Causal Reasoning:
Disseminating New Curricula with Online Courseware

Presented at the American Education Research Association, April 2003

Richard Scheines[1]

Dept. of Philosophy and Human-Computer Interaction
Carnegie Mellon University

1. Introduction

Online courses offer a huge opportunity for improving access and quality. Even though the digital divide is still significant (), public schools, public libraries and community centers have now made access to the world wide web fairly ubiquitous. Unlike textbooks and mail-order video courses, online courses can be made to be *interactive* on every level, incorporating virtual discussion rooms, multimedia, virtual labs that support open-ended exploration, and virtual tutors that can react intelligently to student input. Cognitive and computer scientists have shown that educational technology, properly designed and deployed, can significantly improve learning outcomes on the K-12 as well as the college level (Anderson, et al., 1995; Graesser et al., 2002; Koedinger, et al., 1997; Mostow, 2001; Ur & VanLehn, 1995). Especially in contexts in which high quality human educators are not available in sufficient numbers and are not going to become available anytime soon, the opportunity for improving access and learning outcomes is substantial.

The challenges to delivering on this promise are equally as large, however. Although producing and delivering an online course is now relatively inexpensive (Hayes, et al., 2000),[2] creating one that demonstrably improves learning and is widely disseminated is still very expensive. In all the cases I am aware of, educational technology that significantly and demonstrably improved learning outcomes was the result of many years

[1] This work is funded by the Andrew W. Mellon, the William and Flora Hewlett and the James S. McDonell foundations.

[2] The Educational Program for Gifted Youth at Stanford, http://www-epgy.stanford.edu/, claims to be able to produce an online course for under $50,000.

of an intensive cycle involving design → implementation → empirical research → re-design and re-implementation → further empirical research, etc. Although efforts are underway to create information technology infrastructure that will make this cycle more efficient,[3] it will undoubtedly remain expensive for some time to come.

If educational technology that does improve learning were widely disseminated, however, the high cost of creating it could be easily amortized. For example, it would cost approximately $5/student to produce a demonstrably effective six week mini-course in Introductory Economics on the theory of free markets, if we allow $1,000,000 for development, ongoing technical maintenance, and delivery, if we assume that 4% of the at least one million college students who take Introductory Economics each year were exposed to a such an online course, and we also assume that the course was used for only five years and then discarded. Used by only four universities in classes of approximately 100 for the same period however, the cost is $500/student, or $50,000 per course. In this light, the real challenge to cost-effectively improving access and learning with online education is dissemination.

In the last four years, several colleagues and I have developed an online course in causal and statistical reasoning[4] that includes approximately 18 textbook chapter length modules, 100 short case studies, and a Causality Lab. The material has been used by almost 2,000 students at 21 different colleges or universities in 37 different separate courses. In this paper I discuss several issues that we have confronted in disseminating Causal and Statistical Reasoning (CSR). We have had to accommodate a variety of different institutions, the material we are disseminating is entirely new and does not always fit neatly into typical university curricula, we are trying to simultaneously serve introductory and advanced audiences, and we are trying to provide some clients with a whole course and others with only part of a course. We are also trying to create a sustainable community of use in which the involved faculty wants to customize and in some cases extend the material, but are prevented from doing so by a software environment that is state-of-the-art, but far from user friendly. I begin by providing some background on CSR and then deal with these issues in turn.

---

[3] See, for example, Koedinger, Aleven, and Hefferman (2003),

[4] See www.phil.cmu.edu/projects/csr, or http://www.cmu.edu/oli/.

2. Causal and Statistical Reasoning Online

Every day the media brings reports of medical discoveries, scientific developments, and the results of surveys and experiments about causal questions. These claims concern what to eat, how much to exercise, how to help sustain the environment, where to put our tax dollars, whether capital punishment is a deterrent to crime, and so on. In order to make rational decisions about matters of social policy, we must be able to assess critically--even if informally--the causal and statistical reasoning used in these reports. Our curriculum aims to provide the knowledge and skill to do just this. It aims to illuminate, at least on a qualitative level, the scientific reasoning that underlies the "studies" that shape our social policies.

In a tradition begun by Sewall Wright (1934) and refined by Herb Simon (1953,54), Hubert Blalock (1961), and David Heise (1975) in the 1950s, 60s, and 70s, the real intellectual sources of the curriculum are in converging work in the last twenty years on "graphical" models done by statisticians, computer scientists, and philosophers. Beginning with statistical work around 1980, N. Wermuth and S. Lauritzen, (1983) along with Kiiveri and. Speed (1982, 1984) connected Markov random fields to diagrammatic representations of causal hypotheses. Speed and his students connected the directed graph formalism with linear causal models in the social sciences, epidemiology and econometrics. Working separately in computer science, J. Pearl's, *Probabilistic Reasoning in Intelligent Systems* (1988) developed fast procedures for calculating the probabilistic implications of graphical models introduced by the statisticians, provided algorithms for deciding indistinguishability (by observation) of graphs, and described search algorithms for special cases. Pearl also connected the formalism with models of categorical data, called "Bayes nets" in the computer science literature. P. Spirtes, C. Glymour, and R. Scheines, in *Causation, Prediction and Search,* (1993, 2000) axiomatized the connection between graphs explicitly interpreted to be causal and probabilistic independence, introduced calculations for interventions in causal systems, related such calculations to the "Rubin framework" for prediction (1974), well known in statistics, and extended search algorithms to a much larger class of alternative models involving latent variables. Pearl's *Causality*, (2000), reviews these developments and connects the framework with work on causation in philosophical logic. Applications abound,[5] several high quality conferences feature it regularly,[6] and in short the theory of statistical causal modeling has recently moved from adolescence into early adulthood.

[5] See, for example, Scheines (2000).

Although the theory is deep and often heavily technical, its rudiments are quite simple and can be developed with no more than arrow diagrams and high-school algebra. Because of this, and the disturbing fact that the subject was completely absent in high school or college curricula in the late 1990s, we decided we should develop an accessible curricula and begin the process of disseminating it. Convinced it was a subject that cried out for heavily "interactive" and "hands-on" learning – and that our best bet for achieving critical mass lay in a continuously updatable medium (not textbooks), we decided to develop an "online" course. With generous funding from the Fund for the Improvement of Post-Secondary Education and the Andrew W. Mellon Foundation, a team involving philosophers, education researchers, and programmers from Carnegie Mellon,[7] the University of California, San Diego,[8] and the University of Pittsburgh[9] created 18 highly interactive chapter length "concept modules," a repository of over 100 short "case studies," and a "Causality Lab" that allows students to setup and carry out simulated social science experiments, construct hypotheses to explain the data collected, and attempt to discover the hidden, "true" causal model.

Students using our material typically work through several of the concept modules as part of a whole course. The modules each cover between four and eight sizable concepts, for example, "random assignment of treatment in a medical trial." The concepts are each presented declaratively (usually text and graphics), and then illustrated and reinforced with an interactive simulation, Causality Lab exercise, or case study. Each page or two, the student is asked to do a short comprehension check to make sure they are getting the ideas, and at the end of each module there is a graded quiz that typically takes students 15 to 20 minutes to complete.

Beginning in the spring of 2000 and continuing through the following year, the material was taught to almost 400 students at UCSD and the University of Pittsburgh in five different courses. In each course we conducted a systematic experiment in which we compared lecture vs. online delivery of identical material. In each case students met with instructors weekly, but in the online condition they confined contact to discussion sections over case studies and questions. Over the five experiments, which involved four different lecturers six different TAs and two different universities, the online condition did no worse and typically better than the lecture condition (Scheines, Leinhardt, Smith,

---

[6] For example, Uncertainty in Artificial Intelligence, Artificial Intelligence and Statistics, and NIPs.
[7] Richard Scheines, Clark Glymour, Joel Smith, Joe Ramsey, Juan Casares, and Peter Spirtes.
[8] Sandra Mitchell, David Danks, Mara Harrell, and Willie Wheeler.
[9] Gaea Leinhardt, Kwang-su Cho, Rob Goldberg, Dan Steel and Francis Longworth.

and Cho, 2002). Final exam scores were approximately one half a standard-deviation better for online students, even controlling for pre-test, which was significant at approximately .07. Although evaluation and iterative improvement are continuous and ongoing, we began in academic year 2001-2002 to turn our attention more to dissemination, and it is there I now turn.

3. Early Dissemination Results

In the spring of 2001, we recruited college professors from schools not involved in developing or experimenting on the course[10] to use and test our material:

Table 1: Non-Development Schools 2001-2002

School	Department	Students
Kansas State University	Philosophy	80
Ohio State University	Philosophy	100
University of Chicago	Business	11
Univ. of Washington, Seattle	Statistics	23
Wichita State	Philosophy	15
William Jewell College	Philosophy	20

In fact each of the professors involved had gotten a PhD in philosophy, and was a personal contact with someone in the main development effort. As with UCSD and Pitt, each used the material as a whole course devoted either to scientific or statistical methodology. Except for William Jewell, each location is a large research oriented university. This dissemination experiment brought two dimensions to the surface that matter: getting the course approved as a way to satisfy a distribution requirement, and the level of the student taking the course.

At KSU, OSU, and William Jewell, the course was approved to satisfy a quantitative skills distribution requirement. In each of these three cases, the political battle to get the course accepted by the college or university as a way to satisfy a general requirement was a bruising, difficult, and protracted struggle. In two cases I was asked to intervene

[10] UCSD, CMU, and Pitt.

because turf battles had become intractable. One issue that fueled this debate was curricular. Psychologists, statisticians and mathematicians, who typically guard the "quantitative skills" gate, were loathe to approve a course that involved material not from the standard canon.

The quantitative skills gatekeepers were also hesitant to allow an "online" course the same status as one taught by traditional methods, even though evidence had been produced showing learning outcomes were, if anything, better for the novel delivery method, and even though, the idea of online instruction as we produce it is to free up more time for instructors to spend with small groups of students and less time with lecture.

The major dimension along which the success of our first dissemination experiment varied was the level of the student taking the course, however. In the University of Chicago, the material was offered as an elective graduate course to a small group. In Wichita State the material was offered as an honors seminar to upper level undergraduates. In William Jewell, the course was offered to freshman and sophomores who were worried about being excessively challenged in another quantitative skills course like discrete math. In Ohio State and KSU, upper level undergraduates took the course to satisfy the quantitative skills requirement. The experience was much as you might expect. The students and instructors in Chicago and Wichita were quite positive, those in Ohio State and KSU still positive but less so, and the students at William Jewell unhappy. Were it not for an enthusiastic instructor at William Jewell, the course undoubtedly would have been a failure.

Since our intent was primarily, as we said above, to provide widely accessible material that would, at least on a qualitative level, teach the scientific reasoning that underlies the "studies" that shape our social policies, we could not simply restrict our future dissemination efforts to upper level undergraduate/lower level graduate students at research universities. Although tempting, just making the material easier was not a real option. The first reason is that watering down the material would bore the more apt students, and the second reason is that it would force us to abandon two out of three of our curricular goals.

We intend our material to (eventually) satisfy each of the following three curricular needs:

- An introductory course on scientific method or critical thinking, with a focus on research methods in the behavioral and social sciences - typically the only course a student will take on "research methods"

- A pre or co-req to a traditional statistics sequence
- A foundation/introduction to the modern theory of statistical causal models (Bayes Networks)

In fact the material - as disseminated in 2001-2002, was still thin on statistical topics (typically difficult to teach), and was clearly best suited for the third niche - as it was used at the University of Washington, Seattle in a statistics elective - but still a little too elementary in the list of topics it actually covered. We were thus faced with three substantial challenges to wider dissemination:

1. How to make the material appropriate for both inexperienced and more advanced students, and
2. How to incorporate enough topical flexibility to cover each of the three curricular niches we had targeted, and
3. How to overcome faculty resistance to curricular novelty

4. Early Adjustments

Over the summer of 2002, we made two modifications to address the first two of these challenges. To address the first difficulty, our plan was to support students of disparate experience and ability by moving to more of a mastery as opposed to a standard assessment design. Each of the end-of-module quizzes in the first iteration of the course was a one-shot assessment with no learning feedback. Over the summer of 2002, we tripled the number of quiz questions and added extensive feedback to each question, giving the students three attempts at a quiz instead of one. We also built software that would construct each quiz attempt randomly from items that had not yet been seen by a particular student. The idea was to allow upper level students to quickly process the material and show mastery immediately, but to also allow lower level students the opportunity to take an assessment, learn from its feedback, go to class and learn more, and then get two more attempts at mastery before frustration would set in.

To address the second difficulty, we 1) added several optional case studies to the early modules, 2) followed each with essay questions targeted at the first curricular purpose, and 3) developed recitation lesson plans around still more case-study analysis for the first five modules, and 4) began development on additional modules targeted at purposes two and three.

5. Current Dissemination

In the spring of 2002 we advertised nationally for a summer workshop on using our material - complete with stipends for participation funded by the Mellon foundation. After receiving almost 70 applications, we selected twelve participants representing a wholly different range of institutions and departmental affiliations:

Table 2: Non-development Schools: 2002-2003

School	Department	Students
Allegheny College	Psychology	24
Bethel College	Physician Assistant	10
Chaminade University	Nursing	20
Dominican University	Nursing	23
Duquesne University	Speech Pathology	23
Marymount	Criminal Justice and Sociology	15
Regis College	Psychology	20
Rocky Mountain College	Math	5
University College of the Cariboo	Philosophy	19
Ursinus	Political Science	25
Washington & Jefferson	Math	21
Regis College	Psychology	20

The modification towards a mastery design - allowing students to attempt the end-of-module quiz multiple times until they satisfy an instructor designated level of mastery was a great success. Students from smaller schools in less quantitative disciplines like Nursing or Speech Pathology were almost 10 percentage points below the previous year's mean on the first five modules, but improved an average of almost 18% between first and the second attempt, and almost 10% between the second and third attempts. In over 80% of the modules, students were able to achieve their instructors designated level of mastery, compared with approximately 50% the prior year. Although data is not yet available for analysis this academic year, our informal contact with instructors reveals much less differentiation in student and instructor satisfaction as a function of the level of the student or the course than was the case the prior year.

What did emerge from this round of dissemination were two crucial challenges for the future, and both involve customization. First, as opposed to 2001-2002, in 2002-2003 many of our adopters chose not to offer an entire course based on our material, but rather to include some subset of our material as a unit in an already extant course. Second, several professors wanted to customize material to their own tastes, usually with respect to the cases analyzed.

Below is a histogram that depicts the wide range of demand by number of modules used for all the contexts of use since 2000.

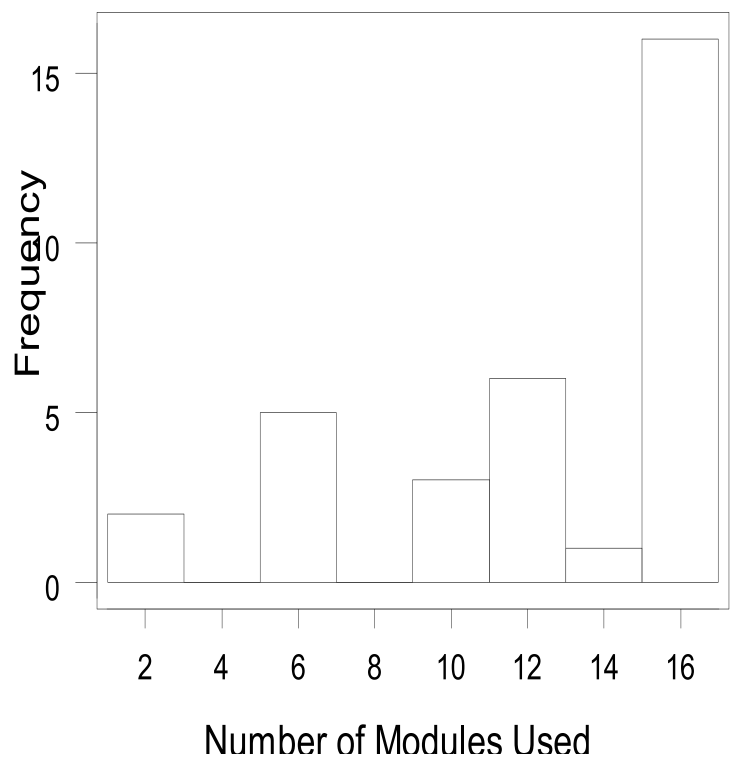

Figure 1: Distribution of Number of Modules Used

Approximately half of users chose to teach a whole course. For the other half, the challenge is to create subsets of material that is self-contained. If, however, an instructor wants to teach a module on "Experiments," which depends on material in eight other modules as pre-requisites, for example, then this need is exceedingly difficult to meet. Several solutions are possible. In one approach, favored by David Yaron in Chemistry and John Miller in Economics, both members of the OLI group at Carnegie Mellon, online material ought to focus on interactive homework or laboratory work, and not on presenting declarative material per se. Their strategy is to leave professors to their own idiosyncrasies for choosing declarative content - but support them with highly interactive homework or lab software that is customizable to a wide variety of uses. This is an excellent strategy, but one that can only work for already extant and mature curricula like introductory economics and a first year course in college chemistry. Where new material is being introduced, such a strategy is much more difficult to implement. Our Causality Lab is capable of being run as a supplementary "lab" to a course on research methods, but cannot be understood or used effectively with zero exposure to our material.

The approach I favor is to allow finer grained concept level assembly of material. If an instructor teaching a course in Psychological Research Methods wants to introduce students to the qualitative ideas about why randomized assignment eliminates the problem of unmeasured confounders in establishing a causal relationship, he or she ought to be able to do so by isolating the minimal set of concepts required to do so and by assembling these concepts into a mini-course of their own design. At the Open Learning Initiative,[11] we are taking precisely this strategy. A similar approach is already mature in the ActiveMath project.[12] By semantically tagging our content with concept level information that includes pre-requisite relationships, and by allowing users to construct their own syllabus by assembling concepts, choosing examples of these concepts, interactive activities for them, assessments for them, etc., and by using the software to enforce pre-requisites that need to be included, instructors will be able to much more flexibly tune our content to their own needs with respect to using only part of our material as part of their own course.

A second challenge that emerged for wider dissemination also involved customization. As you can see from Table 2, the disciplines involved in 2002-2003 range from Physician Assistant programs to Math to Psychology to Political Science. Instructors were eager to make the material relevant by including examples from their own discipline. Although they were of course free to do so outside the confines of the online material - they were much more interested in actually producing examples that we could add to the online case repository, and exercises we could add to the Causality Lab that were of their own design.

[11] See http://www.cmu.edu/oli/.
[12] See http://www.activemath.org/

These points reinforce both suggestions from Gaea Leinhardt and the work now being pursued by the Connexions project at Rice University (Baraniuk, et al., 2002). Leinhardt suggests that we approach dissemination by building a community of users who could then become disseminators, etc. To do so, she suggests we provide ways in which uses could become invested in the project by becoming producers as well as consumers.

The Connexions project at Rice[13] is focused on building the information technology infrastructure to do just this: making online course authoring a collaborative, community-based activity as opposed to a loner sport. By making it easy to re-purpose extant content, that is, to modify and customize it for your own purpose, to build upon rather than reinvent the wheel yet again, Connexions envisions supporting the sort of collaboration that Leinhardt argues will produce effective dissemination. Our experience in disseminating online courseware on causal and statistical reasoning support both views entirely. In order for instructors to use *new curricular* material, the material must support a wide range of use and be modifiable. Dissemination requires buy-in from an intellectual community, and the best way to achieve buy-in is through facilitating genuine input from the whole community.

6. Conclusions

There are a few challenges to dissemination that we thought we might face but that we did not. For one, we thought the success of the course would depend partly on the size of the class. In Figure 2 we show the distribution of class size, which has a large mode at around 20 because many classes are purposely capped at either 20 or 25. It turned out that class size had no discernible relation to the course's success or the student's pre post-test performance. As long as large classes provide sufficient TA support to break into discussion groups of around 20, the students get the human tutoring they need.

[13] http://cnx.rice.edu/

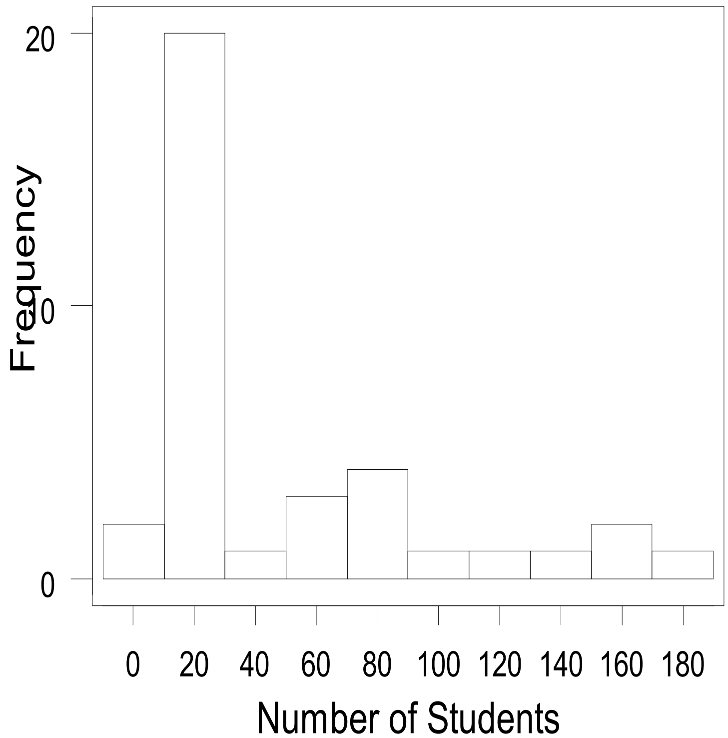

Figure 2: Distribution of Course Size

Another suspicion was that, since different universities provide different levels of technical support and have different levels of technical sophistication generally, that an online course would play better in more technical universities. Seemingly this is not so. If anything, we have had more technical difficulties at large, technically proficient universities than we have had at small ones, mostly because large universities are less nimble about fixing things and less concerned about servicing individuals than are small ones. Online courses that depend on relatively standard features of web browsers, as ours do, are relatively trivial to support at the client site. Although they are difficult to support at the host site, they are uniformly difficult.

Besides the usual challenges to dissemination, e.g., appropriate technical support, high quality material, etc., the challenges that have emerged for us in two years of delivering a novel curriculum on causal and statistical reasoning to over 20 separate institutions of higher education can be enumerated as follows:

1. The same material must be both accessible to inexperienced students as well as challenging to upper level students.

2. The online material must be usable in chunks that range in size from a week to a full semester length course (or more).
3. It must be possible to reconfigure the sequencing of material in order to support more than one curricular function.
4. The material itself must be easily customizable and extendable.

By providing repeated mastery opportunities and lots of "voluntary" content, we have successfully addressed the first challenge: making material accessible to different levels of students. The other challenges are more demanding, and will not, I believe, be met until information technology infrastructure exists that supports 1) fine grain concept level assembly of material by the course instructor, 2) assembly that automatically enforces pre-requisite relationships, and 3) collaborative authoring and use by a wide intellectual community.

In combination with the Connexions project and the ActiveMath project, the Open Learning Initiative at Carnegie Mellon is attempting to provide just such an infrastructure.

7. References

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences*, 4 (2), 167-207.

Baraniuk, R.G., Burrus, C., Hendricks, B., Henry, G., Hero III, A., Johnson, D., Jones, D., Nowak, R., Odegard, J., Potter, L., Reedstrom, R., Schniter, P., Selesnick, I., Williams,D., Wilson, W. (2002). Connexions: Education for a Networked World, in *IEEE International Conference on Acoustics, Speech and Signal Processing – ICASSP'02*, Orlando, May.

Blalock, H. (1961). *Causal Inferences in Nonexperimental Research*. University of North Carolina Press, Chapel Hill, NC.

Clark, R., and Mayer, R. (2003). *E-learning and the Science of Instruction*. Jossey-Bass, San Francisco, CA.

Corbett, A.T., Koedinger, K.R. and Hadley, W. S. (2002). Cognitive Tutors: From the research classroom to all classrooms. In P. Goodman (Ed.) *Technology enhanced learning: Opportunities for change*. Mahway, NJ: Lawrence Erlbaum Associates.

Erlbaum, B., McIntyre, C., and Smith , A. (2002). *Essential Elements: Prepare, Design, and Teach Your Online Course*. Atwood Publishing, Madison, WI.

Graesser, A.C., Person, N., Harter, D., & TRG (2002). Teaching tactics and dialog in AutoTutor. *International Journal of Artificial Intelligence in Education*, 12, 257-279.

Hayes, H.M., Jackson, J., and Anderson, D.V., (2000). Producing Effective Internet Courses with inFusion, *Proceedings Learning '00*, Madrid, Spain, October.

Heise, D. (1975). *Causal Analysis*. Wiley, New York.

Kiiveri, H. and Speed, T. (1982). Structural analysis of multivariate data: A review. *Sociological Methodology*, Leinhardt, S. (ed.). Jossey-Bass, San Francisco.

Kiiveri, H., Speed, T., and Carlin, J. (1984). Recursive causal models. *Journal of the Australian Mathematical Society* 36, 30-52.

Koedinger, K. R., Aleven, V., & Heffernan, N. (2003). Toward a rapid development environment for Cognitive Tutors. Accepted for the *Artificial Intelligence in Education conference.*

Koedinger, K. R., Anderson, J.R., Hadley, W.H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education, 8*, 30-43.

Mostow, J., & Aist, G., (2001). Evaluating tutors that listen: An overview of Project LISTEN. In K. Forbus & P. Feltovich (Eds.), *Smart Machines in Education* (pp. 169-234). Menlo Park, CA: MIT/AAAI Press.

McNaught, C., Kenny, J., Kennedy, P., & Lord, R. (1999). Developing and evaluating a university-wide online distributed learning system: The experience at RMIT University. E*ducational Technology and Society*, 2(4), 70-81.

O'Connor, C.J. (2002). Interactive, Multimedia Online Courses: Meeting Administrative, Development, and Dissemination Challenges Technology Based Learning and Research. Available at: www.aace.org/conf/site/pt3/paper_3008_812.pdf

Palloff, R.M., & Pratt, K. (1999). *Building Learning Communities in Cyberspace: Effective Strategies for the Online Classroom*. Jossey-Bass.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems***.** Morgan and Kaufman, San Mateo.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66, 688-701.

Scheines, R (2000). Effects of lead on children's IQ, *Handbook of Data Mining*, Pat Hayes, editor, Oxford University Press.

Scheines, R., Leinhardt, G., Smith, J., and Cho, K. (2002) "Teaching and Learning with Online Courses.". *Technical Report No. CMU_PHIL-135,* Dept. of Philosophy, Carnegie Mellon University, Pittsburgh, PA, 15213.

Simon, H. (1953). Causal ordering and identifiability. *Studies in Econometric Methods*. Hood and Koopmans (eds). 49-74.Wiley, NY.

Simon, H. (1954).Spurious correlation: a causal interpretation. *JASA*. 49, 467-479.

Spirtes, P., Glymour, C., Scheines R., (2000). *Causation, Prediction and Search*, 2nd Edition, MIT Press, Cambridge, MA.

Ur, S., & VanLehn, K. (1995). STEPS: A simulated, tutorable physics student. *Journal of Artificial Intelligence and Education*, 6(4), 405-437

Wermuth, N. and Lauritzen, S. (1983). Graphical and recursive models for contingency tables. *Biometrika* 72, 537-552.

White, K.W., & Weight, B.H. (Eds.) (1999). The Online Teaching Guide: A Handbook of Attitudes, Strategies and Techniques for the Virtual Classroom. Allyn & Bacon.

Wright, S. (1934). The method of path coefficients. *Ann. Math. Stat.* 5, 161-215