CPATH $\tau\epsilon\chi\nu\eta$ Evaluation: Background, Experiment Design and Current Status

Robert J. Schalkoff
$\tau\epsilon\chi\nu\eta$ External Evaluator
Department of Electrical and Computer Engineering
Clemson University
Clemson, SC 29634
rjschal@clemson.edu

ABSTRACT

This paper describes the design and implementation of Level 2 and Level 3 evaluation procedures for the NSF CPATH $\tau\epsilon\chi\nu\eta$ project. Issues addressed include background, mandated evaluation parameters, cohort identification and selection, experiment design, instruments and evaluation timetables. Preliminary results of a 'dry run' in the Fall 2008 semester are shown. Future efforts are indicated.

Categories and Subject Descriptors

K.3 [**Computers and Education**]: Computer and Information Science Education

Keywords

Quasi-experimental design, NSF CPATH, Computing Education Assessment

1. BACKGROUND AND HISTORY OF THE EVALUATION EFFORT

In the Summer of 2007, the author was invited to serve as an external evaluator for an NSF-supported project: *TEXNH - Evaluation, Adoption and Extension*. This document does not address specific $\tau\epsilon\chi\nu\eta$ implementation aspects or methodologies, but rather focuses on the mandated project evaluation component. Two aspects of $\tau\epsilon\chi\nu\eta$ evaluation are addressed:

Aspect #1: Quasi-experimental design (Level 2)

Aspect #2: Other designs (Level 3); e.g., surveys

1.1 CPATH and the America Competes Act

In May of 2007, the Academic Competitiveness Council (ACC) released recommendations to integrate and coordinate federal education programs in science, technology, engineering and mathematics (STEM). This led to a congressional act (the 'America Competes' Act) requiring formal, quantitative review of all federal programs with a focus on math and science education[3]. CPATH is part of the STEM program.

A consequence of this act was an NSF-mandated Level 2 (quasi-experimental) evaluation for future STEM projects. A decision was made to 'retrofit' some existing CPATH projects with this requirement and external evaluators were chosen to facilitate an independent, quantitative evaluation. The $\tau\epsilon\chi\nu\eta$ project was in this group.

1.2 The 'Rosslyn 11'

SRI International hosted a NSF/CPATH Evaluators Meeting in July, 2008, at their headquarters near Rosslyn, VA. 11 project external evaluators were invited to explore rigorous (i.e., Level 2) evaluation methods for a number of diverse computing education projects[2]. This group became known as the 'Rosslyn 11'; this author was one of the original group.

The group intended to identify a common framework for collecting data measures for a project and program monitoring system. These would support an evaluation that determines the impact and effectiveness of CPATH. The focus was on identifying aspects within individual projects that facilitate a comparative (Level 2) study.

1.3 Potential Instruments

Data considered by the 'Rosslyn 11', NSF and SRI attendees[2] included:

1. Enrollments (and diversity)
2. Data typically used for ABET evaluations
3. Student attitude and belief data
4. Student retention and grades
5. Student employment
6. Faculty attitudes, beliefs, practices, collaboration, publications, and attrition
7. On-line electronic portfolio that are used for recruitment

The $\tau\epsilon\chi\nu\eta$ evaluation described herein focuses on item 2 for the Level 2 effort and item 3 for the Level 3 effort.

ACMSE '09 March 19-21, 2009, Clemson, SC, USA.

1.4 The 'Super-Evaluator'

Shortly after the July 2008 meeting, SRI International was named as the overall CPATH program evaluator. Web resources to support and disseminate the individual and overall CPATH evaluations are available[1]. Following this, SRI and NSF sponsored a combined PI/evaluator meeting in Washington, D.C. in November of 2008.

2. THE $\tau\epsilon\chi\nu\eta$ PROJECT

The NSF-funded CPATH $\tau\epsilon\chi\nu\eta$ project[1] is based upon problem-based instruction in the domain of computer generated visual media. The design process begins with an identification of course goals in terms of fundamental concepts and abilities to be acquired by the student during the course. Semester-long problems from the visual domain that will invoke an exploration of these structures are identified. *Developing ability and enthusiasm for computational problem-solving is of paramount importance.*

Two of the most significant current goals of $\tau\epsilon\chi\nu\eta$ are to work with colleagues at regional campuses of the University of North Carolina system in the adoption and evaluation of the $\tau\epsilon\chi\nu\eta$ approach and to extend the $\tau\epsilon\chi\nu\eta$ approach to include senior division computing courses and courses offered to non-computer science majors.

Some early assessment had been embedded into the $\tau\epsilon\chi\nu\eta$ project, including a diverse potpourri of assessment instruments and approaches (primarily written comments in which students related their perceptions of the class environment). A principal anecdotal finding was a heightened enthusiasm for computing among both students and instructors. This was considered in forming the evaluation questions in Section 4.

3. PRELIMINARY EVALUATION ISSUES

It became clear that the mandated evaluation based upon a quasi-experimental design would present many challenges. Perhaps the most significant was the identification of cohorts. A significant secondary issue concerns evaluation validity and threats to this validity. Especially significant is the effect of instructor variability.

A project review meeting was held at Clemson University on July 28, 2008 to define assessment procedures and tools. $\tau\epsilon\chi\nu\eta$ PIs were appraised of the mandated evaluation procedure and agreed to participate in the process. This meeting served as the catalyst for the subsequent overall project evaluation design and implementation.

4. EVALUATION QUESTIONS

The $\tau\epsilon\chi\nu\eta$ proposal makes a number of claims regarding $\tau\epsilon\chi\nu\eta$'s broad and positive impacts on the treatment group. These include:

1. It will attract more students to computing.
2. It will attract better students to computing.
3. It will motivate and thus retain the students that it attracts, and the students it produces will be well equipped to take on the challenges and opportunities provided by the rapidly evolving field of computing.

[1]The Texnh project is supported by the National Science Foundation under grant CPATH EAE: TEXNH - Evaluation, Adoption and Extension, award number 0722313.

In order to quantitatively verify these assertions, the formal evaluation process began with the design of a set of related questions:

1. Are there quantitative (*incremental ?*) benefits from the TEXNH approach? (student motivation, learning outcomes, instructor motivation)
2. Can TEXNH be exported to other institutions? (institution type/degree program/student background?)

5. LEVEL 2 DESIGN AND ASSUMPTIONS

The proposed Level 2 evaluation design is a Nonequivalent Group with Pre/Post Test Design[2], typically diagrammed as:
NR O_1 X O_2
NR O_1 O_2
where X represents $\tau\epsilon\chi\nu\eta$ (the treatment).

5.1 Issues

Issues which arose almost immediately concerned:

1. The identification of control groups;
2. Small sample class sizes and single sections (power);
3. Elimination of alternative hypotheses ('no plausible alternative explanation'); and
4. The fact that intervention was already underway (This is common to all 2007 CPATH projects within the 'Rosslyn 11' group).

5.2 Level 2 Instrument

As noted in Section 1.3, many institutions employ embedded questions in course exams for ABET evaluation purposes. Not all institutions involved seek ABET accreditation. Nevertheless, we propose the use of ABET-like embedded instruments to be incorporated in the $\tau\epsilon\chi\nu\eta$ (treatment) and (similar) non-$\tau\epsilon\chi\nu\eta$ course final exams. The pragmatics of this process are addressed in Section 9.

5.3 Identification of Cohorts

Cohorts were developed in the Summer and Fall of 2008 and fall into 4 main computing categories:

- Programming for Non-Majors
- CS I
- CS III
- CS DBMS

The specific cohorts proposed and actually used are described in Section 7.1.

6. LEVEL 3 DESIGN AND ASSUMPTIONS (ATTITUDINAL SURVEYS)

In addition, post-course attitude surveys were developed to collect feedback about student attitudes in both $\tau\epsilon\chi\nu\eta$ and (comparison) non-$\tau\epsilon\chi\nu\eta$ courses, especially possible $\tau\epsilon\chi\nu\eta$ motivational impact. The focus of the survey is on

[2]For the trials in Fall 2008, it will be a Post-test only.

the examples used to convey major concepts and student response to software development exercises. In the Fall of 2008, it was administered to students in both $\tau\epsilon\chi\nu\eta$ and (comparison) non-$\tau\epsilon\chi\nu\eta$ courses. The topics covered included satisfaction, effectiveness, opportunity to learn, assignments, engagement, learning. The survey format was 19 questions with an estimated completion time of 10 minutes. 3 sample questions used are shown in Figure 1.

7. I feel my classroom experience in this course generated enthusiasm for the subject.
A) Very little B) A little C) Somewhat D) A lot E) A great deal

8. I feel my software development experience in this course generated enthusiasm for the subject.
A) Very little B) A little C) Somewhat D) A lot E) A great deal

9. I feel the software development experience in this class used real-world examples.
A) Very little B) A little C) Somewhat D) A lot E) A great deal

Figure 1: Sample Questions From Attitudinal Survey

7. DATA COLLECTION, INCLUDING BASELINE DATA

7.1 Cohorts: Pragmatics

Initial cohorts for the quasi-experimental design were developed in the Fall of 2008 and fall into 4 main course categories, as shown in Table 1. A number of pragmatic concerns led to the exclusion of groups **QE3: CS III** and **QE4: DBMS** for the Fall 2008 evaluation. Note also that several comparison groups span more than a single semester.

Ref	School	Course	Coordinator	Pedagogy	Semester
Group QE1: Programming for NON-CS					
1	UNC-W	CSC-112	Narayan	$\tau\epsilon\chi\nu\eta$	
2	CLM	CPSC 111	Duchowski	conventional	F08
3	WCU	CS 140	Dalton	conventional	F08
Group QE2: CS I					
4	CLM	CPSC 101	Westall	$\tau\epsilon\chi\nu\eta$	
5	WCU	CS 150	Holliday	conventional	F08
6	UNC-W	CSC 121	Narayan	conventional	
Group QE3: CS III					
7	CLM	CPSC 212	Geist	$\tau\epsilon\chi\nu\eta$	F08
8	WCU	CS 351	Dalton	conventional	Sp09
9	UNC-W	CSC 332	Narayan	conventional	
Group QE4: DBMS					
10	CLM	CPSC 462	Pargas/Wang	$\tau\epsilon\chi\nu\eta$	F08/Sp09
11	WCU	CS 453	Holliday	$\tau\epsilon\chi\nu\eta$	Sp09
12	UNC-W	CSC 455	Narayan	conventional	

Table 1: Expanded Matrix of Quasi-experimental Design Cohorts and Course Coordinators, (Fall 2008). For Multiple Sections, Embedded Exam Results and Surveys are Combined.

8. COURSE COORDINATORS AND DATA ANONYMITY

8.1 Coordinator Roles

Coordinators for each course to be used as a cohort were identified. Course coordinators have a role in both the Level 2 and Level 3 evaluations. Long-term, the coordinators of each group are responsible for: collectively identifying a set of common objectives for learning outcomes in the course group; designing a collection of common final exam questions to evaluate success in meeting those outcomes; and forwarding the outcome data to the external evaluator. In addition, coordinators participate in the design and administration of attitudinal surveys and survey questions.

8.2 IRB Concerns

Since we are evaluating human subjects (students), a significant concern is conformance to possible IRB constraints. This issue may (and in fact did) vary with institution. IRB approval of activities was obtained for all schools involved. Generally, collected data from student exams and surveys needs to be 'anonymized' and aggregated. To this end, coordinators forward only aggregated and anonymized outcome data to the external evaluator.

9. PROGRESS IMPLEMENTING NEW DESIGN

For the trial evaluation in the first semester (Fall 2008 only), there were no pre-tests nor pre-surveys given as part of the external evaluation process. This was necessary since IRB approval had not yet been secured for all participants. Common exam questions (problems) in the respective groups were developed by the groups and this set was forwarded to the external evaluator. Using this set, a subset of 5 questions for QE-1 and 5 questions for Group QE-2 were selected by the evaluator and forwarded to the PIs for incorporation. The basis for problem selection was breadth of the problem space and balanced representation among the problem-generating cohorts. Problems were graded by the individual course instructors.

10. MAJOR CHALLENGES TO DATE

The Fall 2008 'dry run' was illustrative and should improve our design and implementation for Spring 2009 and subsequent semesters. Noteworthy in the data reporting and survey implementation from the Fall 2008 were the pragmatics of collecting, selecting and distributing sets of common exam problems at the very end of a semester, as well as administering surveys.

One potential difficulty is ensuring long-term cooperation with instructors, especially those not affiliated with the $\tau\epsilon\chi\nu\eta$ project. This appears to be a common concern within the 'Rosslyn 11' projects.

11. CURRENT STATUS OF $\tau\epsilon\chi\nu\eta$ EVALUATION (SPRING 2009)

The sample data presented from Fall 2008 collection is still under analysis, thus it is presented here without comment. No assessment is implied. A quantitative analysis is underway.

11.1 Preliminary Data from QE-1 and QE-2 Level 2 Instrument (Embedded Questions)

Sample and composite embedded question result data is presented in Figures 2 and 3. In these Figures, the y-axis shows an average of the embedded problem grades. The integer grading scale used ranged from 0 (no solution or clueless) to 4 (perfect score).

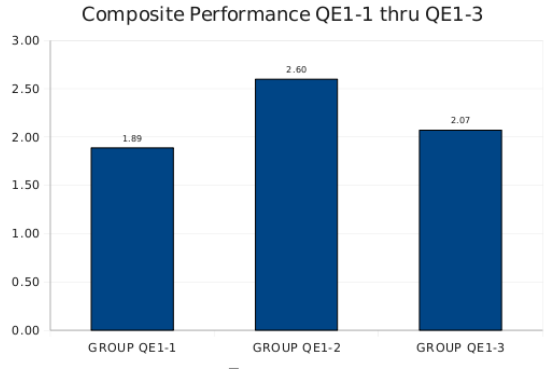

Figure 2: Fall 2008: QE1 Preliminary Composite Data, from Embedded Problems.

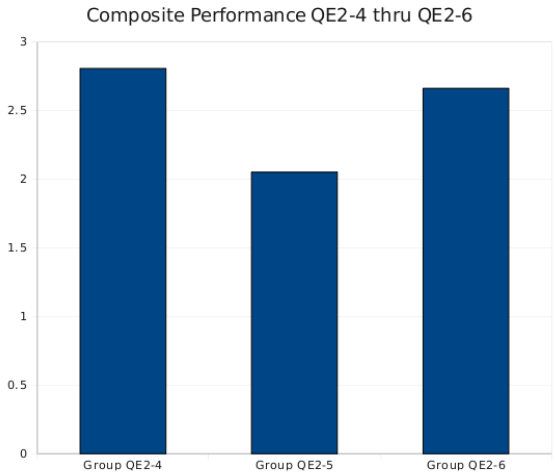

Figure 3: Fall 2008: QE2 Preliminary Composite Data, from Embedded Problems.

11.2 Preliminary Data from QE-1 and QE-2 Level 3 Instrument (Attitudinal Survey)

Using the resulting student data from several of the questions shown in Section 6, Figures 4 and 5 were developed for illustration. In Figure 4, the y-axis indicates the per-unit response to a single question (#7). In Figure 5, the y-axis represents an average of the weighted responses, where a 1 represents a response of `A) Very little` and a 5 indicates a response of `E) A great deal`.

12. FUTURE EVALUATION EFFORTS

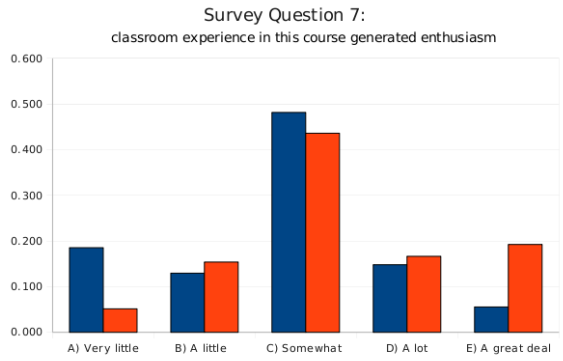

Figure 4: Example of Distribution of Responses for $\tau\epsilon\chi\nu\eta$ and Conventional Survey. Question #7 Shown. Leftmost (Darker) Column is $\tau\epsilon\chi\nu\eta$ Data.

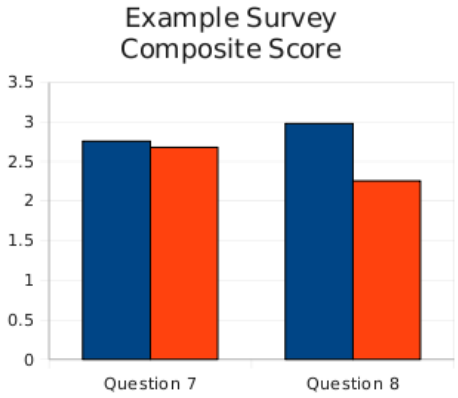

Figure 5: Example of $\tau\epsilon\chi\nu\eta$ and Conventional Survey Composite Data. Questions #7 and #8 Summarized. Leftmost (Darker) Column is $\tau\epsilon\chi\nu\eta$ Section(s).

We feel we have made significant progress towards a meaningful Level 2 evaluation of the $\tau\epsilon\chi\nu\eta$ project. Many open issues remain, including:

1. Continued development and refinement of the evaluation instruments. This includes both common exam and survey questions in collaboration with identified coordinators;
2. Complete evaluation of data from the Fall 2008 Semester;
3. Evaluation of data, including Pre-Tests, for Spring 2009;
4. Identification of cohorts and courses for Fall 2009 and beyond; and
5. Possibility of correlating surveys with embedded problem performance.

13. ACKNOWLEDGMENTS

The $\tau\epsilon\chi\nu\eta$ project is supported by the National Science Foundation under grant CPATH EAE: TEXNH - Evaluation, Adoption and Extension, award number 0722313. The investigators gratefully acknowledge this support without which the project not have been possible.

The author also gratefully acknowledges the help of the numerous $\tau\epsilon\chi\nu\eta$ PIs and their colleagues at the academic institutions noted in Appendix A. In addition, conversation and commiseration with other 'Rosslyn 11' evaluators was invaluable. Guidance from Ms. Adrian Tyler and Dr. Raymond McGhee at SRI is appreciated. Finally, the assistance of ECE work-study student Candace Pringle in the time-consuming process of compiling survey results is appreciated.

14. REFERENCES

[1] `cpathmonitor.org`. This site contains SRI overall project evaluation material as well as individual CPATH project evaluation information.

[2] Executive summary of nsf/cpath evaluators meeting, 2008.

[3] H. Taylor. Slide presentation at initial cpath evaluators meeting, July 2008.

APPENDIX

A. CONTACTS: PIS, COORDINATORS AND SCHOOLS

$\tau\epsilon\chi\nu\eta$ person	email	school
Mike Westall	`westall@cs.clemson.edu`	Clemson
Tim Davis	`tadavis@cs.clemson.edu`	Clemson
Andrew Duchowski	`andrewd@cs.clemson.edu`	Clemson
Robert Geist	`rmg@cs.clemson.edu`	Clemson
Pradip K Srimani	`srimani@cs.clemson.edu`	Clemson
James Wang	`jzwang@cs.clemson.edu`	Clemson
Mark Holliday	`holliday@email.wcu.edu`	WCU
William Kreahling	`wkreahling@email.wcu.edu`	WCU
Sridhar Narayan	`narayans@uncw.edu`	UNC-W
Andy Dalton	`adalton@email.wcu.edu`	WCU

Table 2: CPATH $\tau\epsilon\chi\nu\eta$ Course Coordinators and Schools (Fall 2008-Spring 2009).