# Threats to Validity in a Study of the Effects of Hypermedia Instruction on Learning Outcomes - a Switched Replications Experiment

**Malgorzata S. Zywno**
**Ryerson University**

Abstract

Educational researchers in university settings face many difficulties in trying to conduct controlled action research studies on the effects of hypermedia on learning outcomes, where threats to validity and reliability are often beyond the influence of the investigator. Switched Replications experiment and another, semester-long study, where all students received hypermedia instruction in which it was embedded, were specifically conducted to address several validity threats. In each of the two weeks of the experiment, one group of students received a hypermedia lecture, while the other group received a conventional lecture on the same topic. The two lectures were immediately followed by a test. Two instructors were involved. The hypermedia-instructed group performed significantly better both times, regardless of which instructor made presentations. However, there were no statistically significant differences in overall academic achievement of the students in the course, offered in a hypermedia mode for 12 out of its 13 weeks. This indicated that the choice of instructional media had a strong effect on student achievement, while the effect of instructor differences was negligible. Selection bias, novelty factor, differences in instructional design and social threats to the internal validity of the study were also rejected as a possible explanation for the observed differences in achievement.

## I. Introduction

### Background

The study was situated in the sixth semester Control Systems course (ELE639) in an undergraduate program in Electrical and Computer Engineering at Ryerson University in Toronto, Canada. Hypermedia (text, graphics, video and sounds, linked in a non-linear, associative manner) have been introduced into the course to support experiential learning[1, 2]. A pilot project[3] with classroom hypermedia presentations took place in 1999. The academic performance of students was significantly higher, compared with those registered in the conventional version of the course. As well, positive attitudes towards technology-aided instruction were observed. A formal comparison study conducted in 2000 confirmed these findings, and pointed towards hypermedia instruction accommodating a wider range of learning styles than conventional instruction, as a possible explanation of observed differences[4]. In 2001 the hypermedia instruction was extended to all students registered in the course, collecting more data for the analysis of the relationship between learning styles, hypermedia and academic achievement. The results are described at length in the companion paper[5].

Hypermedia and Student Achievement

Existing studies on the efficacy of instructional hypermedia are still inconclusive[6, 7, 8, 9]. A recent meta-analysis[8] of 46 studies of the effects of hypermedia on student achievement found 60% of them reporting positive results of hypermedia instruction, while 40% reported no significant differences or negative results. Few of the studies reported in the literature meet even rudimentary scientific requirements for selection, manipulation and control of potential mediating variables[6, 7]. As well, educational researchers face many difficulties in trying to conduct controlled studies in university settings, where threats to validity and reliability are often beyond the influence of the investigator[10, 11]. As a result, often when positive effects of hypermedia on student achievement were identified, a compelling alternative hypothesis could not be rejected[6, 12].

## II. Methods

Participation in the study was voluntary. The students were not exposed to any risks, or reprisals for refusal to participate. The two-week Switched Replications experiment had 80% participation rate (n=102) and was embedded in the semester-long study[5], which took place between January and April 2001. All students registered in the course received hypermedia instruction throughout the semester, except of one week of the two-week experiment. Student achievement in the course, a primary dependent variable, was evaluated through standard academic assessments.

Due to the course logistics, students enrolled in ELE639 have always been assigned to two lecture groups taught by different instructors. The course has been tightly coordinated because of the requirements of the design project. The instructors have been following the same instructional design and the same timetable regardless of the instructional media, be it a conventional “chalk & talk” lecture, or hypermedia presentations. In 2001, both instructors made use of the same hypermedia materials. Lecture presentations used HTML pages with graphics, animations, JavaScript interactivity, embedded video clips, Java applets to help visualize concepts and to illustrate the behavior of real-life systems, and software simulations. Students obtained course notes prior to the lectures. The instructors posted course materials, moderated a bulletin board, and maintained an online student database with grades. All lecture materials, including interactive multimedia as well as additional supporting materials (lab and assignment tutorials, past tests and exams, etc.), were accessible for asynchronous review. The course was supported online through WebCT, a popular web management software package.

Random selection was not possible for the Switched Replications experiment, and it used existing (intact) groups. A quasi-experimental design was assumed. However, because the students were originally randomly assigned into their sections upon registering at the university, the design was in fact close to a random experiment. During the two-week experiment, both instructors (referred to as Instructor A and Instructor B) covered the same lecture material and the same application examples, but used different media. In the first of its two weeks, Instructor A presented a topic using hypermedia instruction, while Instructor B presented the same topic using conventional instruction. In the following week, to present the next topic, the instructors switched the instructional media. Students in both groups were given a test at the end of each of the two lectures. During the experiment, students' online access to the relevant hypermedia materials was blocked until after the lecture. After the second lecture, all students were asked to review the material asynchronously online. The following week, both groups were given a third test, which combined topics from the previous two.

Throughout the semester, students were given the same tests, worked on the same design project and assignments (with randomized individual parameters), and responded to the same surveys. Since lectures for both groups always took place at the same time, the integrity of testing was not affected, and students could not “double up” on their exposure to different instructional media or instructors. Course instructors used a team approach for all test evaluations, with each instructor “blind grading” half of any test items across all students and across both groups. Evaluations of the design project and assignments required subjective judgements and were not only team-graded but also double-vetted.

Based on the literature and on observations from the author's previous studies[3, 4], three specific hypotheses were formulated. The first hypothesis was that hypermedia-instructed students would perform significantly better, as compared with the conventionally instructed students, on each of the first two tests of the Switched Replications experiment, regardless of who was the instructor. The second hypothesis was that there would be no significant differences in the student performance between the two groups on the third test of the experiment. The third hypothesis was that there would be no significant differences in the student performance between the groups, instructed by Instructor A and Instructor B respectively, on the overall measure of the course achievement, the course grade CG.

The results and a detailed analysis of possible threats to the study validity was then to be used to either accept, or reject, the main hypothesis formulated from this study and the studies preceding it[3, 4]. The hypothesis was that the significantly better academic achievement observed among hypermedia-instructed students in the 1999-2001 experiments was a result of the choice of instructional media, i.e. hypermedia technology, and not a result of the lack of controls, or alternative explanations. These, as quoted in the literature, include selection bias, sample size, instructor differences, novelty factor, differences in instructional design, and social threats to internal validity of the study.

## III. Results

The 2001 study used a measure of previous academic achievement PAA as a benchmark[5]. PAA was based on a term grade point average in the semester preceding ELE639. A one way analysis of variance (ANOVA) was used to analyze the PAA scores of the two lecture groups, as well as of the two smaller groups of students who participated in the Switched Replications experiment (Table 1). Group differences in both cases were statistically negligible.

Table 1: Comparison of PAA Scores

| | Semester of Hypermedia (12 out of 13 weeks) | | Switched Replications (2 weeks) | |
|---|---|---|---|---|
| | Instructor A Group | Instructor B Group | Instructor A Group | Instructor B Group |
| Pooled Mean | 73.95 | | 73.86 | |
| Pooled STD | 6.677 | | 6.741 | |
| Group Mean | 73.52 | 74.41 | 73.29 | 74.35 |
| Number of students | 66 | 62 | 47 | 55 |
| ANOVA Statistic | F=0.567, df=1, p=0.453 | | F=0.624, df=1, p=0.432 | |
| Effect Size ES | -0.13 | | -0.16 | |

The experiment took place halfway through the course (eighth and ninth week) and both lectures covered advanced topics that were unfamiliar to the students. Only post-tests were therefore administered, and ANOVA was again used to analyze the results. Test 3 consisted of topics from both previous tests, which allowed estimating the student improvement on test items associated with Test 1 and Test 2, following the asynchronous review. Table 2 shows the results of all three tests, as well as components of Test 3, corresponding to the topics from Test 1 and Test 2 (referred to as Test 3a, and Test 3b, respectively). The outcomes were also coded in terms of the effect size ES, used in meta-analysis approach to create a common scale of measurement[13]. It is defined as the difference between the mean score of two groups, divided by the pooled standard deviation[8, 14]. The American Educational Research Association (AERA) recommends that all comparison studies should quote the effect size[15]. ES of 0.30 is considered a moderate, but significant group difference[8, 14].

Table 2: Statistical Analysis of Grades (in %) in Switched Replications Experiment

| | Test 1 | Test 2 | Test 3 | Test 3a | Test 3b |
|---|---|---|---|---|---|
| Pooled Mean | 60.05 | 41.67 | 76.82 | 85.36 | 62.58 |
| Group 1 | 70.28 | 35.46 | 77.93 | 85.53 | 65.25 |
| Group 2 | 51.31 | 46.97 | 75.87 | 85.21 | 60.3 |
| Pooled STD | 20.945 | 21.493 | 14.445 | 12.7614 | 28.7875 |
| ANOVA statistic | F=25.925 p=0.0005** | F=7.748 p=0.006** | F=0.510 p=0.477 | F=0.016 p=0.900 | F=0.746 p=0.390 |
| ES | 0.91 | -0.54 | 0.14 | 0.03 | 0.17 |

** significant at .01 level (2 tailed)
* significant at .05 level (2 tailed)

The analysis of the results revealed statistically significant differences in academic performance on

Test 1 and Test 2, with the hypermedia-instructed group performing better each time, with ES=0.91 and 0.54, respectively (Table 2). After all students were given an opportunity for an asynchronous review of both topics using the course website, differences between the groups were statistically negligible (Table 2). This finding is consistent with the beneficial effects of hypermedia treatment on the comprehensive course achievement[3, 4], observed in 1999 and in 2000. Relatively large standard deviations for Test 1 and Test 2 reflect a single event nature of the experiment, as opposed to the comprehensive measures, such as in PAA, where the standard deviation was much smaller (Table 1). However, the standard deviation of Test 3 was reduced, particularly on items corresponding to the simpler Test 1, reflecting an average improvement in comprehension.

While care was taken to choose topics where comparable instructional media would be used in both presentations, due to the logistics of the course schedule the topics were not of comparable difficulty. The first lecture covered concepts of polar plots and phase and gain margins. The second lecture covered a much more challenging, and theoretical, topic (Cauchy's Theorem and Nyquist criterion of stability). Subsequently, lower marks and smaller, though still significant, differences were observed in Test 2. However, the average student improvement on items associated with Test 1 and Test 2 was similar. The improvement on Test 1 items was 25 points (avg. mark 85.4%, group difference ES=0.03), and the improvement on items associated with Test 2 was 20 points (avg. mark 62.6%, group difference ES=0.17).

Table 3 shows the group difference in the comprehensive measure of achievement in the 13-week course where, except for one of the two weeks of the Switched Replications experiment, all students received hypermedia instruction. Analysis of covariance (ANCOVA) was used to assess differences in the course grade CG between groups to adjust for variance in PAA and to increase precision[16]. As is customary for the F-ratio statistic from ANCOVA, the group means in CG were adjusted for the covariate PAA to allow a more meaningful interpretation of the results[16]. As Table 3 shows, differences in course grade CG, as well as in the final exam grade were negligible.

Table 3: Statistical Analysis of Course Grades (in %)

| | Comprehensive Course Grade CG | | Final Examination Grade | |
|---|---|---|---|---|
| | Instructor A Group | Instructor B Group | Instructor A Group | Instructor B Group |
| Pooled Mean | 75.38 | | 60.57 | |
| Pooled STD | 6.521 | | 11.093 | |
| Group Mean | 75.23 | 75.53 | 60.49 | 60.64 |
| ANCOVA Statistic | F=0.068, df=1, p=0.795 | | F=0.005, df=1, p=0.944 | |
| Effect Size ES | -0.05 | | -0.01 | |

## IV. Discussion of Threats to Study Validity

There are four types of validity[17, 18]: conclusion validity, internal validity, construct validity, and external validity. Conclusion validity is concerned with a question whether the relationship between two variables exists. Internal validity is concerned with a question whether the relationship is a causal one. Construct validity is concerned with a question whether the treatment reflected well a construct of the treatment and whether the measure reflected well an idea of the construct of the measure. External validity is concerned with a question whether the observed effects can be generalized to other persons, times, and places. Threats to validity must be addressed when designing any evaluation study.

Internal validity is one of the most important issues as it deals with inferences about cause-effect relationships. Cook and Campbell[18] list three conditions that must be met before one can infer that a cause-effect relationship exists. They are: covariation- changes in the presumed cause must be related to changes in the presumed effect; temporal precedence - the presumed cause must occur prior to the presumed effect; and no plausible alternative explanations - the presumed cause must be the only reasonable explanation for changes in the outcome measures. The third condition is the most difficult to meet. A number of plausible alternative explanations, or threats to internal validity of a study, may exist. In the multiple group design (such as used in the 1999-2001 studies), threats to internal validity describe any factor other than the treatment that leads to differences between groups. Threats may be minimized through a combination of argument, measurement and observation, design, and analysis[9, 19].

In the literature, comparison studies are often criticized for the lack of adequate controls[7, 9 10, 11] and various compelling alternative explanations for group differences are suggested[6, 9, 11, 12, 17]. Presence of significant group differences, such as observed in the Switched Replications experiment (Table 2) and in the previous studies[3, 4], in itself is not sufficient to confirm the hypothesis that hypermedia instruction results in better student achievement. All threats to the validity of the study have to be rejected as well. The internal validity threats include sample size, selection bias, instructor differences, non-equivalent instructional method, novelty factor, and social threats. These threats will be next discussed in detail.

### Sample Size

Whether in a randomly allocated group design, or in a nonequivalent group design, where the two groups are presumed to be similar, reasonable sample sizes are required for the probabilistic equivalence of the groups to take place. The larger sample size also increases statistical power of the subsequent analysis, and improves external validity. Small sample size is among the most often reported problems in empirical studies. In the 1999-2001 studies conducted by the author, reasonable sample sizes were attained because of high participation rates, drawing from a pool of all students registered in the course. In the Switched Replications experiment, the sample size was n=102. In the full semester experiments, the sample sizes were n=94 in 1999 and in 2000, and n=128 in 2001.

Selection Bias

Selection bias is the main multiple group threat to the internal validity, and refers to the fact that the groups may not have been comparable before the study[17]. No meaningful comparisons can be made if the two populations being compared are significantly different in their demographic, previous knowledge, or academic, attributes. Some studies compare an undergraduate class receiving on-campus instruction, with a graduate or distance education class (typically adult professionals) receiving Internet instruction[20, 21, 22]. Lockee, Burton and Cross[23] point out that results of such studies are biased towards the "no significant difference" phenomenon. This may be more a reflection of graduate and adult students' motivation and determination than of the equivalency of the instructional method. Similarly, when students represent different disciplines[24] or when groups are self-selected[25], meaningful analysis is difficult. Glinkowski et al.[25] report a failure to demonstrate continuing positive effects of hypermedia instruction, when in the second year of experiment, lower achieving students seeking to boost their grades enrolled in a hypermedia-instructed section, which was considered to be "better". Often studies fail to account for large differences in prior student achievement between the treatment and control group[26, 27].

In contrast, the study used existing groups, in which the populations had similar prior academic achievement (Table 1). Students enrolled in the course represented a cross-section of Toronto's multicultural population. They studied full-time and overwhelmingly entered university directly from high school. There were no students with a workplace experience beyond summer jobs, no students with differing majors, and no graduate students. The course represented an introduction to a subject area (control systems) to which none of the students were exposed before in their studies. The failure rate in the course has been traditionally close to zero, and as a result there were virtually no individuals who repeated the course, thus having a significant previous knowledge of the subject. All students functioned in a networked computing environment, used simulation software in design projects, digital signal processing-based data acquisition systems in the lab, and email. Gender differences were not studied, due to a very small number of female students in the program (less than 10%). Demographic background, prior experience, and exposure to technology were therefore similar for the whole population in the study. As explained in the Methods section, because of the original registration process the makeup of the groups was close to randomized. Therefore, the two groups were assumed to be probabilistically near equivalent, and thus the selection bias threat could be rejected. The same argument could be made for the populations in the 1999 and 2000 studies[3, 4].

Instructor Differences

Instructor attributes are pointed out as a possible source of group differences[12, 28]. In studies dealing with large populations of introductory classes several different instructors are used in teaching different sections of both treatment groups so that an averaging effect can take place[29]. However, such large groups usually have other serious confounding factors to deal with (different majors, mix of undergraduate and graduate students, mix of liberal arts and science students, varying levels of motivation in freshman classes, full-time and part-time, etc.). Where smaller populations are involved, teaching both the experimental and the control group by the same instructor is recommended to remove the instructor bias[12].

However, one could argue that teaching by the same instructor would in fact introduce bias, based on personal preferences for one type of the instruction. It is undeniable that expertise plays a significant role - for example, comparisons between groups taught by a teaching assistant, typically a graduate student, and a tenured professor with many years of experience, would be unreliable. However, when the instructors have comparable professional expertise, deemed acceptable by their teaching institution, and teach the course following the same instructional design and evaluation schemes, their personal attributes such as teaching style, personality traits, etc. should have little effect on the student achievement.

Having either several different instructors or just one instructor for ELE639 was not possible, for administrative reasons. Since 1996, three different instructors were involved in the course, two at a time. They had different personal traits and learning and teaching styles. However, they all had comparable professional attributes (tenure, years of teaching experience industrial expertise, area of specialization), similar teaching evaluations, were aware of the purpose of the study, and presumably were teaching to the best of their abilities. All three have taught the course before using the conventional approach (i.e. "chalk & talk" lectures), and two have used hypermedia instruction (Table 4).

Table 4: Effect Sizes for Instructors Differences: A, B, C- Instructors; h, c - Instructional Media; '01a, '01b - Switched Replications Test 1 and Test 2

| '96 | '97 | '98 | '99 | '00 | '01 | '01a | '01b |
|---|---|---|---|---|---|---|---|
| A-c | A-c | B-c | A-h | A-h | A-h | A-h | A-c |
| C-c | B-c | C-c | C-c | B-c | B-h | B-c | B-h |
| ES=0.03 | ES=-0.05 | ES=-0.18 | ES=0.68 | ES=0.71 | ES=-0.05 | ES=0.91 | ES=0.54 |

Table 4 shows that despite different personal attributes, no statistically significant differences were noted whenever the same instructional media was used, regardless of which instructors were teaching the course. However, in all instances of instructors using different instructional media, hypermedia instructed groups performed significantly better. This suggests that the choice of instructional media had a strong effect on student achievement, while instructor differences were a weak effect, and thus could be rejected as the threat to the internal validity of the study.

## Differences in Instructional Design

The lack of methodological control between instruction in hypermedia and control groups is among most often quoted compelling alternative explanations for group differences in comparison studies[7, 12, 28, 30]. Many studies reported in the literature, especially comparing online education with classroom education, fall into this category, providing fuel for critics[11, 23]. It is often stated in the literature that the time and effort invested in creating a technology-enhanced course involves instructional design improvements for the students in treatments with the newer media, which are not implemented for those receiving traditional instruction[6, 12, 28, 30].

In a well-known exchange with Kozma[31], Clark[30] argued that comparison studies confuse media with methods, thereby risking unexamined important rival hypotheses. He attributed differences reported in empirical studies to dissimilar teaching methods and argued that a properly designed and controlled study would show no significant differences between a hypermedia treated group and a control group. Clark's argument was that research should focus on teaching-learning methods, not on questions of the media. On the other hand, Kozma argued that technology is not irrelevant and that research should focus on which technologies are best for supporting the best methods of teaching and learning. The truth is, as always, somewhere in between. The educational technology itself does not produce learning; and what matters is how it is used. The evidence is accumulating that hypermedia is most effective in the context of student-centered education, where it has to be grounded firmly in curriculum goals and incorporated into the instructional process[32, 33, 34]. However, to make valid comparisons, the same principles should be implemented in control group, where no hypermedia instruction is used.

As explained in the Methods section, ELE639 was tightly coordinated, and care was taken to make sure that the course used the same progressive, student-centered instructional design, regardless of the choice of the instructional media in the classroom. All students engaged in active, collaborative, design-based learning (lab projects, computer assignments). The same schedule of lecture topics was followed and all academic assessment tools (tests, projects, and final examination) were prepared collaboratively. In 2001, the Switched Replications experiment was embedded in an environment where all students received hypermedia treatment throughout the semester, except for the two weeks of the experiment. During those two weeks, both instructors covered the same lecture material and the same application examples, but used different media. Thus different instructional design could be rejected as a plausible explanation for the observed differences both in the 1999-2000 semester-long studies as well as in the Switched Replications experiment in 2001.

Novelty Factor

Another alternative explanation for observed performance gains in a hypermedia-instructed group could be a novelty factor. In some studies, while performance improved when the new format was introduced, once the novelty of the instructional method wore off, the student achievement dropped to previous levels[35, 36]. It is also sometimes asserted that the impact of the treatment in comparison studies may be reduced as the use of computer technology is becoming widespread, directly, as well as indirectly, in education[37]. However, there is no indication that this happened in the study. Based on the university data, as well as on the results of exit surveys, since 1999 the student exposure to the WebCT environment in their liberal studies has increased significantly. As well, the level of home computer ownership and the Internet access has grown from 50 % to almost 90%. Yet the overall group differences were similar in 1999, in 2000 as well as in the 2001 Switched Replications experiment (Table 4). To specifically avoid the novelty effect, the latter was conducted after seven weeks of a uniform exposure to hypermedia instruction. Since the student cohort in the study was close to probabilistically equivalent, the novelty factor would have had an equal bias on both groups. Thus, the novelty factor could be rejected as the explanation for the observed differences.

Social Threats

Classroom action research is a human activity. As such, its results are affected by the human interactions involved. Social threats to internal validity refer to the social pressures in the research context that can lead to group differences. They include diffusion or imitation of treatment, compensatory equalization of treatment, compensatory rivalry, and resentful demoralization[17]. The best way to minimize social threats is to isolate the two groups. However, this is typically impossible when experiments are conducted within the constraints of the institutional setting.

Diffusion or imitation of treatment occurs when participants from the two groups share experiences or imitate the program. Compensatory equalization of treatment involves people who manage the research project, or administrators, who may feel compelled to allow students' reassignments between the groups or to equalize the perceived unequal treatment in some way, thus directly negating the treatment effect. Diffusion and compensatory equalization tend to equalize the results, biasing them towards the “no difference” finding. Compensatory rivalry and resentful demoralization are two opposite effects. In the first case, the comparison group develops a competitive attitude towards students receiving treatment. This results in equalization of the post-test results. In the second case, the comparison group gets discouraged, frustrated and gives up. This may result in exaggerated group differences.

Compensatory equalization did not occur in 1999-2001, as both course instructors consistently adhered to the agreed-upon format, the administration did not interfere, and no reassignments between groups were allowed. However, a possible presence of diffusion, compensatory rivalry or resentful demoralization was difficult to detect or dismiss. As the literature review suggests, students are very much aware of differential treatment, especially in the consecutive offerings of the course, when a perception of a “better” course begins to build around the hypermedia-aided offering[25, 38]. They actively collaborate to achieve the maximum perceived advantage in the course. Such behavior effectively negates the intended effect and may lead to a diffusion of the results. It is very difficult to prevent such actions, since that could contravene departmental policies (for example, by conducting attendance checks).

In 1999-2000, the diffusion of treatment most likely did take place, as the two treatment groups were not, and could not, be isolated. Although a casual observation of lecture attendance patterns did not suggest any students attending lectures in hypermedia classes against their original assignment, students could have shared access to the website, which theoretically was restricted to the experimental group only, or could have shared materials printed off the website. Students of different levels of achievement and with different learning styles collaborated on lab projects, as intended in the course design, but they also most likely studied together and collaborated on assignments, not only within, but also across the treatment groups. However, since the diffusion of treatment tends to reduce group differences, this threat did not affect the study validity. As well, in the 2001 Switched Replications experiment, the immediate assessments following the lectures excluded the possibility of it having taken place.

Over two consecutive offerings of ELE639 (1999-2000), a perception of a "better" course did build around the hypermedia version. There were a few individual complaints when requests for reassignment to the experimental group were denied. As the awareness of the differential effects of treatment, with better results for the hypermedia group, filtered to the student population, the possibility increased that compensatory rivalry or resentful demoralization could become a problem should the differential treatment continue. However, the hypermedia instruction was extended to all students in 2001 because withholding treatment from one of the groups in light of the previous positive results would have been incompatible with the objective of increased learning, and simply unethical. The uniform treatment thus excluded the possibility of either of these two social threats taking place in 2001.

<u>Construct Validity</u>

Construct validity refers to the degree to which inferences can be made from the operationalizations in the study to the theoretical constructs on which they were based[17]. The use of expert opinion minimizes threats to construct validity[17, 18]. Construct validity of the study, one aspect of which is the relationship between learning styles and hypermedia, is supported by the fact that the observed relationships[3, 4, 5] confirm theoretical assumptions of the Felder Learning Model[39, 40], which is thought to be particularly useful in engineering education.

Construct validity also hinges on the quality of measurement to capture the effects of the treatment. In the study, expert opinion was used in all measurements (standard academic evaluations, an attitude survey, and a learning style inventory). Student performance in the course was evaluated using a series of instruments developed to meet the criteria set by the Canadian Engineering Accreditation Board (CEAB) for an accredited engineering undergraduate program, and those of the department. They included tests, assignments, projects and final exams. All course evaluations were developed collaboratively by two domain experts teaching the course, who were tenured professors with many years of professional and teaching experience in the field of control systems. The attitude survey was developed with the assistance of an expert in psychology. A subsequent analysis of 2001 survey results showed strong internal consistency of the scales, with Cronbach alpha of 0.86. The study also used a learning style inventory[41] developed for the Felder Learning Model, which in 2001 received top reviews from MERLOT

(Multimedia Educational Resources for Learning and Online Teaching).

Cook and Campbell[18] list additional threats to construct validity, such as mono-operation bias and interaction of different treatments. Mono-operation bias threat is a possibility that only a narrow aspect of the treatment method is used. This threat was not an issue because the author's expertise in the area of hypermedia courseware development, and considerable time spent on creating the courseware constitute an assurance that different attributes of hypermedia were utilized. Treatment interaction (the effect recorded is in fact due to some other program or treatment) was not an issue because there were no other courses taken by the students in the study that used the hypermedia classroom instruction. Students may have been exposed to online supplementary materials in their elective studies before, but the extent of the hypermedia usage in the classroom would be minimal. As explained in the companion paper[5], the hypermedia-instructed course developed by the author is still unique in engineering programs at Ryerson.

External Validity

External validity is the degree to which the conclusions of a study would hold for other persons, in other places and at other times[17]. External validity deals with generalizations of the conclusions. Threats to external validity are any factors that would preclude making such generalizations to contexts other than the specific study that took place. One way to improve external validity is to include random sampling representative of a large population. This approach is not realistic in studies that use cohorts available in institutional settings. The only other way to increase external validity is to replicate the study in a variety of places, with different people at different times. Group differences were observed when three different cohorts of students took part in the study (1999, 2000, and 2001).

Conclusion Validity

Conclusion validity is the degree to which conclusions we reach about the relationships in our data are reasonable[17]. It is also sometimes referred to as statistical validity. Threats to conclusion validity are essentially two kinds of errors one can make about relationships: the conclusion that there is no relationship when in fact there is one, and the conclusion that there is a relationship when in fact there is none. In the first case, threats could arise from low reliability of measures (poor instrument design or implementation) or the low statistical power. In the second case, the threats involve “fishing” for results and error rate problems (adjusting the significance levels). The way to improve the conclusion validity is to have good statistical power (sample size, effect size, significance level), good reliability and good implementation. It is the author's contention that these conditions have been met in the 1999-2001 studies.

V. Conclusions and Future Work

The course in which the study was located, used a student-centered approach with emphasis on active, experiential, problem-based learning, and had a significant design project where teamwork and communication skills were important. The project required a tight coordination of the course. All students functioned in the same computer-assisted learning environment (access to computer network, simulation software, computer-controlled laboratory setup). Progressive approach, project orientation and coordination ensured that the instructional design of the course remained the same regardless of the instructional media used in the lectures. As discussed, treatment groups were probabilistically near equivalent. Two previous comparison studies, conducted by the author[3, 4] showed that hypermedia-instructed students had better academic achievement than their conventionally instructed counterparts did. The results of the Switched Replications experiment in 2001 confirmed that previously observed treatment effects were repeatable. Over time, no significant group differences were observed when instructors used the same instructional media (either "chalk & talk" lectures, or hypermedia presentations), but there were significant differences between groups when the instructors used differentiated instructional media (Table 4). The hypermedia-instructed group performed significantly better each time, regardless of which instructor made presentations. This suggests not only that the treatment effects were repeatable, but that the choice of instructional media had a strong effect on achievement, while the effect of instructor differences was negligible.

Discussion of the threats to the study validity allowed for a rejection of differences in the instructional design, the instructor differences, as well as the novelty factor and the social factors, as possible alternative explanations for the observed differences. The hypothesis that the hypermedia instruction, and not any alternative explanations, is responsible for the observed improvement in academic achievement was thus confirmed.

The results showing increased learning from the use of instructional hypermedia need to be interpreted carefully. Many academic institutions face challenges of having to improve outcomes and accessibility while dealing with shrinking budgets, which forces them to look to technology for answers. However, the rush towards new technologies without first asking the hard questions about appropriate educational goals, may end with disappointing and wasteful results. As the literature suggests, and the results of the study confirm, for the technology to be effective, it has to support a sound educational approach[32, 33, 34]. Research on distance education, including the newer form of WWW-based asynchronous instruction, generally reports that programs for adult learners produce learning outcomes and achievement at least equal to conventional instruction[22, 42, 43].

Yet replacing face-to-face instruction in an undergraduate environment with large Internet-based classes as a cost-cutting measure does not guarantee academic success. A community of peers and personal contact with instructors are major contributing factors to the learning process of a student whose world experience, motivation and maturity are still developing. However, there is an indication that some routine materials can be delivered through WWW, provided that the liberated classroom time can be spent on higher value-added activities, such as mentoring, experiential and problem-solving, design-based activities, to the benefit of the students[44]. In the Switched Replications experiment, results of Test 3 showed that asynchronous online review could be an effective tool in helping students "catch up" on the material not covered, or not adequately covered, in the class. However, the in-class and online activities need to be well integrated into the course design and assessment scheme and students should not be overburdened with the requirement of independent study on top of the class activities. Being required to explore new material outside the traditional classroom lecture may create a negative view towards the process[45] especially if the Web-supported course takes up a disproportionate amount of students' available time.

The author concludes that hypermedia instruction in the classroom can improve learning outcomes and enhance the quality of student-instructor interactions, as long as educational and learning goals are paramount. It is quite telling that the students consider the contact with the instructor as one of the most integral parts of learning. In the exit survey, the vast majority (77% in 2001, 73% in 2000) indicated that technology should supplement, and not replace, student-instructor interactions. This is consistent with the findings in the literature[46, 47]. We should therefore remember that technology is not a panacea for problems in the educational system, and that hypermedia alone cannot equitably replace human interactions that contribute to learning.

Bibliography


1. Zywno, M.S. & Pereira, D., Innovative Initiatives in Control Education at Ryerson Polytechnic University - Fuzzy-Logic Control of the 3D-Helicopter Simulator, Proceedings of 2000 American Control Conference, Chicago, IL, (2000).
2. Zywno, M.S. & Kennedy, D.C., Integrating the Internet, Multimedia Components, and Hands-on Experimentation into Problem-Based Control Education, Proceedings of 30th ASEE/IEEE Frontiers in Education Conference, Session T2D-5, Kansas City, Missouri, (2000). Online at: http://fie.engrng.pitt.edu/fie2000/papers/1120.pdf
3. Zywno, M.S., & Waalen, J. K., Analysis of Student Outcomes and Attitudes in a Technology-enabled Control Education at Ryerson - a Case Study, Global Journal of Engineering Education, Vol. 5, No.1, pp. 49-56, Australia, (2001). Online at: http://www.eng.monash.edu.au/uicee/gjee/vol5no1/Zwyno&Waalen.pdf
4. Zywno, M.S., and Waalen, J.K., The Effect of Hypermedia Instruction on Achievement and Attitudes of Students with Different Learning Styles, Proceedings of 2001 ASEE Annual Conference and Exposition, Session 1330, Albuquerque, NM, (2001).
5. Zywno, M.S., Instructional Technology, Learning Styles and Academic Achievement, 2002 ASEE Annual Conference and Exposition, Session 2422, Montreal, Quebec, (2002).
6. Ayersman, D.J., Reviewing the Research on Hypermedia-Based Learning, Journal of Research on Computing in Education, Vol. 28, No.4, pp. 500-525, (1996).

7. Dillon, A. & Gabbard, R., Hypermedia as an educational technology: A review of the quantitative research literature on learner comprehension, control and style, Review of Educational Research, Washington, Vol. 68, no 3, pp. 322-349, (1998).
8. Liao, Y.C., Effects of Hypermedia on Students' Achievement: A Meta-Analysis, Journal of Educational Multimedia and Hypermedia, Vol. 8, No. 3, pp. 255-277, (1999).
9. Reeves, T.C., Enhancing the Worth of Instructional Technology Research through "Design Experiments" and Other Development Research Strategies, Session 41.29, Annual Meeting of the American Educational Research Association, New Orleans, LA, (2000). Online at: http://itech1.coe.uga.edu/~treeves/AERA2000Reeves.pdf
10. Reeves, T.C., Pseudoscience in computer-based education: The case of learner control, Journal of Computer-Based Instruction, Vol. 20, No. 3, pp.39-46, (1993).
11. Neal, E., Does Using Technology in Instruction Enhance Learning? Or, the Artless Stateof Comparative Research. Technology Source [online], (1998). Available at: http://horizon.uncs.edu/ts/commentary/1998-06.asp
12. Weller, H.G., Assessing the Impact of Computer-Based Learning in Science, Journal of Research on Computing in Education, Vol. 28, No.4, (1996).
13. Glass, G.V. Meta-Analysis: An Approach to the Synthesis of Research Results. Journal of Research in Science Teaching, Vol. 19, No. 2, pp. 93-112, (1982).
14. Kadiyala, M. & Crynes, B.L., A Review of Literature on Effectiveness of Use of Information Technology in Education, Journal of Engineering Education, Vol. 89, No. 2, pp. 177-184, (2000).
15. Thompson, B., AERA Editorial Policies Regarding Statistical Significance Testing: Three Suggested Reforms, Educational Researcher, Vol. 25, No. 2, pp. 26-30, (1996).
16. Wildt, A.R. & Ahtola, O., Analysis of Covariance, Sage University paper series on Quantitative Applications in Social Science, series no. 07-012, Beverly Hills and London: Sage Publications, (1978).
17. Trochim, W., Research Methods Knowledge Base, 2nd ed., Cornell University, (1999). Online at: http://trochim.human.cornell.edu/kb/
18. Cook, T.D. & Campbell, D.T., Quasi-Experimentation: Design and Analysis for Field Settings, Rand McNally, Chicago, Illinois, (1979).
19. Trochim, W. & Land, D., Designing Designs for Research, The Researcher, Vol. 1, No. 1, pp. 1-6, (1982).
20. Davis, J.L., Computer-Assisted Distance Learning, Part II: Examination Performance of Students On and Off Campus, Journal of Engineering Education, Vol. 85, No. 1, pp. 77-82, (1996).
21. Madviwalla, M. & Hovav, A., Adapting business process redesign concepts to learning processes, Business Process management Journal, Vol. 4, No. 3, (1998).
22. Evans, R.M., Murray, S.L., Daily, M. & Hall, R, Effectiveness of an Internet-Based Graduate Engineering Management Course, Journal of Engineering Education, Vol. 89, No. 2, pp. 63-71, (2000).
23. Lockee, B.B., Burton, J.K. & Cross, L.H., No Comparison: Distance Education Finds a New Use for "No Significant Difference", Educational Technology Research and Development, Vol. 47, No.3, pp. 33-42, (1999).
24. Weinberger, C., and Mutharasan, R., Fundamentals of Manufacturing- MultiMedia Modules for Contextual Learning, Session F2B - Learning technologies II, Proceedings of the 28th ASEE/IEEE Frontiers in Education Conference, Tempe, AR, (1998).
25. Glinkowski, M., Hylan, J.& Lister, B., A New, Studio-based, Multimedia Dynamic Systems Course: Does it really work?, Proceedings of the 27th ASEE/IEEE Frontiers in Education Conference, Pittsburgh, PA, (1997).
26. Borchert, R., Jensen, D., Yates, D., Development and Assessment of hands-on and Visualization Modules for Enhancement of learning in Mechanics, Sesion 1368, 1999 ASEE Annual Conference & Exposition, Charlotte, NC, (1999).
27. Siegle, D. & Foster, T., Effects of Laptop Computers with Multimedia and Presentation Software on Student

Achievement, Proceedings of the Annual Meeting of the American Education Research Association, New Orleans, LA, (2000). Available online at: http://www.nsoe.uconn.edu/siegle/Conferences/Siegle_Foster.htm

28. Fletcher-Flinn, C.M. & Gravatt, B., The Efficacy of Computer-Assisted Instruction (CAI): A Meta-Analysis, Journal of Educational Multimedia and Hypermedia, Vol. 12, No. 3, pp. 219-242, (1995).
29. Oakley, B. II, A Virtual Classroom Approach to Teaching Circuit Analysis, IEEE Transactions on Education, Vol. 39, No. 3, pp. 287- 296, (1996).
30. Clark, R.E., Media Will Never Influence Learning, Educational Technology Research and Development, Vol. 42, No. 2, pp. 21-29, (1994).
31. Kozma, R.B., Will Media Influence Learning? , Reframing the Debate, Educational Technology Research and Development, Vol. 42, No. 2, pp. 7-19, (1994).
32. Catalano, G. D. & Catalano, K. C., Transformation: From Teacher-Centered to Student-Centered Engineering Education, Journal of Engineering Education, Vol. 88, No. 1, pp. 59-64, (1999).
33. Bransford, J., Brophy, S. & Williams, S., When Computer Technologies Meet the Learning Sciences: Issues and Opportunities, Journal of Applied Developmental Psychology, Vol. 21, No. 1, pp. 59-84, (2000).
34. Poindexter, S.E. & Heck, B.S., Using the Web in Your Courses: What Can You Do? What Should You Do? IEEE Control Systems Magazine. Vol. 19, No. 1, pp. 83-92 (1999).
35. Kashy, E., Thoennessen, M., Tsai, Y., Davis, N.E & Wolfe, S.L., Using Networked Tools to Enhance Student Success Rates in Large Classes, Proceedings of the 27th ASEE/IEEE Frontiers in Education Conference, Pittsburgh, PA, (1997).
36. Pankuch, B.J., Multimedia in Lectures and on the World Wide Web, Issues of Education at Community Colleges: Essays by Fellows in the Mid-Career Fellowship at Princetown University, (1998). Online at; www.eclipse.net/~pankuch/Nesletter/Pages_News/Archive/MM-Prin-FINAL2.pdf
37. Christmann, E., Badgett, J. & Lucking, R., Progressive Comparison of the Effects of Computer-Assisted Instruction on the Academic Achievement of Secondary Students, Journal of Research on Computing in Education, Vol. 29, No. 4, pp. 325-337, (1997).
38. Coleman, J. N., Kinniment, D.J., Burns, F.P., Butler, T.J. & Koelmans, A.M., Effectiveness of Computer-aided Learning as a Direct Replacement for Lecturing in Degree-Level Electronics, IEEE Transactions on Education, Vol. 41, No. 3, pp. 177- 184, (1998).
39. Felder, R.M. & Silverman, L.K., Learning and Teaching Styles in Engineering Education, ASEE Journal of Engineering Education, Vol. 78, No. 7, pp. 674-681, (1988).
40. Felder, R. M. Matters of Style, *AEEE Prism*, December Issue, pp. 18-23. (1996).
41. Felder, R.M. & Soloman, B.A., Index of Learning Styles Questionnaire, North Carolina State University, (1998). Online at: http://www2.ncsu.edu/unity/lockers/users/f/felder/public/ILSdir/ILS-a.htm
42. Smeaton, A.F. & Keogh, G., An analysis of the use of virtual delivery of undergraduate lectures, Computers & Education, Vol. 32, pp.83-94, (1999).
43. Wright, V.H. & Marsh, G.E., Technology and Teaching: A Turning Point, (2000). Available at: http://www.computed.coe.wayne.edu/Vol5/Wright&Marsh.html
44. Wallace, D.R. & Weiner, S.T., How Might Classroom Time Be Used Given WWW-Based Lectures?, Journal of Engineering Education, Vol. 87, No. 3, pp. 237-248, (1998).
45. Smith, M. & Komerath, N., Learning More From Class Time: Technology Enhancement in the Classroom, Session 3202, Proceedings of the 2000 ASEE Annual Conference & Exposition, St. Louis, MO, (2000).
46. McIntyre, D.R., Wolff, F.G., An experiment with WWW interactive learning in university education, Computers And Education, Vol: 31, No. 3, pp. 255-264, (1998).

47. Sheppard, S., Reamon, D., Friedlander, L., Kerns, C., Leifer, L., Marincovich, M., and Toye, G., Assessment of Technology-Assisted Learning in Higher Education: It Requires new thinking by Universities and Colleges, Session T2B, Proceedings of the 28th ASEE/IEEE Frontiers in Education Conference, Tempe, AR, (1998).

MALGORZATA S. ZYWNO
Malgorzata Zywno is a Professor in the Department of Electrical and Computer Engineering at Ryerson University in Toronto, Canada. She received her M.Eng. degree in Electrical Engineering from the University of Toronto in 1990 and is currently in a doctoral program at Glasgow Caledonian University in Glasgow, Scotland, UK. She expects to complete her degree in 2003. Her teaching and research interests include control systems, modeling and system identification, and more recently, technology-aided pedagogy. She has been developing multimedia and Internet-based courseware, and has published and made presentations on the issues of implementations and efficacy of technology-mediated teaching, and faculty attitudes towards e-learning. Her other research interests include issues of recruitment and retention strategies for women in engineering. Professor Zywno is a member of the American Society for Engineering Education (ASEE), of the Institute of Electrical and Electronic Engineers (IEEE), and a registered Professional Engineer in the Province of Ontario, Canada.