the subjects devise to overcome their lexical problems when they write' (p. 7), the author discusses, in particular, communication strategies, which have most often been considered in relation to oral production. Arguing that L1 and L2 writers undergo the same processes, but with L2 writers facing more discrepancy between what they want to say and what they can say, Garcia analyses two variables in his study – the language (L1



Spanish or L2 English) and L2 proficiency (elementary or upper intermediate). He poses 12 research questions with accompanying hypotheses to explore whether lexical problems faced and strategies employed differ according to these two variables. Participants were 31 Spanish-speaking EFL students from a Mexican university. Two groups were created based on scores on the Oxford Placement Test, and on Nation's Vocabulary Productive levels test: a higher proficiency group of 16 students who had either taken or were taking advanced courses in English and had mostly studied in the US and a lower proficiency group of 15 students, none of whom had studied in an English speaking country.

Two writing tasks were used, one in English, one in Spanish, completed by all participants, with one week between them. Participants were given 45 minutes to write, and were instructed to approach the task as they would in an examination. They could use any dictionary they wished. The main form of data elicitation was a think-aloud protocol (in the L1 or L2, as participants wished), concurrent with the writing task. A number of retrospective interviews were also conducted to further explore with think-aloud participants instances where problems seemed to occur. Questionnaires on writing practices were also completed.

Garcia identified 804 instances of lexical problems in the transcripts and classified them using an adapted version of Hemmati's (2001) taxonomy. This resulted in thirteen individual problems, of three main types: 'no-word' problems, where the L2 word was either unknown or could not be retrieved; 'one-word' problems, where the required word was known but there were problems with it, such as mis-spelling; and 'more-than-one-word' problems, where more than one word was retrieved and the student had difficulties in choosing between them. Problems were also categorised as either competence-based (the word is unknown) or performance-based (the word is known but cannot be fully retrieved). A taxonomy of eight strategies used to solve these problems was also created and grouped into four categories: 'trying to solve the problem, de-problematising the problem, giving up and modifying the content, and postponing the solution' (Garcia 2006: 254).

For lexical problems, Garcia reports that writers paid more attention to and acted upon more lexical problems when writing in the L2 compared with L1, facing mainly 'no-word' problems in the L2 and mainly 'one-word' problems in the L1. In the L2, both proficiency groups faced more competence-based than performance-based problems, although in the lower proficiency group there was a higher frequency of competence-based problems, and in the higher proficiency group, more performance-based ones, as might be expected. Both groups were motivated by a desire to compensate for insufficient lexical resources and/or to enhance what they had written. There were significant differences in how often the groups met certain problems (Garcia 2006: 258) – 'delayed retrieval of some aspect of a word' and 'selection dilemma of two or more suitable candidates in the L1' were more common in the higher proficiency group, and 'no L2 form for a given meaning' was more common in the lower proficiency group. Contrary to what Garcia anticipated, both groups tackled lexical problems at the moment they occurred, rather than waiting until they had developed their ideas.

Regarding L1/L2 strategy use, the dictionary was widely used for solving lexical knowledge problems, although less often in the L1 and with more success than in the L2 (again, as one might expect). For lexical retrieval and enhancement problems in both languages, students tended to turn to their own resources (e.g. using the text they had already written). Overall,

and perhaps unsurprisingly, students were more successful in solving lexical problems in the L1 than in the L2, although success tended to depend on the knowledge source used to help solve the problem. However, Garcia argues that the problems learners faced in using the dictionary in the L2 underlines the importance of training learners to use this resource effectively. Finally, higher proficiency learners seemed to be more flexible in their strategy use, to be more efficient in their use of the dictionary and to retrieve words more accurately.

Garcia concludes by relating his findings to theories of writing and argues, with justification, that lexical problems are particularly present in the formulation stage of composing, which has received little research attention. He adds that these problems and the strategies used to solve them are not merely compensatory but are also aimed at enhancement of what is written. He suggests that his findings be used to help teachers plan instructional materials that help students to deal with lexical problems, to transfer L1 writing communication strategies, and to use a dictionary more effectively, recommendations which are largely supported by his study.

#### 4.1 Summary and further commentary

This group of theses indicates that, in spite of the wealth of studies that already exists into attitudes and motivation in language learning, the topic continues to be a focus of interest among novice researchers. Two of them, Bavendiek (2006) and Hemmings (2006), seem to come to similar conclusions that pre-existing attitudes to language study or autonomous behaviour are difficult to change. By contrast, Stolyar's (2006) study suggests greater possibilities for the fostering of positive attitudes through a teaching approach that focuses on the key area of learner interest. In the area of learner strategies, that all three studies were concerned with lexis in one form or another may reflect a growing interest in the research community in the area of vocabulary.

All six theses used a mixed-methods approach, collecting both quantitative and qualitative data. These were used most effectively when one set of data complemented the other, as was particularly the case in Stolyar's (2006) and Wiliwan's (2006) study. Quantitative data were used most convincingly when the researchers justified their choice of the statistical test employed and demonstrated that their data met the necessary conditions for the use of the test. These steps were rarely taken, however, in the theses reviewed.

### 5. Discussion and concluding remarks

Here, we briefly discuss how the studies used theory, and we make a few general remarks about the reporting in the theses and about our review process itself.

The use of theory in educational research has been widely discussed, in terms of whether the theories used are falsifiable, whether research designs allow them to be falsified, and whether theory testing or building is possible if the research is genuinely useful to practitioners (Pring 2000; Thomas 2005; Marsden 2007). Space constraints do not allow us to discuss the nature

of the theories used in the theses we have reviewed. Here we briefly describe HOW theory was used.

None of the theses reviewed set out to test a learning theory, in the sense that the research designs did not allow for data to be elicited which could falsify a specific theory (though Bavendiek's aims were to investigate the 'claimed effects' of a process-model of learner autonomy). One study (Hill) claimed to build a new theory by describing a new perspective ('sociocognitive metaphorm') on how learners might understand and learn tense and aspect. Several theses collected data which were compatible with or illustrated a theory: Sheen's data were compatible with the Noticing Hypothesis and the Interaction Hypothesis (though could not falsify them as the control group did not have the same stimulus materials with no interaction); Stolyar's study was underpinned by a constructivist view of learning; Wiliwan draws on linguistic theories of discourse processing; and Walters' study is largely underpinned by theories of vocabulary acquisition (e.g. de Bot, Paribakht & Wesche (1997), who suggest that inferencing from context aids the process of creating and filling an empty lemma structure).

Several theses used one or more theoretical frameworks, drawing on their terminology, categories and assumptions, rather than providing valid evidence for or against a theory. For example, Hemmings reviewed a range of theories relating to learning styles, motivation and 'The Good Language Learner' (Naiman et al. 1978). A few studies used theory to provide a rationale for a teaching approach or pedagogical decisions, rather than providing insights into (e.g. supporting or refuting) a specific learning or teaching theory. For example, Tuncel found evidence that his classroom activities helped learning, and used a pedagogical perspective (CLT) as a backdrop to describing these activities. Skoufaki used linguistic idiom theory (e.g. defining metaphor types) to describe the idioms and presentation groups she used, and suggested that a cognitive perspective on learning could explain how figurative language is learnt. Ciobanu drew on several SLA theories to justify certain decisions he took in designing his CALL programme. Niño did not draw on learning theory in the design or analysis of her study, but provided evidence compatible with her own pedagogical hypothesis that using machine translations would be useful for learners.

One feature which struck us on reading so many theses together was the sheer volume of data collected. This included observation notes, classroom recordings, performance and achievement data (production and comprehension, in oral and written modalities), teacher and learner interview transcripts, and questionnaire data. For example, Niño created a learner corpus of 128 post-edited and 128 translated texts, an overall size of 56,893 tokens; Al Ghamdi collected 35 hours of lesson observation notes; Hill produced a 30,000-word database of transcriptions of 13 hours of audio-recorded L2 learner oral interaction. Most of these data are digital, and, therefore, uploadable to online repositories. We suggest that that making such data, and the collection instruments, available to researchers' would improve the replicability of postgraduate research and the usefulness of the findings for teaching and research.

During the initial 'key word' search, it became obvious that abstracts were not as helpful as one might expect. Some titles and abstracts did not mention generic terms such as 'teaching', 'learning', 'pedagogy', 'second' or 'foreign'. On the other hand, some abstracts did not provide key details, such as research questions, details of the methods (e.g. participants' age, proficiency, native or target language), or main findings. In their place were often detailed

rationales and/or theoretical backgrounds. Similarly, the methods and procedures used were not always as accessible in the body of the theses as one would wish, and key details could be hidden in unexpected places. We do, however, acknowledge that we have not been able to report fully on the detail of the theses that we have reviewed. We strongly recommend readers pursue the publications that have emanated from these doctoral studies, but which space constraints did not allow us to reference.

We have shown that 2006 was a prolific year for second and foreign language education Ph.D.s from English universities with very wide range of substantive areas and methodological approaches. However, a few key areas did not emerge in that year, such as: process–product research into teaching languages in primary schools; studies relating to task-based teaching and learning; research into content and language integrated learning; investigations of relations between socio-cultural awareness, attitudes and language learning; and psycholinguistic research into how input is processed by learners, such as parsing, segmentation or attention and related pedagogical issues. We might also hope for more studies directly related to foreign and L2 education in the UK, particularly among primary and secondary school pupils. Finally, although generativist perspectives were well-represented in the initial group of theses, it would be useful if doctoral studies could connect more clearly with other specific theories of L2 acquisition and could be designed to test, for example, theories from cognitive, functional and socio-cultural perspectives (at least, those which are testable).

## References

- Al-Ghamdi, H. (2006). The effectiveness and appropriateness of a medical English course. Ph.D. dissertation, University of Essex.
- Andersen, R. W. & Y. Shirai (1994). Discourse motivations for some cognitive acquisition principles. *Studies in Second Language Acquisition* 16, 133–156.
- Athanasopoulos, P. (2006). The effects of second language acquisition on cognition: The case of number. Ph.D. dissertation, University of Essex.
- Bavendiek, U. (2006). Portfolios and learner autonomy: The case of undergraduates learning German. Ph.D. dissertation, University of Liverpool.
- Bin Yusof, Y. (2006). The teaching of Arabic language at the International Islamic University Malaysia: The instructors' and students' perceptions. Ph.D. dissertation, University of Exeter.
- Brereton, A. (2006). Opportunities for participation: Sign language use with hearing children in an early years classroom. Ph.D. dissertation, University of Cambridge.
- Carlson, A. (2006). Computers, literacy and the bilingual/bicultural child. Ph.D. dissertation, University of Kent.
- Cheng, H-F. (2006). Motivational teaching practice of Taiwanese English teachers. Ph.D. dissertation, University of Nottingham.
- Ciobanu, D-I. (2006). Acquiring reading skills in a foreign language in a multilingual corpus-based environment. Ph.D. dissertation, University of Leeds.
- Coluzzi, P. (2006). Minority language planning and micronationalism in Italy: Three case studies. Ph.D. dissertation, University of Bristol.
- de Bot, K., T. S. Paribakht & M. B. Wesche (1997). Toward a lexical processing model for the study of second language vocabulary acquisition. *Studies in Second Language Acquisition* 19, 309–329.
- Doughty, C. & J. Williams (1998). *Focus on form in classroom second language acquisition*. Cambridge: Cambridge University Press.
- Edwards, P. (2006). Willingness to communicate among Korean learners of English. Ph.D. dissertation, University of Nottingham.

- Field, M. (2006). How do cultural expectations impact on the use of ICT for second language learning university students in Japan? Ph.D. dissertation, University of Cambridge.
- Garcia, S. S. (2006). Lexical strategies in L1 and L2 writing: A study with Mexican university EFL students. Ph.D. dissertation, University of Essex.
- Ginnis, P. (2002). *The teacher's toolkit: Raise classroom achievement with activities for every learner*. Bancyfelin: Crown House Publishing.
- Goglia, F. (2006). Communicative strategies in the Italian of Igbo-Nigerian immigrants in Padova (Italy): A contact linguistic approach. Ph.D. dissertation, University of Manchester.
- Grabe, W. & F. Stoller (2002). *Teaching and researching reading*. London, Longman.
- Granger, S. (2003). Error-tagged learner corpora and CALL: A promising synergy. *CALICO* 20, 465–480.
- Guo, X. (2006). Verbs in the written English of Chinese learners: A corpus-based comparison between non-native speakers and native speakers. Ph.D. dissertation, University of Birmingham.
- Hammerley, M. (1997). Educational research and teaching: A response to David Hargreaves' TTA lecture. *British Educational Research Journal* 23, 141–161.
- Hemmati, F. (2001). Vocabulary problems in the EFL writing of Iranian students: Taxonomies and strategies. Ph.D. dissertation, University of Essex.
- Hemmings, E. (2006). An exploration of the relationships between attitudes, learning styles and Year 7 pupils' experiences of language learning. Ph.D. dissertation, University of Nottingham.
- Hill, K. (2006). Sociocognitive metaphorm. Ph.D. dissertation, University of Nottingham.
- Hoey, M. (1991). *Patterns of lexis in text*. Oxford: Oxford University Press.
- Kasemsin, G. (2006). Interlanguage refusals: A cross-sectional study of Thai EFL learners' refusals in English. Ph.D. dissertation, University of Leeds.
- Kolb, D. A. (1984). *Experiential learning: Experience as source of learning and development*. Englewood Cliffs, NJ: Prentice Hall.
- Laufer, B. & P. Nation (1999). A vocabulary-size test of controlled productive ability. *Language Testing* 16, 33–51.
- Lee, H-C. (2006). Using children's literature in particular Arthur Ransome's *Swallows and Amazons* for teaching EFL in Taiwan. Ph.D. dissertation, University of Birmingham.
- Liang, S-C. (2006). Teaching English vocabulary in the primary classroom in Taiwan: A case study. Ph.D. dissertation, University of Essex.
- Lonsdale, M. dos S. (2006). Effects of the typographic layout of reading examination materials on performance. Ph.D. dissertation, University of Reading.
- Lu, Y. (2006). A discourse analytic study of EFL test-takers' spoken discourse competence and its impact on their oral proficiency and spoken grammatical competence. Ph.D. dissertation, University of Reading.
- Manners, L-M. (2006). A Medau approach to promoting language skills in pre-school children with SLI and EAL. Ph.D. dissertation, Liverpool John Moores University.
- Marsden, E. (2007). Can educational experiments both test a theory and inform practice? *British Educational Research Journal* 33.4, 565–588.
- McKenny, J. A. (2006). A corpus-based investigation of the phraseology in various genres of written English with applications to the teaching of English for academic purposes. Ph.D. dissertation, University of Leeds.
- Mertin, P. A. (2006). The role of culture in second language learning with special reference to the Japanese students at the international school of Düsseldorf, Germany. Ph.D. dissertation, Goldsmiths College, University of London.
- Naiman, N., M. Fröhlich, H. H. Stern & A. Todesco (1978). *The good language learner*. Toronto: Ontario Institute for Studies in Education.
- Niño, A. (2006). Evaluating the suitability of using raw machine translation output as input for foreign language written production. Ph.D. dissertation, University of Manchester.
- Norris, J. & L. Ortega (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning* 50, 417–528.
- Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. Boston: Heinle & Heinle.
- Pring, R. (2000). *Philosophy of educational research*. London: Continuum.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics* 11, 129–158.

- Schmitt, N., D. Schmitt & C. Clapham (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing* 18, 55–88.
- Sheen, Y. (2006). Corrective feedback, individual differences, and the acquisition of English articles by second language learners. Ph.D. dissertation, University of Nottingham.
- Skoufaki, S. (2006). Investigating L2 idiom instruction methods. Ph.D. dissertation, University of Cambridge.
- Souza, A. B. B. de. (2006). 'Should I speak Portuguese or English?' – Ethnic and social identity construction in the language choices of Brazilian mothers and their mixed-heritage children. Ph.D. dissertation, University of Southampton.
- Spada, N. (1987). Relationships between instructional differences and learning outcomes: A process–product study of communicative language teaching. *Applied Linguistics* 8, 137–161.
- Sternberg, R. J. (1997). *Thinking styles*. New York: Cambridge University Press.
- Stolyar, R. (2006). Project-based learning: Affect and effect. Ph.D. dissertation, Manchester Metropolitan University.
- Thepsiri, K. (2006). Scaffolding in project work: A study of a project-based course implementation in a Thai university context. Ph.D. dissertation, University of Newcastle upon Tyne.
- Thomas, G. (2005). *Strangers in paradigms: Education and theory in the twentieth century*. Maidenhead: Open University Press.
- Thomas, L. F. & E. S. Harri-Augstein (1985). *Self-organised learning: Foundations of a conversational science of psychology*. London: Routledge & Kegan Paul.
- Thorndike, R. L. & E. Hagen (1986). *Cognitive abilities test* (2nd edn.). Windsor: NFER-Nelson.
- Ting, K-Y. (2006). The effect of using multiple intelligences approaches and Internet resources in English Language courses for university students in Taiwan. Ph.D. dissertation, University of Exeter.
- Tonkyn, A. (2006). Towards progress-sensitive measures of L2 speaking proficiency: A case study of instructed learners. Ph.D. dissertation, University of Reading.
- Torgerson, C. & D. Torgerson (2003). The design and conduct of randomised controlled trials in education: Lessons from health care. *Oxford Review of Education* 29, 67–80.
- Tuncel, E. (2006). An evaluative study of the relationship between instructional differences and learning outcomes: A survey based and experimental study of a more communicative approach to language teaching in a traditional setting. Ph.D. dissertation, University of Essex.
- van Dijk, T. A. & W. Kintsch (1983). *Strategies of discourse comprehension*. London: Academic Press.
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Walters, J. M. (2006). Beyond exhortation: Training students of English as a second language to infer meaning from context. Ph.D. dissertation, University of Nottingham.
- Ward, J. W. (2006). The lexical aspect of reading English as a foreign language for engineering undergraduates. Ph.D. dissertation, University of Birmingham.
- Wiliwan, S. (2006). Lexical cohesion and metacognitive strategy training: An integrated approach to main idea comprehension. Ph.D. dissertation, University of Leeds.

EMMA MARSDEN is a lecturer in L2 education in the Department of Educational Studies at the University of York, England, UK. Her work includes investigating L2 grammar pedagogy, morphosyntactic and lexical development, and input processing. Her publications have appeared in *Language Learning*, *Second Language Research*, *The British Educational Research Journal*, *Language Learning Journal*, and several book chapters. She was a founding co-convenor of the British Association of Applied Linguistics' Language Learning and Teaching Special Interest Group.

SUZANNE GRAHAM is Senior Lecturer in Language and Education at the Institute of Education, the University of Reading, England, UK. Her publications include articles in a range of journals (e.g. *Language Learning*, *The Modern Language Journal*, *System*, *Language Learning Journal*), books and systematic reviews in the areas of language learner strategies, motivation and listening comprehension. She is a member of the Editorial Board for the *Language Learning Journal* and has acted as reviewer for a number of international journals in the field of applied linguistics. She is an active member of the UK Project on Language Learner Strategies research group.