

# Analyzing recent CALL effectiveness research: Towards a common agenda.

## **Abstract**

*This paper presents a comprehensive picture of what has been investigated in terms of CALL effectiveness over the period 2000-2004 and determines why this question is still such a difficult one to answer. The study looks at both strengths and weaknesses in this body of work and singles out models of good design practice. While a perfect research design remains elusive, a positive trend towards combining quantitative and qualitative methods and including learning processes in the investigation can be observed. There is also a move away from simple comparison studies (teaching with or without computers) in favor of complex designs including a variety of variables under investigation and comparing different types of CALL. In this context, the usefulness of meta-analyses to determine clear cause and effect results, relying solely on effect sizes, appears questionable. The author argues that a series of systematic syntheses of findings related to one particular variable such as learning strategy or writing quality might produce more valuable insights into the potential impact of technologies on learning processes and outcomes and proposes a research agenda along those lines.*

## **Introduction**

The effectiveness of computers in education has been a long-standing concern (see Dunkel, 1991), recently emerging as a major theme in a comprehensive survey of unanswered questions in CALL by Hubbard (2003). Among the large number of invited international experts who responded to Hubbard's survey, only a handful proposed valid research designs to answer their particular questions related to effectiveness (which was part of the overall brief). There is obviously still a pressing need among the educational research community for models of successful designs in this area. This is especially the case for postgraduate students embarking on studies in this important but hitherto under- (or ill-) researched field. There is also an urgent need for educators and educational institutions to understand the full extent of the impact of ICT on learning.

One of the problems in determining the effectiveness of ICT in general is that research to date is relatively scarce (compared with other disciplines) and lacks scholarly rigour (Dunkel, 1991; McIssac & Gunawardana, 1996; Soe, Koki, & Chang, 2000; Windschitl, 1998; Zhao, 2001). The difficulty of generalizing results characteristic of the previous generation of Computer Assisted Learning (CAL) and CALL continues to exist (see Chapelle, 1997; Debski, 2003; Levy, 2002), and not surprisingly such research has produced ambiguous results (Dunkel, 1991; Miech, Nave, & Mosteller, 1996). Judgments in the area vary widely. At one end of the spectrum are early positive reports from the authors of several large meta-analyses of CAL (Kulik, Kulik, & Cohen, 1980:538): "the computer did its job quickly – on average in about two-thirds the time required by conventional teaching methods"; and of CALL, where Basena & Jamieson (1996:19) stated that "the newer technologies show promise to be able to provide feedback in multiple modes, such as listening and

reading”, even though they cautioned that “the results are difficult to interpret, and the designs and measures do not lend themselves to reproduction or generalizability”. At the other end are dismissive, unsubstantiated comments such as: "Study after study seems to confirm that computer-based instruction reduces performance levels and that habitual Internet use induces depression" (Noble, 1998:2).

The most often cited collection of research results, claiming no significantly different findings between classroom and distance learning using ICT (Russell, 1999), includes 355 research papers. However, the book itself raises questions about poor research designs – problems which have long been pointed out by educational researchers (Clarke, 1983, 1985; Joy II & Garcia, 2000; Maddux, 1995).

Key shortcomings of many investigations are outlined in a review of the effectiveness of distance learning in higher education in the US (Phipps & Merisotis, 1999:3-4) which echo concerns expressed by Chapelle and Jamieson (1991) in relation to CALL:

- “Much of the research does not control for extraneous variables and therefore cannot show cause and effect.”
- “Most of the studies do not use randomly selected subjects.”
- “The validity and reliability of the instruments used to measure student outcomes and attitudes are questionable.”
- “Many studies do not adequately control for the feelings and attitudes of the students and faculty – what the educational research refers to as ‘reactive effects’.”

Naturally, one can go too far in the demand for the application of rigorous conditions to educational research, and the above study has been criticized for doing precisely that (Brown & Wack, 1999). After all, if we managed to control for every possible confounding variable in an experimental design we would be left with the technology itself as the independent variable, when in today’s learning environments this is inextricably linked to the instructional method and the context in which the learning takes place (see Chapelle & Jamieson, 1991; Clarke, 1985 for detailed discussions).

The most conclusive results so far can be gained from sound published studies on the relationship between computer use and academic achievement (most of these have examined performance in science and mathematics). They indicate that technology use can improve outcomes (Becker, 1994; Christmann & Badgett, 1999; Hativa, 1994; Kozma, 1991; Kulik & Kulik, 1987; Liao, 1992; Niemiec & Walberg, 1987, 1992; Ryan, 1991; Van Dusen & Worthren, 1994). In their research synthesis on computer-based instruction (CBI), for example, Niemiec and Walberg (1992) calculated a positive average CBI effect on achievement of 0.42 standard deviations. Ryan (1991) computed a mean effect size (see Appendix B for definition) of 0.31 in a meta-analysis of 40 published and unpublished studies on computer use and achievement in elementary schools.

Only three meta-analyses specifically related to CALL were located in our search. Zhao (2001) analysed nine studies on various achievement outcomes, such as vocabulary, grammar and writing in research conducted between 1997 and 2001 in tertiary settings. Soe et al (2000) analysed results of 17 studies investigating the single

variable *reading* in school settings between 1982 and 1999. Blok et al (2001), reported in great detail on five studies which evaluated word learning programs for elementary school students. While Zhao and Soe et al reported overall positive effects associated with the various technologies described, Blok et al reported no significant effect for four of the five studies in their meta-analysis.

However, since newer, constructivist applications of educational technology are designed to develop higher order thinking and problem-solving skills, learning outcomes cannot readily be assessed through standardized tests. Rumberger (2000) suggests that such skills are best assessed through performance-based measures carried out over extended periods of time (Lesgold, Lajorie, Brunzo, & Eggen, 1992). But establishing the reliability and validity of such assessments is problematic, and even performance-based assessments may not provide an appropriate measure of the lifelong learning that may result from these newer forms and uses of educational technology (Brown, Bransford, Ferrara, & Campione, 1983). It is therefore imperative to also look at studies that focus on the processes of learning rather than outcomes alone. This became apparent during the author's work at the Open University in the UK in the process of designing a collaborative research project aimed at ascertaining the impact of a voiced online tutorial (Lyceum) on language learning. Whereas we originally thought it would be possible to measure outcome effects of the audiographic device under investigation, it became clear very early in the discussions that what was occurring during the process of using such a device was of equal importance and also needed to be recorded and analyzed.

Naturally this complicates research even further, and it is not surprising that this problematic field is dominated by confusion and uncertainty. There are very few studies to date which set out to investigate ICT effectiveness related specifically to learning processes (see also Dunkel, 1991), and most of those located in our search have restricted their investigations to gauging students' perceptions on the variables under investigation. While these studies are valuable in their own right, they do not however, provide conclusive evidence on ICT effectiveness.

The purpose of our study is to:

- Throw light on a continuously confusing body of research and determine why the question of ICT effectiveness is still such a difficult one to answer.
- Identify strengths and common problems in the research.
- Single out valid designs that may serve as models for future research.

Our aim is to:

- Produce a comprehensive picture of what exactly is being investigated in the area by categorizing studies according to research design, technology used, educational setting, subject taught and variables under investigation. Because the body of research proved to be rather substantial, we have limited our investigations here to studies concerned with CALL in the broadest sense over the last four years (earlier studies are discussed in detail in Dunkel, 1991; Kulik, 2003; Miech et al, 1996). It is hoped that this work will be of use to researchers in planning their studies on the basis of what has already been done in their area of interest and thus strengthen existing findings.

Work of this nature is delicate in that there is some potential of misinterpreting information, especially relating to studies where scarce context information is given. We would like to state emphatically that all care was taken in checking our facts, and that our motivation at all times was to produce a piece of helpful research by dealing with the body of work in a constructively critical manner. Since our investigations are ongoing and expanding, we welcome feedback of any kind.

### ***Definition***

The Webster dictionary defines effective (from *efficere*) as having an effect, producing a result, bringing something to pass. Efficacy (synonymous with effectiveness) is defined as the power to produce effects or intended results. This suggests a strong causal relationship between an intervention, such as the use of a particular item of technology in a learning situation and a discernible change in the learning process, the learning climate or the learning achievement. Definitions given in the other major dictionaries support this view.

### ***Literature Search***

A number of resources were used to search for studies relevant to this project. Searches on the following computerized databases were conducted: Proquest, Educational Resources Information Centre (ERIC), Expanded Academic ASAP, Educational Research Abstracts Online (ERA), Linguistics and Language Behavior Abstracts, CSA Internet Database Service. In addition, the search engines Google and Vivisimo were used to locate studies. Key words used for these searches were “effectiveness AND computer assisted language learning”, “effectiveness AND language learning”, “effectiveness AND multimedia”, “effectiveness AND Internet”, “effectiveness AND World Wide Web”, “effectiveness AND www”. Websites of well known journals on CALL, such as CALL-EJ Online, CALICO, ReCALL and Computer Assisted Language Learning, were also searched. Finally, studies already known to the researcher were included in the analysis. These searches resulted in more than 150 studies. Many of these were discarded because they were concerned with learning in areas other than CALL or did not investigate an aspect of effectiveness of ICT on language learning. The final number of studies included here was 52.

### ***Criteria for inclusion***

To be included in this analysis a study had to be published after December 1999 and to report the results of research about the effectiveness of the use of ICT on language learning processes or outcomes in the widest sense. Since our goal was to give a comprehensive overview of types of research conducted, we did not cull these studies to include only those with sound research designs. The search for this write-up was concluded in December 2004.

### ***The file drawer problem***

A limitation of this project is that it deals exclusively with published research studies and hence may have missed important information contained in unpublished work. Furthermore, there is some suspicion among statisticians and researchers that more studies have been published that contain significant findings than studies reporting non-significant results, thus presenting biased results of actual research carried out

(see Rosenthal, 1979; Soe et al, 2000; Zhao, 2001). Again, we would be grateful for any important additions to the body of work included here.

## **Categorization of Studies**

Our aim was to provide the widest possible breadth of information on the studies included. Because of limitations of table space and the great variation in the many characteristics of each study, this proved tremendously difficult, especially where variables under investigation were concerned. In order to present so much information in an intelligible table format we have sometimes had to group several variables under an umbrella category. For example, many studies looked at different aspects of *Reading* (*close reading, critical reading, reading comprehension* etc). These have all been grouped together as *reading skills*. Other variables, such as *Concentration Level, Use of L1 Translation, and Student Anxiety*, were the subject of investigation in one study only. These appear in Table 3 as “Other”. (Appendix A gives the most detailed overview). The major categories chosen are:

- Number of participants
- Research design used
- Technology used
- Educational setting
- Language taught
- Subject/skill taught
- Variable under investigation

We would have liked to include *instructional method* and *context* (i.e. how well supported the use of ICT was in the particular setting), but information on these very important elements remains very scarce indeed. (see also Hubbard, 2004; Jamieson & Chapelle, 2004; Levy, 2004)

## **Overview and Strength of the Effectiveness Research**

Taking into account earlier studies (see Basena & Jamieson, 1996; Dunkel, 1991; Soe et al, 2000) we are beginning to see a substantial body of qualitative and quantitative data related to CALL effectiveness. Our search produced 52 studies published over the past four years alone (see Appendix D for a distribution of studies in Journals). While concerns expressed by Phipps & Merisotis (1999) regarding the lack of random assignment of subjects and control for extraneous variables are still present they are no longer sustainable to the same extent. We are finding an increasing number of sound quasi-experimental and experimental studies, a substantial number of which include randomly selected control groups (see Table 1).

In our sample the distribution between experimental (including quasi-experimental) and non-experimental (including pre-experimental) studies is 24 to 29 (one study used 2 research designs). This is indeed a strength of the research since both categories include a large number of excellent studies and their combined results advance the field more substantially than a continuous series of highly controlled studies. This is especially the case since the latter are predominantly very short-term (as brief as a 40 minute treatment) and most often investigate a single variable.

[INSERT TABLE 1 HERE]

As predicted by Dunkel (1991), there is a noticeable move towards the inclusion of learning processes (see also Collentine, 2000). This is not surprising since teaching methods and technologies have changed in a way that no longer allows for easy numeric measurement of outcomes. If we look at the 16 studies that investigate writing (see Table 2), for instance, 9 deal with writing quality (see Table 3) and are spread fairly evenly across the spectrum of research designs. While it is not the scope here to synthesize these results, it was our intention to provide opportunities for this. Building up a body of clustered research results of this nature ought to give more detailed insights into the impact of ICT on any particular process or outcome. Looking down the list of variables in Table 3 and Appendix A shows that there is a distinctly higher percentage of investigations concerned with quality and process measures than with quantity. This trend is also reported by Hoven (2004).

[INSERT TABLES 2 & 3 HERE]

Although investigations in tertiary settings still dominate the field (which was also reported by Liu et al (2002) who reviewed CALL studies pre 2000), we are seeing an increased number of studies emerging in the school environment (11.5% in our sample, excluding the 17 contained in the meta analysis of Soe et al 2000). This, too, lends strength to the body of research since it provides important opportunities for investigating potentially differential impacts of ICT on learning in elementary, secondary and tertiary settings.

We are also beginning to see languages other than those that have always been prominent in the field in CALL emerging, i.e. Japanese, Indonesian, Chinese and Russian (see Table 4). Interestingly the distribution appears unchanged from when we looked at Web-based resources (Felix 2001) with ESL/EFL, Spanish and French far outweighing any other language. However, the information given in Table 2 shows a useful spread of languages included in investigations of many variables.

[INSERT TABLE 4 HERE]

## ***Common Problems in the Effectiveness Research***

### **Misleading titles**

Many studies claim to be investigating effectiveness of a form of ICT but are really looking at its viability as a teaching and learning environment on the basis of surveys of student perceptions. Titles can often be very misleading when they include the keyword effectiveness. For example, the otherwise interesting study by Dolle and Enjelvin (2004) is entitled *Investigating “VLE-effectiveness” in languages*, when a more representative title would have been *Investigating student perceptions of a VLE as a viable learning environment*. Since no measurements or comparisons of learning outcomes were carried out the study cannot produce information on “key factors in enhancing learning effectiveness in a VLE”, as the authors claim at one point (p 486). Instead it gives valuable insights into the students’ perceptions of the teaching and learning environment effected and affected by the technologies. It is one of the few studies that describes in detail the instructional methods and the technologies used. In studies like these it is important to remember (and the authors themselves list a number of limitations) that students’ statements such as “formally assessed Web-

based tasks have improved my overall coursework grade” (p 485) cannot be meaningful in the absence of any form of controlled comparison. The fact that the statistical analyses established a significant relationship between this statement and the one that said “assessed Web-based tasks encouraged me to engage more in the learning process” (p 485), does not establish enhanced learning effectiveness but that students think they are doing better. Although this information is useful and interesting, there is no way of knowing whether a different kind of intervention would not have produced the same result, even without possible Hawthorne and Pygmalion effects (see Appendix B for definition) being at play.

### **Poor description of the research design**

A thorough description of the procedures, including information on the subjects, materials, technologies, treatments, tests, statistical analyses and anything else pertinent to the particular investigation is absolutely essential in rigorous research practice (Nutta et al., 2002 is an excellent example here). However, this is far from standard practice, and it is sometimes surprising how studies in which procedures are not fully explained get published. For example, a great limitation of the otherwise thorough study by Chikamatsu (2003) is the lack of information about the subjects. In a design where the effects of computer use on Japanese writing is investigated, information on students previous experience of computer-assisted writing, as well as their IT literacy is crucial. The study also gives no background as to current or previous teaching methods and uses of CALL in the context of comparing handwritten and computer written tests when some or all of these constitute potential threats to internal validity. Another study (Myers, 2000) provides so little information on the research design of a study reporting on the effects of voice recognition software on perception and speech production in second language use that it was impossible to fit it into our classifications grid.

### **Failure to investigate previous research**

A surprisingly large number of studies do not begin with a thorough investigation of what has already been done in the area to be studied. Notable exceptions are the excellent review on *writing and CALL* in Chikamatsu (2003) and *reading and CALL* in Soe et al (2000). This often leads to duplications of efforts when building on and incorporating existing findings would advance the field more substantially. It is most surprising to find a statement such as the following when investigating student perceptions of CALL represents the largest area of activity to date (see Table 3).

*In spite of the widespread use of computer-assisted language learning (CALL) and its perceived facilitative role in second language (L2) learning, there is little data on how learners feel, experience, or think about CALL in the L2 learning context. (Suh, 2002:669)*

### **Poor choice of variables to be investigated**

There are a number of studies comparing excellent CALL activities with poor non-CALL exercises or inferior CALL. This is especially disappointing in studies where much effort has been invested in designing well-controlled procedures. An interesting case in point here is the project by Yeh and Lehman (2001) which we had singled out as one of the best experimental designs in the current body of work. Among others the study looked at the important variable of *learner control* in CALL but what was compared was a system that allowed students a great deal of freedom in viewing and repeating video and text segments in any order and check a glossary with a system in

which pace and sequence were completely fixed and repetition was not possible. The study simply confirmed what we already know through a relatively large body of literature when the rigorous design would have allowed for more interesting probing of learner control.

### **Overambitious reporting of results**

Many studies still lack a discussion of limitations which, in an environment where perfectly controlled designs are near impossible, is most surprising. There is a definite trend for excellent designs being accompanied by detailed discussions of possible confounding variables and cautious reporting of findings (even overly cautious as in 1. below) and for poor designs not recognizing threats to internal or external validity and reporting results in very certain terms, such as “this study proves” or “as has been demonstrated, shown” and so on. Since numerous books and articles have been written on the subject, we will not include here yet another outline of potential threats to the validity of findings but refer the reader to the detailed discussion in Chapelle and Jamieson (1991). A simple check of what other than the particular *treatment* might have an influence on results would take researchers a long way in setting up studies and writing up results realistically.

### ***Design Models***

#### **General comments**

Because there is such a large scope for research in this area, there cannot be a single best design model. What is imperative, though, is that researchers match the design to the research questions, the context in which the study takes place, the time-frame available, the variables under investigation, their capacity of statistical analyses and their ability to control for confounding elements. A short-term fully controlled experimental design, for instance would be suitable to measure individual well defined outcome effects (see 1. below), while a longer-term non-experimental study using qualitative measures such as observational procedures and think-aloud protocols would yield important data related to effects on learning processes. A combination of various data collection methods within one single study will help in strengthening confidence levels about results (see 2. below).

We have singled out below examples of good design practice in four distinct and important areas including both outcomes and processes: (1) a study of differential effects of multimedia elements on vocabulary learning; (2) an investigation of the effectiveness of CALL on language proficiency; (3) a project examining whether threaded discussions can be effective in realizing constructivist principles; and (4) a case study examining the role of visually rich technology in facilitating children’s writing.

#### **1. Within-Subject Design**

Al-Seghayer (2001) used a within-subject design (n=30) to test the effect of multimedia annotation on vocabulary acquisition under three conditions: printed text definitions; printed text definitions linked to still pictures; and printed text definitions linked to video clips. Data was gathered in a variety of complementary modes. In qualitative terms, students were asked to complete a questionnaire and a face-to-face interview, giving their own view of which condition might be most conducive to vocabulary learning or conveying of meaning. In quantitative terms, a recognition and

a production test was administered and data processed by analysis of variance procedures. Results of both qualitative and quantitative data led to the conclusion that video clips are more effective than still pictures in teaching unknown vocabulary.

This study represents a sound approach to the problem. It contains an extensive and useful literature review of related studies and sets out to fill a gap in the research. While the review gives strong evidence for the effectiveness of both still pictures and video clips in a variety of language learning activities, the new study represents the first attempt to compare the two modes experimentally.

A great deal of care was taken in controlling for confounding variables and in describing procedures in every detail. Each subject served as their own control by taking part in each of the conditions. The annotated items of vocabulary used in each condition were controlled for frequency, grammatical category, morphological category, visual complexity and for whether they represented abstract or concrete concepts. An unknown reading passage was chosen and adapted for intermediate ESL learners applying the criteria *text length*, *syntactic complexity* and *content* and pre-tested with subjects not taking part in the study.

The author outlines the following limitations: (1) The small sample sheds doubt on the validity of the observed significance; (2) Assessment of the learning outcome was measured only with multiple choice and production tests; (3) The study did not analyse individual performance data such as *study path* or *reaction time*; and (4) Only short-term retention was studied (p. 227). It would therefore be interesting to replicate this study with many more subjects, and specifically address these limitations. Another useful measure to add in a larger study would be to carry out a correlation analysis between the quantitative findings and the students' perceptions of the most effective mode. Could it be that they performed better in this mode because they believed they would?

## **2. Experimental Design**

The study by Nutta et al (2002) is an excellent example of a well-designed experimental investigation. It is one of the very few that was carried out over an extensive period of time (three hours by 25 weeks for the experimental treatment, 13 months in all including an equal instruction period and collection of qualitative data) and also one of a limited number conducted in an elementary school setting (n=28). The study set out to test the effect of multimedia materials on proficiency in Spanish. Students were assigned at random to a treatment and a control group with the same instructor carrying out the teaching in both. A great deal of care was taken in providing students with near identical activities.

The extensive quantitative measures included pre- and posttests on proficiency, scored by an independent native speaker, and a criterion-referenced posttest on achievement, including oral and written items, developed by another independent researcher. The achievement test was checked for internal consistency of the written and oral sections combined and also re-administered in delayed mode. Qualitative measures were equally extensive, albeit limited to a smaller number of subjects for some procedures, and consisted of whole-group observations, interviews with students, teachers, administrators and parents and think-aloud protocols. Because of the nature of the qualitative investigations this study gives more detailed information

on instructional methods and context than is usually found in experimental studies, especially those that are limited to a one-off treatment. The interviews with parents, for instance, supported the often voiced ethical dilemma in this sort of design in which students might be deprived of a potentially better learning environment. One parent noted “My child says that she does not get to use the computer” (p. 304).

No statistically significant difference was found at posttest between the experimental group and the control group (who had used printed and audio materials). A significant difference, however, was reported on the delayed posttest in favour of the treatment group. The qualitative investigations “detected differences in language behavior, with the students who used multimedia spending more time to stop, check, and revise their language production, leading to greater precision in pronunciation and the use of larger chunks of language when repeating phrases” (p. 293).

The authors point out that their results are not generalisable because of the small number of case studies in the qualitative investigations. They also discuss potentially influencing factors such as attrition, IT failures and the later time in which experimental students were taught. There are, of course other potential threats to internal validity, such as the teacher inadvertently treating the experimental group differently and students having differential exposure to learning opportunities outside the experiment. However, such effects would more likely have been reflected in the immediate posttests rather than the delayed one (unless, of course, all the extra exposure had taken place in the period between the two tests and was exclusive to the experimental group).

A serious problem of the study is the decision to set the alpha level for statistical significance at .10. This is highly unusual, it is normally set at either .01 or .05. Had this been observed the results on the delayed test (given as  $U=.071$ ) would also not have been significantly different. [The U value is meaningless on its own and can only be a whole number. The alpha value, (or the two-tailed significance value) is the important value and this must be less than .05 to be significant (see Gravetter & Wallnau, 2000:640-648).

The authors’ discussion on possible reasons for the superior performance in the delayed tests, supported by previous findings in the area, is nevertheless interesting and worthy of further research.

### **3. Situated Study Design**

Weasenforth et al (2002) conducted an interesting investigation over three semesters examining whether threaded asynchronous discussion group activities might be effective in realizing constructivist objectives (as defined by Bonk & Cunningham, 1998) in the context of postgraduate ESL ( $n=52$ ). This is a descriptive (situated) study in which close observations of students’ performance on frequent task-based assignments, coupled with student surveys, led to systematically documented changes in procedures and instructors’ roles (interventions) in order to realize the constructivist learning objectives which formed the framework of the study.

Naturally a project of this nature cannot claim cause and effect results in terms of achievement. It does, however, provide valuable insight into the extent to which asynchronous discussions, as mediated by tutors, might promote various social and

cognitive skills as well as address affective factors and motivational differences in the students. The study also contains useful models for group discussion assignments and evaluation forms.

#### **4. Case Study Design**

This study by Vincent (2001), conducted over five weeks in an elementary school (n=6), investigates the impact of a multi-media software called *MicroWorlds* on writing proficiency of children who strongly favour a visual learning style. It is an excellent example of what can be done to increase validity in a study with a very limited number of subjects and with so much scope for outcomes having been produced by elements other than the treatment. Procedures are described in great detail. Participating children were selected by rigorous selection criteria including scores from recognized (and referenced) visual, verbal and spatial tests, interviews with children and some parents and a log of classroom observations. Five children were identified as fulfilling the criteria for selection and one child performing at the opposite end of the scale, i.e. favouring a strongly verbal style, was selected as the “negative case”. Assessment of the narrative writing tasks during the treatment period followed the standardized criteria devised by the Victorian Board of Studies.

The children took part in three different writing activities: (1) writing without a visual input in which they wrote by hand with the final version produced on a word processor; (2) setting up a writing task with a drama stimulus which included some use of graphics and animation with children given the choice to present their work either on paper or as a *MicroWorlds* presentation; and (3) a task in which it was compulsory to work with *MicroWorlds* software. In addition to the formal assessments of the tasks, the teacher also kept an observational journal yielding interesting information of the differences in attitude towards the tasks between the children. The design is somewhat similar to 1. above, an added problem here, however, is the cumulative learning effect (a standard threat to within-subject designs). While the researcher does not consider this in the discussion of the results which supported activity (3), only very careful and tentative conclusions are put forward. Studies such as these are important starting points for larger experimental investigations.

#### ***Concluding Remarks and Suggestions for Future Research***

Judging from the body of research included here, a perfect research design for studying the effectiveness of ICT on language learning remains elusive. This is, of course, not surprising since the many extraneous variables present in natural educational settings have not decreased since criticisms of poor methods were voiced more than a decade ago but have increased even further. Technologies, settings and teaching methods have not only become more complex but inextricably linked and outcomes therefore even more difficult to measure. The ever pursued question of the impact of ICT on learning remains unanswerable in a clear cause and effect sense. There exists a noticeable trend away from the comparison studies carried out during the 1980s that wanted to find out whether teaching with computers was better than teaching without them. One of the reasons for this is surely the difficulty of carrying out valid research of this kind in natural settings. The most obvious reason, though, is that in an environment where computers have become a natural part of the educational

experience and in which we have learnt that teachers will not be replaced by them, the question is no longer as interesting. What remains interesting to investigate is how technologies are impacting learning processes and as a consequence might improve learning outcomes. It is worth noting that we are beginning to see studies that compare the differential impact of CALL methods (Bell & LeBlanc, 2000; Lo, Wang, & Yeh, 2004; Ma, Nicolas, Gierts, & Vandamme, 2002; Smith, Alvarez-Torres, & Zhao, 2003) rather than looking at overall effectiveness of CALL.

We are beginning to see enough data in CALL that suggest positive effects on vocabulary development, reading and writing (the last is strongly supported by Kulik's latest meta-analysis in 2003 and by Liu 2002). There is also a substantial body of data that indicates that student perceptions of CALL are on the whole positive, provided technologies are stable and well supported. In these studies salutary effects on motivation, computer literacy, target culture awareness, reading and listening skills, and classroom climate in terms of enjoyment and comfort have been observed. On the negative side there are still concerns about technical difficulties interfering with the learning process; older students not feeling comfortable with computers; younger students not possessing the necessary metaskills for coping effectively in these challenging environments; training needs in computer literacy for both students and teachers; problems with group dynamics; and time constraints (Felix, 2004). It is sometimes forgotten that some or all of these are likely to have an influence on research results.

Now that we can examine clusters of research investigating the same or similar variables in a variety of ways, we are in a better position to ascertain how a specific item of ICT might impact a specific environment, outcome or process. A lot remains to be done, though. We need to build on existing knowledge, re-investigate established findings in different settings, replicate excellent studies using more subjects, and design sound new projects in areas and languages that have not yet been included. A design that is vastly underused in CALL research but highly recommended for longer term studies is *time-series analysis* in which the same group of students is involved in the experimental and control treatment for a certain amount of time and then switched – more than once if possible (see Felix, 1989; Warschauer, 1996). Many more studies are needed in the school environment and in the Vocational Education and Training (VET) sector. We also recommend inclusion of delayed tests in designs where achievement tests are carried out, fully recognizing, of course, that it is sometimes difficult to get access to the same students.

While no single study, nor any meta-analysis on its own can so far give a definitive answer on ICT effectiveness, a series of systematic syntheses of findings related to one particular variable such as learning style or writing quality might produce more valuable insights into the potential impact of technologies on learning processes and outcomes. These would need to incorporate qualitative findings rather than rely on effect sizes alone. An approach like this would begin to establish a research agenda in ICT effectiveness rather than continue the series of isolated single studies on different topics from which it is difficult to draw firm conclusions (see Miech et al 1996 in Kulik 2003:34-35). We hope that this project has provided a small impetus towards this goal.

## References

- Al-Seghayer, K. (2001). The Effect of Multimedia Annotation Modes on L2 Vocabulary Acquisition: A Comparative Study. *Language Learning & Technology*, 5(1), 202-232.
- Basena, D., & Jamieson, J. (1996). CALL Research in Second Language Learning: 1990-1994. *CAELL*, 7(1/2), 14-22.
- Becker, H. J. (1994). Mindless or mindful use of integrated learning systems. *International Journal of Educational Research*, 21, 65-79.
- Bell, F. L., & LeBlanc, L. B. (2000). The language of glosses in L2 reading on computer: Learner's preferences. *HISPANIA*, 83(2), 274-285.
- Blok, H., van Daalen-Kapteijns, M. M., Otter, M. E., & Overmaat, M. (2001). Using Computers to Learn Words in the Elementary Grades: An Evaluation Framework and a Review of Effect Studies. *Computer Assisted Language Learning*, 14(2), 99-128
- Bonk, C., & Cunningham, D. (1998). Searching for learner-centered, constructivist, and sociocultural components of collaborative educational learning tools. In C. Bonk & K. King (Eds.), *Electronic collaborators* (pp. 25-50). Mahwah, NJ: Lawrence Erlbaum.
- Brown, A. L., Bransford, J. D., Ferrara, R., & Campione, J. (1983). Learning, remembering, and understanding. In J. Flavell & E. Markman (Eds.), *Mussen handbook of child psychology* (pp. 77-166). New York: Wiley.
- Brown, G., & Wack, M. (1999). The Difference Frenzy and Matching Buckshot with Buckshot. *The Technology Source*, May.
- Chapelle, C. (1997). CALL in the year 2000: Still in search of research paradigms? *Language Learning and Technology*, 1(1), 19-43.
- Chapelle, C. C., & Jamieson, J. J. (1991). Internal and External Validity Issues in Research on CALL Effectiveness. In P. Dunkel (Ed.), *Computer-Assisted Language Learning and Testing*. (pp. 37-59). New York: Newbury House.
- Chikamatsu, N. (2003). The effects of computer use on L2 Japanese writing. *Foreign Language Annals*, 36(1), 114-127.
- Christmann, E. P., & Badgett, J. L. (1999). A comparative analysis of the effects of computer-assisted instruction on student achievement in differing science and demographic areas. *Journal of Computers in Mathematics and Science Teaching*, 18, 135-143.
- Clarke, R. E. (1983). Reconsidering Research on Learning from Media. *Review of Education Research*, 53(4), 445-459.
- Clarke, R. E. (1985). Evidence for Confounding in Computer Based Instruction Studies: Analysing the Meta-Analyses. *Educational Technology Research and Development*, 33(4), 235-262.
- Collentine, J. (2000). Insights into the Construction of Grammatical Knowledge Provided by User-Behavior Tracking Technologies. *Language Learning & Technology*, 3 (2). Retrieved 12/7/2004, from <http://lt.msu.edu/vol3num2/collentine/>
- Debski, R. (2003). Analysis of research in CALL (1980-2000) with a reflection on CALL as an academic discipline. *ReCALL*, 15(2), 177-188.
- Diamond Bullet. *Usability Glossary*. Retrieved 16April, 2004, from [http://www.usabilityfirst.com/glossary/main.cgi?function=search\\_page](http://www.usabilityfirst.com/glossary/main.cgi?function=search_page)
- Dolle, J., & Enjelvin, G. (2004). Investigating "VLE-effectiveness" in Languages. *Computer Assisted Language Learning*, 16(5), 469-490.

- Dunkel, P. (1991). The effectiveness of research on computer-assisted instruction and computer-assisted language learning. In P. Dunkel (Ed.), *Computer-Assisted Language Learning and Testing*. (pp. 5-36). New York: Newbury House.
- Felix, U. (1989). *An Investigation of the Effects of Music, Relaxation and Suggestion in Second Language Acquisition in Schools*. Unpublished doctoral thesis, Flinders University, Adelaide, Australia.  
<http://www-personal.monash.edu.au/~ufelix/accel.htm>
- Felix, U. (2001). *Beyond Babel: Language Learning Online*. Melbourne: Language Australia.
- Felix, U. (2004). A Multivariate Analysis of Secondary Students' Experience of Web-Based Language Learning. *ReCALL*, in press.
- Gravetter, F. J., & Wallnau, L. B. (2000). *Statistics for the behavioral sciences* (5th ed.). Australia ; London: Wadsworth/Thomson Learning.
- Hativa, N. (1994). What you design is not what you get (WYDINWYG): Cognitive, affective, and social impacts of learning with ILSC An integration of findings from six-years of qualitative and quantitative studies. *International Journal of Educational Research*, 21, 81-111.
- Hoven, D. (2004). *Methods in our madness? Re-examining the methodological frameworks of our field*. Paper presented at the Eleventh International CALL Conference, University of Antwerp.
- Hubbard, P. (2003). A survey of unanswered questions in CALL. *CALL*, 16(2-3), 141-155.
- Hubbard, P. (2004). *Some Subject, Treatment, and Data Collection Trends in Current CALL Research*. Paper presented at the Eleventh International CALL Conference, University of Antwerp.
- Jamieson, J., & Chapelle, C. (2004). *Issues in CALL Evaluation*. Paper presented at the Eleventh International CALL Conference, University of Antwerp.
- Joy II, E. H., & Garcia, F. E. (2000). Measuring Learning Effectiveness: A new Look at No-Significant-Difference Findings. *Journal of Asynchronous Learning Networks*, 4(1), 33- 39.
- Kozma, R. B. (1991). Learning with media. *Review of Educational Research*, 61, 179-211.
- Kulik, J. A. (2003). *Effects of Using Instructional Technology in Colleges and Universities: What Controlled Evaluation Studies Say* (No. P10446.003). Arlington, Virginia.: SRI International.
- Kulik, J. A., Kulik, C. C., & Cohen, P. A. (1980). Effectiveness of computer-based college teaching: A meta-analysis of findings. *Review of Education Research*, 50(4), 525-544.
- Kulik, J. A., & Kulik, C.-L. C. (1987). Review of recent literature on computer-based instruction. *Contemporary Education Review*, 12, 222-230.
- Lesgold, A., Lajorie, S. P., Brunzo, M., & Eggen, G. (1992). Possibilities for assessment using computer-based apprenticeship environments. In J. S. Regian & V. J. Shute (Eds.), *Cognitive approaches to automated instruction* (pp. 49-80). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Levy, M. (2002). CALL by design: products, processes and methods. *ReCALL*, 14(1), 58-85.
- Levy, M. (2004). *Interpretations of Context in CALL Research: The Goals, the Data and the Methods*. Paper presented at the Eleventh International CALL Conference, University of Antwerp.

- Liao, Y. K. (1992). Effects of computer-assisted instruction on cognitive outcomes: A meta-analysis. *Journal of Research on Computing and Education*, 24, 367-380.
- Liu, M., Moore, Z., Graham, L., & Lee, S. (2002). A look at the research on computer-based technology use in second language learning: A review of the literature from 1990-2000. *Journal of Research on Technology in Education*, 34(3), 250-273.
- Lo, J. J., Wang, H. M., & Yeh, S. W. (2004). Effects of confidence scores and remedial instruction on prepositions learning in adaptive hypermedia. *Computers & Education*, 42(1), 45-63.
- Ma, L., Nicolas, V. V., Gierts, S., & Vandamme, F. (2002). Evaluation of a Virtual Environment: The Elementary Spoken Chinese Prototype (ESCvar Prototype). *Communication and Cognition*, 35(3-4), 257-275.
- Maddux, C. D. (1995). Research in Educational Computing: Problems of Internal and External Validity. *Computers in the Schools*, 11(3), 7-10.
- McIssac, M. S., & Gunawardana, C. N. (1996). Distance education. In D. Johnassen (Ed.), *Handbook of research for educational communications and technology* (pp. 403-437). New York: Macmillan.
- Miech, E. J., Nave, B., & Mosteller, F. (1996). *On CALL: A Review of Computer-Assisted Language Learning in U.S. Colleges and Universities*. Cambridge, MA: From the Center for Evaluation of the Program Initiatives for Children.
- Myers, M. J. (2000). Voice Recognition Software and a Hand-Held Translation Machine for Second-Language Learning. *Computer Assisted Language Learning*, 13(1), 29-41.
- Niemiec, R. P., & Walberg, H. J. (1987). Comparative effects of computer-assisted instruction: A synthesis of reviews. *Journal of Educational Computing Research*, 3, 19-37.
- Niemiec, R. P., & Walberg, H. J. (1992). The effect of computers on learning. *International Journal of Education*, 17, 99-108.
- Noble, D. F. (1998). *Digital Diploma Mills, part II, The coming battle over online instruction*, from <http://communication.uscd.edu/dl/ddm2.html>
- Nunan, D. (1992). *Research methods in language learning*. Cambridge ; New York, NY, USA: Cambridge University Press.
- Nutta, J. W., Feyten, C. M., Norwood, A. L., Meros, J. N., Yoshii, M., & Ducher, J. (2002). Exploring new frontiers: What do computers contribute to teaching foreign languages in elementary school? *Foreign Language Annals*, 35(3), 293-306.
- Phipps, R., & Merisotis, J. (1999). *What's the Difference? A Review of Contemporary Research on the Effectiveness of Distance Learning in Higher Education*. Retrieved 11/04/2004, from <http://www.ihep.com/Pubs/PDF/Difference.pdf>
- Robson, C. (2002). *Real world research : a resource for social scientists and practitioner-researchers* (2nd ed.). Oxford: Blackwell.
- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, 86, 638-641.
- Rumberger, R. W. (2000). *A multilevel, longitudinal approach to evaluating the effectiveness of educational technology*. Paper presented at the Design meeting on effectiveness in educational technology, Menlo Park, California.
- Russell, T. L. (1999). *The No Significant Difference Phenomenon*. Chapel Hill , NC: Office of Instructional Telecommunications, North Carolina University.
- Ryan, A. W. (1991). Meta-analysis of achievement effects of microcomputer applications in elementary schools. *Educational Administration Quarterly*, 27, 161-184.

- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Singleton Jr., R. A., Straits, B. C., & Straits, M. M. (1993). *Approaches to Social Research* (second ed.). New York: Oxford University Press.
- Smith, B., Alvarez-Torres, M. J., & Zhao, Y. (2003). Features of CMC technologies and their impact on language learners' online interaction. *Computers in Human Behavior*, 19(6), 703-729.
- Soe, K., Koki, S., & Chang, J. M. (2000). *Effect of Computer-Assisted Instruction (CAI) on Reading Achievement: A Meta-Analysis*. Retrieved 18/2/2004, from <http://www.prel.org/products/Products/effect-cai.htm>
- Suh, J.-S. (2002). Effectiveness of CALL Writing Instruction: The Voices of Korean EFL Learners. *Foreign Language Annals*, 35(6), 669-679.
- Van Dusen, L. M., & Worthren, B. R. (1994). The impact of integrated learning system implementation on student outcomes: Implications for research and evaluation. *International Journal of Educational Research*, 21.
- Vincent, J. (2001). The role of visually rich technology in facilitating children's writing. *Journal of Computer Assisted Learning*, 17(3), 242-250.
- Warschauer, M. (1996). Computer-mediated Collaborative Learning: Theory and Practice. *Modern Language Journal*, 18(4), 470-481.
- Weasenforth, D., Biesenbach-Lucas, S., & Meloni, C. (2002). Realizing Constructivist Objectives Through Collaborative Technologies: Threaded Discussions. *Language Learning & Technology*, 6(3), 58-86.
- Windschitl, M. (1998). The WWW and classroom research: what path should we take? *Educational Researcher*, 27(1), 28-33.
- Yeh, S.-W., & Lehman, J. D. (2001). Effects of Learner Control and Learning Strategies on English as a Foreign Language (EFL) Learning from Interactive Hypermedia Lessons. *Journal of Educational Multimedia and Hypermedia*, 10(2), 141-159.
- Zhao, Y. (2001). *Recent developments in Technology and language learning: A literature review and meta-analysis*. Retrieved 18/02/2004, from <http://ott.educ.msu.edu/elanguage/about/literature.asp>

Notes: My grateful thanks to Nina Dow for her contribution to the data gathering and statistical analyses.

An earlier version of this paper was delivered at the Eleventh International CALL Conference, at the University of Antwerp 5-7 September, 2004.

## Appendix A.

### ***General Overview of Articles and Categories***

<b>Proj. Id</b>	<b>Technology</b> <b>Ed Setting</b>	<b>*</b>	<b>Research Design</b>	<b>Language</b>	<b>Subject/lang skill</b>	<b>Variables</b>
<b>1</b> n = 33	Computerized grammar and vocabulary exercises (Dasher); instructional videos, computerized reading program, on-line spell-checker and English-French glossary Tertiary	Q	Nonequivalent Comparison Group Survey/questionnaires	French	Reading Writing Speaking Listening Culture	Proficiency Student Anxiety Student Motivation Student Perceptions Student/Teacher Roles
<b>3</b> n =85	audio- and video-cassettes Tertiary	P	One-group Pre-Post Tests	ESL/EFL	Writing	Writing Quality
<b>4</b> n =30	Computerized reading program using video and static pictures Tertiary	Q	Within-subjects Design Survey/questionnaires	ESL/EFL	Vocabulary	Vocabulary Development Picture or Video annotation
<b>7</b> n =48	Software for "holistic practice" providing "higher level, contextualized practice activities, idea processor and on-line thesaurus Tertiary	E	Factorial Experimental Design	Spanish	Writing	Writing Quality
<b>9</b> n =54	Grammar checker Tertiary	N	Cohort comparison	French	Writing	Student Attitude Grammatical accuracy
<b>14</b> n =20	Email Tertiary	N	Case Study	Japanese	Writing	Writing Quality
<b>17</b> n =20	Synchronous CMC Tertiary	E	Between-subjects Design	Spanish	Vocabulary Listening Writing Reading Speaking	Vocabulary Development

Proj. Id	Technology	Ed Setting	*	Research Design	Language	Subject/lang skill	Variables
18 n =17	Hyperlinks	Tertiary	Q	Between-subjects Design	French	Reading	Vocabulary Development Time on Task Reading Skills
19 n =60	Hyperlinks	Tertiary	E	Between-subjects Design Within-subjects Design	French	Reading	Vocabulary Development Time on Task Reading Skills Concentration Level
20 n =20	Multi-media - ClarisWorks and HyperStudio	Primary	P	One-group Pre-Post Tests	English	Writing	Writing Quality
23 n =12	Integrated Web Materials	Tertiary	N	Analysis of transcripts	Spanish	Speaking Communication	Communication/negotiation Interaction
24 n =181	CD with L1 translation, user control text, hypertext, colour graphics and sound	Tertiary	E	Pre-Posttest Control Group	French	Vocabulary	Vocabulary Development Gender differences L1 Translation Time for look-up
25 n =15	Virtual Learning Environment	Tertiary	N	Survey/questionnaires Non-participant observation	German	General or not specified	Student Perceptions
26 n =26	Computerised visual display	Tertiary	Q	Pre-Posttest Control Group Survey/questionnaires	French	Pronunciation	Prosody
27 n =27	Web Tutor	Tertiary	Q	Between-subjects Design Survey/questionnaires	German	Grammar	Click/Drag/Type
28 n =112	Animated Graphical Annotation	Tertiary	E	Pre-Posttest Control Group Survey/questionnaires	Japanese	Listening Pronunciation	Listening & Pronunciation
37 n =33	Netscape Composer	Tertiary	N	Case Study Non-participant observation Focus Groups Survey/questionnaires	Spanish	Create web page	Student Perceptions Student/Teacher Roles
40 n =58	Computer-networked environment	Tertiary	N	Questionnaires & Teachers Logs	ESL/EFL	Reading	Reading Skills Student Attitude Student Motivation
43 n =88	Adaptive Hypermedia	Tertiary	E	Between-subjects Design	ESL/EFL	Grammar	Adaptive remedial instruction Student Confidence Rating

Proj. Id	Technology	Ed Setting	*	Research Design	Language	Subject/lang skill	Variables
45 n =41	Asynchronous CMC	Tertiary	N	Case Study Non-participant observation Survey/questionnaires Analysis of transcripts	EFL	Speaking Listening Writing	Conversational Interaction Student Attitude
48 n =6	Videodisc	Tertiary	N	Case Study Video observation Analysis of transcripts	Italian	Speaking Listening	Conversational Interaction
55 n =62	Internet; audio and visual annotation of text	Tertiary	E	Pre-Posttest Control Group	French	Vocabulary Reading	Vocabulary Development Time on Task Authoring Multimedia
56 n =28	CD ROM	Primary	E	Pre-Posttest Control Group Case Study	Spanish	Integrated Language Skills	Writing Quality Proficiency
59 n =5	Email	Tertiary	N	Case Study Interviews Non-participant observation Analysis of transcripts	ESL/EFL	Communication Writing	Student Attitude Student Motivation
63 n =56	Interactive Video	Tertiary	Q	Posttest-only Control Group	Spanish	Listening Writing	Factual Accuracy Grammatical accuracy Elaboration/Paraphrasing
64 n =14	Synchronous CMC	Tertiary	Q	Within-subjects Design Survey/questionnaires	ESL/EFL	Communication	Student Perceptions Conversational Interaction Type of CMC
65 n =?	Web-based interactive tool	Tertiary	P	One-group Post Tests Only Survey/questionnaires	Russian	Reading Vocabulary Grammar	Grammatical accuracy Vocabulary Development Reading Skills Writing Quality Student Perceptions
68 n =19	Computer-networked environment	Tertiary	N	Case Study	ESL/EFL	Writing	Student Perceptions

Proj. Id	Technology	Ed Setting	*	Research Design	Language	Subject/lang skill	Variables
70 n =55	Web/Internet/email	Tertiary	N	Cross-sectional survey Non-participant observation Interviews	ESL/EFL Chinese German Indonesian Japanese Russian	Create web page	Learner Autonomy
71 n =27	Interactive Laserdiscs	Tertiary	N	Case Study Interviews/questionnaires/Journals	French Spanish	Integrated Language Skills	Learning Strategies/Styles
73 n =6	Multi-media – “MicroWorlds”	Primary	N	Case Study Interviews Analysis of transcripts Participant observation	English	Writing	Writing Quality
74 n =52	Asynchronous CMC	Tertiary	N	Situated study Survey/questionnaires	ESL/EFL	Writing Reading	Critical reflection
78 n =111	Hypermedia-based interactive video program	Tertiary	E	Factorial Experimental Design Survey/questionnaires	ESL/EFL	Listening Reading	Learner Control Learning Strategies/Styles Student Attitude Advance Organisers Recall
79 n =29	Computer-networked environment	VET	N	Case Study Interviews/questionnaires/Journals Non-participant observation	ESL/EFL	General or not specified	Student Perceptions Proficiency Grammatical accuracy
84 n =47	Multi-media - ClarisWorks and HyperStudio	Primary	E	Pre-Posttest Control Group	English	Writing	Writing Quality
85 n =66	Web Tutor	Tertiary	Q	Posttest-only Control Group Survey/questionnaires	French	Vocabulary	Vocabulary Development
86 n =72	Virtual Learning Environment	Tertiary	N	Cross-sectional survey	French German	General or not specified	Student Perceptions
91 n =72	Email	Tertiary	N	Analysis of transcripts	ESL/EFL	Writing	Writing Quality
92 n =63	Web/Internet/email	Tertiary	N	Survey/questionnaires Interviews	ESL/EFL Japanese Italian	General or not specified	Student Perceptions Learning Strategies/Styles

Proj. Id	Technology	Ed Setting	*	Research Design	Language	Subject/lang skill	Variables
93 n =82	Web/Internet/email	Secondary	N	Survey/questionnaires Interviews	ESL/EFL Japanese French German Spanish	General or not specified	Student Perceptions Learning Strategies/Styles
94 n =27	Synchronous CMC	Tertiary	Q	Within-subjects Design	German	Communication	Oral Discourse Written Discourse Participation Grammatical Competence
95 n =27	Synchronous CMC	Tertiary	N	Analysis of transcripts	Spanish	Communication	Interactional features
96 n=20	Synchronous NBC	Tertiary	N	Analysis of transcripts	Spanish	Communication	Grammatical Competence
97 n=24	Synchronous CMC	Tertiary	Q	Within-subjects Design	Spanish	Vocabulary	Vocabulary Development Student Perceptions
98 n=56	Chatline	Tertiary	N	Analysis of transcripts	Italian	Communication	Communication/negotiation Noticing Modified output
100 n=50	Synchronous CMC		N	Analysis of transcripts Survey/questionnaires	Spanish	Communication	Communication/negotiation Student Attitude
101 n=40	Multi-media and User-behaviour tracking technologies		Q	Nonequivalent Comparison Group	Spanish	Grammar	Grammatical Competence
102 n=5	Multi-media and internet		N	Case Study	French Spanish	General	Student Perceptions
103 n=103	Computerized grammar exercises		Q	Longitudinal Study Nonequivalent Comparison Group	German	Grammar	Grammatical accuracy
105 n=30	Speech Recognition Software		N	Analysis of transcripts	ESL/EFL	Pronunciation	Pronunciation Problems
106 n=60	Synchronous CMC		N	Analysis of transcripts	Spanish	Communication	Communication/negotiation Grammatical accuracy
107 n=11	Multi-media and grammar animations		E	Pre-Posttest Control Group Survey/questionnaires	German	Grammar	Long-term retention

\* E = Experimental; N = Non-experimental; P = Pre-experimental; Q = Quasi-experimental;

## **Appendix B**

### ***Research Design Terminology***

#### **PRE-EXPERIMENTAL DESIGN**

May have pre- and posttreatment tests, but lacks a control group. (Nunan, 1992:41)

#### **QUASI-EXPERIMENTAL DESIGN**

Has both pre and posttests and experimental and control groups, but no random assignment of subjects.(Nunan, 1992:41)

#### **EXPERIMENTAL DESIGN**

Has both pre and posttests, experimental and control groups, and random assignment of subjects.(Nunan, 1992:41)

#### **NONEXPERIMENTAL DESIGN**

Refers to situations in which a presumed cause and effect are identified and measured but in which other structural features of experiments, such as random assignment, pretests and control groups are missing. Instead reliance is placed on measuring alternative explanations individually and then statistically controlling for them.

#### **ONE-GROUP PRETEST-POSTTEST DESIGN**

A single pre-test observation is taken on a group of respondents, treatment then occurs, and a single posttest observation on the same measure follows (Shadish et al, 2002:108)

#### **NONEQUIVALENT COMPARISON GROUP DESIGN**

Uses a treatment group and an untreated comparison group, with both pretest and posttest data gathered on the same units. (Shadish et al, 2002:136)

#### **POSTTEST-ONLY CONTROL GROUP DESIGN**

Incorporates just the basic elements of experimental design: random assignment of subjects to treatment and control groups, introduction of the independent variable to the treatment group, and a post treatment measure of the dependent variable for both groups. (Singleton Jr. et al, 1993:222)

#### **PRETEST-POSTTEST CONTROL GROUP DESIGN**

A design which measures the experimental group before and after the experimental treatment. A control group is measured at the same time, but does not receive the experimental treatment.

#### **WITHIN-SUBJECTS**

A study designed to make a comparison of 2 or more treatments and that compares them by having each user try each treatment, measuring their performance for each. (Diamond Bullet)

### **BETWEEN-SUBJECTS**

A study designed to make a comparison of 2 or more treatments and that compares them by having one set of users try one treatment and another set of users try another treatment, measuring their performance for each. (Diamond Bullet)

### **FACTORIAL EXPERIMENTAL DESIGN**

A design which enables the effects of two or more independent variables to be explored jointly. (Singleton Jr. et al, 1993:225)

### **CASE STUDY**

A strategy for doing research which involves an empirical investigation of a particular contemporary phenomenon within its real life context using multiple sources of evidence. (Robson, 2002:178)

### **CROSS-SECTIONAL SURVEY**

Data on a sample or “cross section” of respondents chosen to represent a particular target population are gathered at essentially one point in time. (Singleton Jr. et al, 1993:254)

### **NONPARTICIPANT OBSERVATION**

An approach to field research in which the researcher attempts to observe people without interacting with them and, typically without their knowing that they are being observed. (Singleton Jr. et al, 1993:520)

### **TIME SERIES**

Refers to a large series of observations made on the same variable consecutively over time. The observations can be on the same unit or on different but similar units. (Shadish et al, 2002:172)

### ***Effect size***

The number of standard deviation units separating scores of experimental and control groups. Values above 0.25 are large enough to be educationally meaningful (Kulik 2003).

### ***Hawthorne effect***

Refers to participants’ awareness of being studied affecting their performance (Singleton Jr., Straits & Straits 1993:29).

### ***Pygmalion effect***

Refers to teachers’ expectations about student achievement becoming self-fulfilling prophecies (Shadish, Cook & Campbell 2002:78).

## Appendix C

### Research Studies

- 1 Adair-Hauck, B., Willingham-McLain, L., & Youngs, B. E. (2000). Evaluating the Integration of Technology and Second Language Learning. *CALICO Journal*, 17(2), 269-306.
- 3 Al-Ahmed, F. H. (2000). Multi-Media Input to a Foundation ESP Course at University Level. *British Journal of Educational Technology*, 31(4), 374-376.
- 4 Al-Seghayer, K. (2001). The Effect of Multimedia Annotation Modes on L2 Vocabulary Acquisition: A Comparative Study. *Language Learning & Technology*, 5(1), 202-232.
- 7\* Borrás, I., & Coronel, M. (2000). *Integrating ARCS Motivational Strategies into CALL Courseware for Holistic Practice: A Study*. Paper presented at the World Conference on Educational Multimedia, Hypermedia and Telecommunications (EDMEDIA).
- 9 Burston, J. (2001). Exploiting the Potential of a Computer-based Grammar Checker in Conjunction with Self-Monitoring Strategies with Advanced Level Students of French. *CALICO Journal*, 18(3), 499-515.
- 14 Chikamatsu, N. (2003). The effects of computer use on L2 Japanese writing. *Foreign Language Annals*, 36(1), 114-127.
- 17 de la Fuente, M. J. (2003). Is SLA Interactionist Theory Relevant to CALL? A Study on the Effects of Computer-Mediated Interaction in L2 Vocabulary Acquisition. *Computer Assisted Language Learning*, 16(1), 47-81.
- 18 De Ridder, I. (2000). Are We Conditioned to Follow Links? Highlights in CALL Materials and Their Impact on the Reading Process. *Computer Assisted Language Learning*, 13(2), 183-195
- 19 De Ridder, I. (2002). Visible or Invisible Links: Does the Highlighting of Hyperlinks Affect Incidental Vocabulary Learning, Text Comprehension, and the Reading Process? *Language Learning & Technology*, 6(1), 123-146.
- 20 Fan, H.-L., & Orey, M. (2001). Multimedia in the Classroom: Its Effect on Student Writing Ability. *Journal of Research on Technology in Education*, 33(5).
- 23 González-Lloret, M. (2003). Designing Task-Based CALL to Promote Interaction: En Busca de Esmeraldas. *Language Learning & Technology*, 7(1), 86-104.
- 24 Grace, C. A. (2000). Gender differences: Vocabulary retention and access to translations for beginning language learners in CALL. *Modern Language Journal*, 84(2), 214-224.
- 25 Hampel, R., & Hauck, M. (2004). Towards an Effective Use of Audio Conferencing in Distance Language Courses. *Language Learning & Technology*, 8(1), 66-82.
- 26 Hardison, D. M. (2004). Generalization of Computer-Assisted Prosody Training: Quantitative and Qualitative Findings. *Language Learning &*

- Technology*, 8(1), 34-52.
- 27 Heift, T. (2003). Drag or Type, but Don't Click: A Study on the Effectiveness of Different CALL Exercise Types. *Canadian Journal of Applied Linguistics/Revue canadienne de linguistique appliquee*, 6(1), 69-85.
- 28 Hew, S.-H., & Ohki, M. (2001). A Study on the Effectiveness and Usefulness of Animated Graphical Annotation in Japanese CALL. *ReCALL*, 13(2), 245-260.
- 37 Lam, Y., & Lawrence, G. (2002). Teacher-Student Role Redefinition During a Computer-Based Second Language Project: Are Computers Catalysts for Empowering Change? *Computer Assisted Language Learning*, 15(3), 295-315.
- 40 Levine, A., Ferenz, O., & Reves, T. (2000). EFL Academic Reading and Modern Technology: How Can We Turn Our Students into Independent Critical Readers? *TESL-EJ*, 4 (4). Retrieved 14 April, 2004, from <http://www-writing.berkeley.edu/TESL-EJ/ej16/a1.html>
- 43 Lo, J. J., Wang, H. M., & Yeh, S. W. (2004). Effects of confidence scores and remedial instruction on prepositions learning in adaptive hypermedia. *Computers & Education*, 42(1), 45-63.
- 45 McIntosh, S., Brault, B., & Chao, T. (2003). A Case Study in Asynchronous Voice Conferencing for Language Instruction. *Educational Media International*, 40(1-2), 63-74.
- 48 Miceli, T., & Kennedy, C. (2000). Call in a communicative context : a study of students' conversation at a multimedia workstation. *Babel*, 35(1), 18-23.
- 55 Nikolova, O. R. (2002). Effects of Students' Participation in Authoring of Multimedia Materials on Student Acquisition of Vocabulary. *Language Learning & Technology*, 6(1), 100-122.
- 56 Nutta, J. W., Feyten, C. M., Norwood, A. L., Meros, J. N., Yoshii, M., & Ducher, J. (2002). Exploring new frontiers: What do computers contribute to teaching foreign languages in elementary school? *Foreign Language Annals*, 35(3), 293-306.
- 59 Sakar, A. (2001). The cross-cultural effects of electronic mail exchange on the Turkish University students of English as a Foreign Language (EFL). *CALL-EJ Online*, 3 (1). Retrieved 14 April, 2004, from <http://www.clec.ritsumei.ac.jp/english/callejonline/6-1/sakar.html>
- 63 Shea, P. (2000). Leveling the Playing Field: A Study of Captioned Interactive Video for Second Language Learning. *Journal of Educational Computing Research*, 22(3), 243-263.
- 64 Smith, B., Alvarez-Torres, M. J., & Zhao, Y. (2003). Features of CMC technologies and their impact on language learners' online interaction. *Computers in Human Behavior*, 19(6), 703-729.
- 65 Soboleva, O., & Tronenko, N. (2002). A Russian Multimedia Learning Package for Classroom Use and Self-Study. *Computer Assisted Language Learning*, 15(5), 483-499.
- 68 Suh, J.-S. (2002). Effectiveness of CALL Writing Instruction: The Voices of Korean EFL Learners. *Foreign Language Annals*, 35(6), 669-679.
- 70 Toyoda, E. (2001). Exercise of Learner Autonomy in Project-Oriented CALL. *CALL-EJ Online*, 2 (2). Retrieved 14 April, 2004, from

<http://www.clec.ritsumei.ac.jp/english/callejonline/5-2/toyoda.html>

- 71 Ulitsky, H. (2000). Language Learner Strategies with Technology. *Journal of Educational Computing Research*, 22(3), 285-322.
- 73 Vincent, J. (2001). The role of visually rich technology in facilitating children's writing. *Journal of Computer Assisted Learning*, 17(3), 242-250.
- 74 Weasenforth, D., Biesenbach-Lucas, S., & Meloni, C. (2002). Realizing Constructivist Objectives Through Collaborative Technologies: Threaded Discussions. *Language Learning & Technology*, 6(3), 58-86.
- 78 Yeh, S.-W., & Lehman, J. D. (2001). Effects of Learner Control and Learning Strategies on English as a Foreign Language (EFL) Learning from Interactive Hypermedia Lessons. *Journal of Educational Multimedia and Hypermedia*, 10(2), 141-159.
- 79 Young, S. S.-C. (2003). Integrating ICT into second language education in a vocational high school. *Journal of Computer Assisted Learning*, 19(4), 447-461.
- 84 Fan, H.-L., & Orey, M. (2001). Multimedia in the Classroom: Its Effect on Student Writing Ability. *Journal of Research on Technology in Education*, 33(5).
- 85 Labrie, G. (2000). A French Vocabulary Tutor for the Web. *CALICO Journal*, 17(3), 475-499.
- 86 Dolle, J., & Enjelvin, G. (2004). Investigating "VLE-effectiveness" in Languages. *Computer Assisted Language Learning*.
- 91 Biesenbach-Lucas, S., & Weasenforth, D. (2001). E-Mail and Word Processing in the ESL Classroom: How the Medium Affects the Message. *Language Learning & Technology*, 5(1), 135-165.
- 92 Felix, U. (2001). A multivariate analysis of students' experience of Web-based learning. *Australian Journal of Educational Technology*, 17(1), 21-36.
- 93 Felix, U. (2004). A Multivariate Analysis of Secondary Students' Experience of Web-Based Language Learning. *ReCALL*, 16(1), 129-141
- 94 Bohlke, O. (2003). A Comparison of Students Participation Levels by Group Size and Language Stages during Chatroom and Face-to-Face Discussions in German. *CALICO Journal*, 21(1), 67-87.
- 95 Darhower, M. (2002). Interactional Features of Synchronous Computer-Mediated Communication in the Intermediate L2 Class: a Sociocultural Case Study. *CALICO Journal*, 19(2), 249-277.
- 96 Pellettieri, J. (2000). Negotiation in cyberspace: The role of chatting in the development of grammatical competence. In M. Warschauer & R. Kern (Eds.), *Networked-based language teaching: Concepts and Practice* (pp. 59-86). Cambridge: Cambridge University Press.
- 97 Perez, L. C. (2003). Foreign Language Productivity in Synchronous versus Asynchronous Computer-Mediated Communication. *CALICO Journal*, 21(1), 89-104.
- 98 Tudini, V. (2004). Chatlines for Beginners: Negotiating Conversation at a Distance. In B. Holmberg, M. Shelley & C. White (Eds.), *Distance Education and Languages: Evolution and Change* (Vol. in press): Multilingual Matters.

- 100 Blake, R. (2000). Computer Mediated Communication: A Window on L2 Spanish Interlanguage. *Language Learning & Technology*, 4 (1). Retrieved, from <http://llt.msu.edu/vol4num1/blake/default.html>
- 101 Collentine, J. (2000). Insights into the Construction of Grammatical Knowledge Provided by User-Behavior Tracking Technologies. *Language Learning & Technology*, 3 (2). Retrieved, from <http://llt.msu.edu/vol3num2/collentine/>
- 102 Bradley, T., & Lomicka, L. (2000). A Case Study of Learner Interaction in Technology-Enhanced Language Learning Environments. *Journal of Educational Computing Research*, 22(3), 347 – 368
- 103\* Hall, C. (2000). *The Effectiveness of CALL in Grammar Teaching. An Evaluation Using Error Analysis*. Paper presented at the Conference of University Teachers of German, University of Keele
- 105 Derwing, T. M., Munro, M. J., & Carbonaro, M. (2000). Does Popular Speech Recognition Software Work with ESL Speech? *TESOL Quarterly*, 34(3), 592
- 106\* Oskoz, A. (2004). *The Interaction Hypothesis and the Sociocultural Theory in Synchronous Computer-mediated Communication (S-CMC): Intertwining Non competing Approaches*. Paper presented at the EuroCALL Conference 2004, Vienna
- 107 Roche, J., & Scheller, J. (2004). On the Efficiency of Grammar Animations in CALL. *TEL & CAL*, 4/2004, 50-55.

\* Conference presentation only; text available from the author.

## Appendix D

### *List of Journals*

Journal	Experimental	Quasi- Experimental	Pre-Experimental	Non-Experimental	Total
Australian Journal of Educational Technology				SQ (1)	1
Babel				CS (1)	1
British Journal of Educational Technology			1		1
CALICO Journal		NCG (1) PCG (1) WS (2)		AT (1) 1	6
CALL-EJ Online				CS (1) CSS (1)	2
Canadian Journal of Applied Linguistics		BS (1)			1
Computer Assisted Language Learning	BS (1)	BS (1)	1	CS (1) CSS (1)	5
Computers & Education	BS (1)				1
Computers in Human Behavior		WS (1)			1
EDMEDIA Conference Proceedings	Factorial Experimental (1)				1
Educational Media International				CS (1)	1
Foreign Language Annals	PPC (1)			CS (2)	3
Journal of Computer Assisted Learning				CS (2)	2
Journal of Educational Computing Research		PCG (1)		CS (2)	3
Journal of Educational Multimedia and Hypermedia	Factorial Experimental (1)				1
Journal of Research on Technology in Education	PPC (1)		1		2
Language Learning & Technology	BS & WS (1) PPC (1)	NCG (1) PCG (1) WS (1)		AT (3) SQ (1) SS (1)	10
Modern Language Journal	PPC (1)				1
ReCALL	PPC (1)			SQ (1)	2
TESL-EJ				QTL (1)	1
TESOL Quarterly				AT (1)	1
TEL & CAL	PPC (1)				1

AT = Analysis of transcripts

BS = Between-subjects

CS = Case Study

CSS = Cross-sectional survey

NCG = Nonequivalent Comparison Group Design

PCG = Posttest-only Control Group

PPC = Pretest-Posttest Control Group

QTL = Questionnaires & Teachers Logs

SQ = Survey/questionnaires

SS = Situated study      WS = Within-subjects.

Table 1: Studies by Research Design\*

Research Design	Pre-Experimental	Quasi-experimental	Experimental	Non-experimental
One-group Pre-Post Tests	3 20 65			
Nonequivalent Comparison Group Design		1 101 103		
Posttest-only Control Group		63 85		
Pretest-Posttest Control Group		26	107	
Within-subjects Design		4 19** 64 94 97		
Between-subjects Design		18 27	17 19** 43	
Factorial Experimental Design			7 78	
Pre-Posttest Control Group			24 28 55 56 84	
Analysis of transcripts				23 91 95 96 98 100 105 106
Case Study				14 37 45 48 59 68 71 73 79 102
Cross-sectional survey				70 86
Questionnaires & Teachers Logs				40
Situated study				9 74
Survey/questionnaires				25 92 93
<b>Total</b>	<b>3</b>	<b>13</b>	<b>11</b>	<b>26</b>

\*see definitions in Appendix B

Note: Numbers represent the studies' ID.

\*\*This study used two research designs

Table 2: Studies by Language and Subject/skills Taught

	Chinese	ESL/EFL	French	German	Indonesian	Italian	Japanese	English	Russian	Spanish	Total
<b>Communication</b>		59 64	88	94		98		90		23 95 96 100 106	<b>11</b>
<b>Create web page</b>	70	70		70	70		70		70	37	<b>2</b>
<b>Culture</b>			1								<b>1</b>
<b>General or not specified</b>		79 92 93	86 93 102	25 86 93		92	92 93			93 102	<b>6</b>
<b>Grammar</b>		43		27 103 107					65	101	<b>6</b>
<b>Integrated Language Skills</b>		79 92	71				2			56 71	<b>5</b>
<b>Listening</b>		45 78	1			48	28			17 63	<b>7</b>
<b>Pronunciation</b>		105	26				28				<b>3</b>
<b>Reading</b>		40 74 78	1 18 19 55						65	17	<b>9</b>
<b>Speaking</b>		45	1			48				17 23	<b>5</b>
<b>Vocabulary</b>		4	24 55 85					77	65	17 97	<b>8</b>
<b>Writing</b>		3 45 54 59 68 74 91	1 9				14	20 73 84		7 17 63	<b>16</b>

Table 3: Studies by Research Design and Variable under investigation

Variable under investigation	Research Design															Total
	Pre-exp.	Quasi-experimental					Experimental				Non-experimental					
		NCG	PCG	PPC	WS	BS	BS	FED	PPC	CS	SQ	AT	CSS	QTL	SS	
Conversational Interaction					64					45 48		98 100 106				
Grammatical accuracy	65	9 103	63							79	9	106				
Grammatical competence		101			94							96				
Learning Strategies/Styles								78		71	92 93					
Listening & Pronunciation			63	26				78	28			105				
Proficiency		1							56	79						
Reading Skills	65				19	18	19	78						40		
Student Attitude								78		45 59	100			40		
Student Motivation		1								59				40		
Student Perceptions	65	1			64 97					37 68 79 102	25 92 93		86			
Student/Teacher Roles		1								37						
Time on Task					19	18	19		55							
Vocabulary Development	65		85		4 19 97	18	17 19		24 55							
Writing Quality	3 20		63					7	56 84	14 73		91				
Other		1			64 19 94	27	19 43	78	24 55 107			23 95 98	70		74	

**NCG** = Nonequivalent Comparison Group Design

**PCG** = Posttest-only Control Group

**AT** = Analysis of transcripts

**CS** = Case Study

**PPC** = Pretest-Posttest Control Group

**FED** = Factorial Experimental Design

**CSS** = Cross-sectional survey

**SQ** = Survey/questionnaires

**QTL** = Questionnaires & Teachers Logs

**BS** = Between-subjects Design

**WS** = Within-subjects Design

**SS** = Situated study

Table 4: Studies by Language Taught

Language	Article Ids	Total for language
Chinese	70	1
ESL/EFL	3 4 40 43 45 59 64 68 70 74 78 79 91 92 93 105	16
L1 English	20 73 84	3
French	1 9 18 19 24 26 55 71 85 86 87 88 93 102	14
German	25 27 70 86 93 94 103 107	8
Indonesian	30 70	2
Italian	48 92 98	3
Japanese	14 28 70 92 93	5
Russian	65 70	2
Spanish	7 17 23 37 56 63 71 93 95 96 97 100 101 102 106	15

**Ref: Felix, U. (2005) Analyzing recent CALL effectiveness research: Towards a common agenda, *CALL Journal*, 18 (1&2), 1-33.**