Lecture Notes on Language and Literature (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/langl.2024.070310
ISSN 2523-5869 Vol. 7 Num. 3

Experimental Research Design in Evaluating Computer Assisted Second Language Learning

Feng Gao[1,a,*], Tom Duan[2,b]

[1]School of English Language, Literature and Culture, Beijing International Studies University, Beijing, China
[2]Department of Recruitment, Admissions and International Development, University of Roehampton, London, UK
[a]gaofeng@bisu.edu.cn, [b]Tom.duan@roehampton.ac.uk
[*]Corresponding author

Keywords: Experimental Research Design, Computer Assisted Second Language Learning, Controlled Trials, Randomised Controlled Trials

Abstract: By reviewing a range of related studies, this article has illustrated the internal validity of controlled trial (CT) design and randomised controlled trial (RCT) design in evaluating computer assisted second language learning. It has also provided a range of explanations for why RCTs are so rare in this topic area. The RCT is the most robust method of assessing effectiveness in terms of its prestigious internal validity. Nevertheless, this is not to deny the merit of quasi-experimentation, especially in evaluating computer assisted second language learning. A quasi-experiment, relaxing some aspects of control, may still yield valuable information and enable a researcher to answer some specific question arising from second language teaching experience.

1. Introduction

Educational researchers use different research methods to confirm or disconfirm hypotheses, because some research questions can be addressed with certain methods but not with others. Compared with other research methods in education, experimental research is the strongest for testing causal relationship [6, 9]. For its enthusiastic advocates, experimental research is the only method of verifying educational improvement, and the only way of introducing improvement without the danger of a faddish discard of old wisdom in favour of inferior novelties [4]. In commonsense language, experimental research is a research technique, which regularly begins with a hypothesis, modifies something in a situation, and then compares outcomes with and without the modification [9]. Experiments encourage researchers to isolate and target the impact arising from one or a few variables, so that experimental studies can claim to show any degree of causality [9]. The high level of control over experiments, however, will decrease external validity of experimental research, because of unnatural research situations and contexts. The other distinct characteristic of experimental research is explicitness of data collection procedures [4]. Precise quantitative data, the norm in experiments, is always generated through carefully designed and focused instruments (e.g. tests, observations and questionnaires).

In the language education realm, an experiment typically focuses on a specific element of the larger process of language learning and teaching. Computer assisted second language learning, like any pedagogical activity, potentially can be evaluated through experimental research. The experimental research designs in this topic area can be categorized into post-test studies, pre- and post-test studies, controlled trials (CTs), and randomised controlled trials (RCTs). The randomised controlled trial (RCT) is a classical experimental design, and the others are three types of quasi-experimental designs. Post-test design is regarded to have no scientific value, in respect of a total absence of control [4]. Pre- and post-test design is also a relatively poor example of experimental research design. Since there is no comparison group existing in pre- and post-test design, any difference between pre-test and post-test can be attributed to history, maturation, the effect of testing, instrumentation decay, statistical regression, or knowledge of being in an experiment [4, 12]. Considering the obvious weakness of post-test design and pre- and post-test design, this essay will only focus on the strength and limitation of controlled trial (CT) design and RCT design, and the feasibility of carrying out RCTs in evaluating computer assisted second language learning.

2. Controlled Trials in Evaluating Computer Assisted Second Language Learning

In CT designs for evaluating computer assisted second language learning, a researcher usually firstly conducts a pre-test to measure the subjects' second language ability, uses two or more groups and makes one or more groups exposed to the intervention, and then evaluates the effects of the intervention by comparing the result of post-test between the group that receives the intervention (the experimental group) and a group allocated to conventional practice, a placebo, or no intervention (the control group). The gathered data will be analysed using statistical tests of significance to determine the subjects' second language achievement.

Coll [5], for instance, undertook a research study to see whether low-proficiency English for Specific Purpose (ESP) learners could benefit from a hypermedia enhanced learning environment, specifically in terms of incidental acquisition of lexical items in the target language. The subjects were 80 lower intermediate ESL level learners (aged 17-21 yr) enrolled in Chemistry and Chemical Engineering English courses at the Universitat Jaume I, CastellÓ, Spain. The 80 students were equally divided into an experimental group and a control group. The subjects of the experimental group received a set of multimedia lessons using courseware — *The World of Chemistry: Selected Demonstrations and Animations* — for two weeks. Vocabulary was taught implicitly by providing subjects visual and auditory input to Chemistry-related English. Pre- and post-testing involved the same multiple-choice, blank-filling, and sentence-making vocabulary exercises. A statistical comparison of these two tests between experimental group and comparison group revealed an improvement in vocabulary for both experimental group and control group. The vocabulary achievement scores of the subjects in experimental group increased significantly from pre-treatment to post-treatment as opposed to those in the control group. The obvious implication of this study is that hyper-media based instruction, if properly designed, can provide an effective learning environment to promote vocabulary acquisition within a second language learning framework.

A cause-and-effect relationship is more than just hinted at, with hyper-media based instruction being presumed cause and an acceleration in the rate of vocabulary acquisition being the effect. Having a comparison group in this CT, the researcher can eliminate some alternative explanations, such as the effect of history and maturation, from the causal inference. Nevertheless, there may be one or more rival hypotheses that prevent a clear interpretation of the study results. The first rival variable is selection bias. The experimental group and the control group in this study were formed by matching their low ability in English. There are still other variables that would affect the dependent variables. For instance, the experimental group might unintentionally contain subjects who have a relatively positive attitude and a good experience of the technical aspects of computer use, and comparatively higher motivation in English learning than those in the control group. Since a researcher cannot foresee all the variables between the two groups that will influence the result, true matching becomes an impossible task. A second uncontrolled rival hypothesis is statistical regression. Because they begin at an unusually low level in English vocabulary (the overall mean scores are 32.6% and 34.6% for the experimental group and the control group in the pre-test respectively), the subjects are unlikely to respond further in the same direction. It is uncertain as to what percentage of the improvement between pre-test and post-test is due to statistical regression. The improvement of the subjects in the control group may attribute to the third rival variable, diffusion of treatment. Experimental group subjects may tell those in the control group about the new courseware they used to learn vocabulary, and then the control group subjects may use it. Alternatively, the subjects in the control group may seek for extra tuition or work harder to reduce differences. This is the fourth rival hypothesis — compensatory behaviour. A fifth confounded variable is Hawthorne effect. The experimental group may improve, not through any intrinsic effect of hyper-media based instruction but due to merely taking part in the experiment. Considering the interval between pre-test and post-test is only two weeks, temporal change is the sixth confounded rival explanation. The increased scores of control group subjects in the post-test prove that people will generally improve in second language learning over time irrespective of any intervention. A seventh rival hypothesis is testing effect. The same types of vocabulary exercises were employed in the pre-test and post-test. If the subjects had remembered the pre-test questions and this affected what they learned in the following two-week hypermedia assisted vocabulary learning, or how they answered questions on the post-test, the researcher could not claim that the treatment alone had affected the dependent variable.

This is only one of many possible examples of CT design, and there might be other uncontrolled threats arising with instrumentation, morality and experimenter expectancy. Only selection bias and regression to the mean can be decreased or removed from the internal validity of CTs by using random allocation, whilst other rival hypotheses can occur after randomisation [13].

3. Randomised Controlled Trials in Evaluating Computer Assisted Second Language Learning

Having all the parts of CT design, RCT design requires that the two or more groups involved in experimental study are formed through random allocation. The RCT is acknowledged to be the ‘gold standard’ of effectiveness research, because of the high quality of causal inferences that can be made from it [6, 13]. Randomised allocation makes the samples representative of a known population, and comparable to each other within limits of sampling error [6]. The RCT design, however, is not a panacea that inevitably rules out all threats to internal validity, and it does not guarantee that the initial comparability between groups will be maintained over the course of an experiment [6].

The aim of Lin *et al*. [7] study is to try to investigate what are the differential effects of computer-assisted instruction (CAI) and a paper-and-pencil approach on automatization of word cognition skills among mildly mentally handicapped and nonhandicapped students. Ninety three sample students were selected from 10 elementary public schools in New York City according to the following criteria: adequate English proficiency, absence of significant sensorimotoric impairments or severe behaviour problems, and achieving accuracy between 45% and 85% in more than 100 seconds in the pre-test. The participants were Caucasian-American and Chinese-American children in various grades. The subjects were randomly assigned to the CAI and paper-and-pencil conditions. The intervention consisted of 10 lessons. For the CAI condition, students were instructed to read the entire computer screen display aloud in the presentation phase. Whilst for paper-and-pencil condition, flashcards were used in place of screen. A 20-item multiple-choice test served as both the pre- and post-test. The subjects were quizzed on the set of words, which had been used in the presentation phase in post-test. The results were quite complicated: both groups of students, particularly paper-and-pencil group students and nonhandicapped students, improved significantly in accuracy; both groups of students, especially CAI group students, nonhandicapped students, and initially slower students, displayed less response time in the post-test. In conclusion, the CAI assisted mildly mentally handicapped and nonhandicapped students to speed up in word recognition, whilst the paper-and-pencil approach is valued for facilitating learners to achieve higher accuracy.

This study represents rare RCTs in evaluating computer assisted second language learning. Since the 93 students were formed into the CAI group and the paper-and-pencil group by randomised allocation, comparability is achieved by equating the average unit within each group. Randomisation does not remove the idiosyncrasy, such as culture backgrounds, socio-economic strata of the subjects, from any one unit, but makes the entire known and unknown idiosyncrasy that could affect outcome equally presented in the CAI group and the paper-and-pencil group. By random chance alone, the extremely badly scored (near to 45% accuracy) students do not concentrate on one group, so that the subjects in each group appear less extreme, and are influenced less by statistical regression. Furthermore, since statistical regression affects both the CAI group and paper-and-pencil group equally, the effect of statistical regression is cancelled out in comparing the improvement of the post-test results between the two groups.

As an individually randomised trial, it is quite hard to decide what kind of role that diffusion of treatment has played in this research study in respect that the subjects of CAI group and paper-and-pencil group will communicate with each other easily. Cluster randomised design will avoid or reduce the risk of contamination between experimental group and control group [14]. Researchers still need to interpret the outcome of RCTs cautiously, because the other threats to internal validity, which are mentioned in the previous section, cannot be eliminated by randomised allocation.

4. Obstacles to Conduct Randomised Controlled Trials in Evaluating Computer Assisted Second Language Learning

Methodologically, the RCT is the most appropriate method that can be used in evaluating computer assisted second language learning; however, a very limited number of RCTs have been carried out in this topic area. Andrews *et al*. [1] reviewed published research on the impact of information communication technology (ICT) on literacy learning for 5-16-year-olds, and identified 169 researcher-manipulated evaluation research studies since 1990. Only one RCT on the impact of ICT on ESL learning met the inclusion criteria of this systematic review project.

There is no conclusive answer for why the RCT is rarely used in the research on this topic to date. The first possible explanation could be that evaluating computer assisted second language learning is classroom-based research. In the real world in which schools and classes exist, serious limitations are placed on the freedom of researchers to allocate subjects individually in a randomised fashion, and manipulate an experimental group and a control group to conduct research. Language program administrators are generally unwilling to disturb their ongoing programs and allow reorganization of classes in order to assign participants to different groups at random [11]. Secondly, to allocate subjects randomly without respect to subjects' preference for one of the interventions will lead to resentful demoralization [2]. Considering children and young adults' general enthusiasm for ICT, it will be difficult to form an equivalent control group without ICT related intervention randomly. The control group will turn into a 'negative treatment' group, and affect the scientific value of trials. Thirdly, RCTs cannot be designed to answer questions about certain kinds of possible causal variables [6]. For instance, it is not possible to assign subjects at random to detect how learners' different attitudes towards ICT influence the effect of CAI on second language learning. Fourth, since RCTs in educational evaluation frequently failed to produce results showing statistically significant advantages for treatments, a researcher might feel more reluctant and pessimistic to conduct a RCT in evaluating computer assisted second language learning, or unwilling to publish 'wrong' answers yielded by well-designed experimental studies. The fifth reason lies in lay public's uncertainty about RCTs, which brings researchers another problem to persuade participants to accept random assignment. In Rajan and Turner's (cited in [10]) study on day care and breast cancer, only 54% parents expressed they knew what the term 'randomisation' meant.

In addition, second language learning is not an instant process, and the results of short-term experiments lack validity and reliability, even when they show positive effects of interventions. To carry out a long-term RCT one has to face the serious problem caused by feasibility with regard to disruption of school routines. Second language learning is also a complex process involving many conditions. Researchers usually found that there are too many confounding variables to control effectively, and it is too difficult to conduct pure experimental studies with human participants in second language learning [3]. Thus, most studies in this topic area tend to be quasi-experimental rather than classical experimental.

RCTs allow researchers to have 'a controlled look at nature' (Paivio and Begg, cited in [8]: 155). Yet this strength limits researchers to examine numerous variables simultaneously [9]. Theoretically, to conduct an experiment by isolating the impact of ICT on second language learning from context will increase generalization and external validity of the experimental outcome. In fact, ICT cannot lead to second language learning directly and independently. Second language learning progress is influenced by a combination of many factors, such as learners' aptitude, motivation, and anxiety. It is still debatable that how far the results of such carefully decontextualized experiments can be generalized to learning in normal classroom [8].

The procedure of conducting a RCT is quite complicated and timebound: to develop and validate major research questions; to carry out pilot work; to make pre-test measurements; to collect, clean and order data; and to analysis data [6]. Hence, it is rarely desirable to conduct a RCT when decisions have to be made rapidly.

5. Conclusion

This essay has illustrated the internal validity of CT design and RCT design in evaluating computer assisted second language learning. It has also provided a range of explanations for why RCTs are so rare in this topic area. The RCT is the most robust method of assessing effectiveness in terms of its prestigious internal validity. Nevertheless, this is not to deny the merit of quasi-experimentation, especially in evaluating computer assisted second language learning. A quasi-experiment, relaxing some aspects of control, may still yield valuable information and enable a researcher to answer some specific question arising from second language teaching experience [8]. Quasi-experimentation is an alternative research technique in case the planned RCT cannot be implemented or breaks down after implementation.

The RCT is the most appropriate method, but not the only method in evaluating computer assisted second language learning. RCTs indicate what will transpire with respect to teaching a second language assisted by computer, whilst research studies with greater qualitative focus state what actually happened in second language classroom. Both research approaches provide pedagogical implications for designing and applying effective CAI in second language learning.

In a recent paper on the study of the impact of ICT on literacy education, Andrews *et al.* [1] suggested that the relationship between CAI and second language learning was not necessarily causal, but might be understood better as ‘symbiotic’. This perspective offers a new direction for research designs: how to measure the symbiosis existing in computer assisted second language learning.

References

[1] Andrews, R., Robinson, A. and Torgerson, C. (2004) Introduction. In R. Andrews (eds) The Impact of ICT on Literacy Education (pp. 1-33). London: Routledge Falmer.

[2] Brewin, C. R. and Bradley, C. (1989) Patient preferences and randomised clinical trials, British Medical Journal, 299, 313-315.

[3] Brown, J. D. and Rodgers, T. S. (2002) Doing Second Language Research. Oxford: Oxford University Press.

[4] Campbell, D. T. and Stanley, J. C. (1969) Experimental and Quasi-Experimental Designs for Research. Chicago: Rand M çNally & Company.

[5] Coll, J. F. (2002) Richness of semantic encoding in a hyper-media assisted instructional environment for ESP: Effects of incidental vocabulary retention among learners with low ability in the target language, ReCALL, 14 (2), 263-284.

[6] Cook, T. D. and Campbell, D. T. (1979) Quasi-Experimentation: Design & Analysis Issues for Field Settings. Boston: Houghton Mifflin Company.

[7] Lin, A., Podell, D. M. and Rein, N. (1991) The effects of CAI on word recognition in mildly mentally handicapped and nonhandicapped learners, Journal of Special Education Technology, 11 (1), 16-25.

[8] McDonough, J. and McDonough, S. (1997) Research Methods for English Language Teachers. London: Arnold.

[9] Neuman, W. L. (2003) Social Research Methods: Qualitative and Quantitative Approaches. 5th edn. Boston: Allyn and Bacon.

[10] Oakley, A. (2000) Experiments in Knowing: Gender and Method in the Social Sciences. Cambridge: Polity Press.

[11] Seliger, H. W. and Shohamy, E. (1989) Second Language Research Methods. Oxford: Oxford University Press.

[12] Spector, P. E. (1982) Research Designs. London: SAGE Publications Ltd.

[13] Torgerson, C. (2003) Systematic Reviews. London: Continuum.

[14] Torgerson, D. (2001) Contamination in trials: Is cluster randomisation the answer? British Medical Journal, 322, 355-357.