

## **AC 2007-1793: A CLOSER LOOK AT THE VANTH/ERC BIOMEDICAL ENGINEERING MODULES: A METHOD TO EXAMINE THE EFFECTS OF SELECTED CORE COMPONENTS OF THE “LEGACY CYCLE” AND HPL MODEL**

### **David Cordray, Vanderbilt University**

David S. Cordray, PhD is Professor of Psychology and Public Policy at Vanderbilt University. He is currently the Thrust Leader in Assessment and Evaluation for the VaNTH ERC. Professor Cordray has written extensively on research and evaluation methodology in education and human services areas. He has conducted experimental, quasi-experimental and meta-analytic assessments of intervention effectiveness in education, health, welfare, and other human service areas.

### **Alene Harris, Vanderbilt**

Alene H. Harris, Ph.D. is a Research Asst. Professor of Education at Vanderbilt University and serves as the Director of Education Programs of the VaNTH ERC. Her research interest is in effective instruction, and she provides workshops in designing and delivering HPL-influenced Legacy Cycle lessons.

### **Thomas Harris, Vanderbilt University**

Thomas R. Harris, MD, PhD is the Orrin Henry Ingram Distinguished Professor of Engineering and Professor of Biomedical Engineering, Chemical Engineering and Medicine at Vanderbilt University. He is currently Chair of the Department of Biomedical Engineering. His current interests focus on the development of learning sciences and learning technology for bioengineering. He is currently the director of the National Science Foundation Engineering Research Center in Bioengineering Technologies.

# A Closer Look at the VaNTH/ERC Biomedical Engineering Modules: A Method to Examine of the Effects of Selected Core Components of the “Legacy Cycle” and HPL Model

## Abstract

Educational innovations developed within the Vanderbilt-Northwestern-Texas-Harvard/MIT Engineering Research Center (VaNTH/ERC) for Bioengineering Educational Technologies have been tested using a sequence of experimental, quasi-experimental, and non-experimental methods. More specifically, many of the innovations were first tested using randomized experiments or high quality quasi-experiments. When modules were shown to be effective under experimental conditions, subsequent assessments relied on non-experimental methods (e.g., pre-to-post gain) or simply the achieved level of average performance.

Our overall all (i.e., meta-analytic) assessment of experimental and quasi-experimental studies have shown that they are effective in enhancing the learning outcomes for college and high school students. This paper demonstrates a set of methods that are used to probe the effects of multiple, sequential assessments of the same innovation, offering a broad scale model for monitoring the educational accomplishments of programs. This approach attempts to capitalize on available experimental and non-experimental as a means of testing effectiveness and monitoring subsequent performance. We also present analyses that show the unique contribution (to learning and other outcomes) of variations in specific core components of the Legacy Cycle and its learning principles.

## I. Introduction

The Vanderbilt-Northwestern-Texas-Harvard/MIT Engineering Research Center (VaNTH/ERC) for Bioengineering Educational Technologies has developed over 60 innovative instructional modules in a variety of bioengineering areas (See Harris et al.<sup>1</sup>). These innovations are based on the model of learning and instruction, (referred to as HPL for “How People Learn”) described by Bransford et al.<sup>2</sup> in a volume issued by the National Academy of Sciences entitled *How People Learn: Brain, Mind, Experience, and School*. Prior analyses have shown that a subset of all VaNTH-sponsored modules and course enhancements are effective. Using meta-analytic methods to quantitatively combine the results from multiple studies, we have shown that HPL-inspired innovations are effective for student learning (see Cordray, Pion, A.Harris, & Norris<sup>3</sup>). An updated meta-analysis, including 19 studies that yielded 28 effect size estimates, revealed an average effect size of 0.644 standard deviation unit. The most common ES is expressed as  $ES = (M^T - M^C) / SD_{pooled}$ , or the mean (M) difference between treatment ( $M^T$ ) and control ( $M^C$ ) conditions, divided by the pooled standard deviation (SD) across conditions. However, additional analyses suggest that the results for well executed experimental studies are somewhat smaller (0.441 standard deviation units) but they are still regarded as moderate effects (See Cordray, Harris & Gilbert<sup>4</sup>). Almost all of the studies that were synthesized used the “Legacy Cycle” as a means of operationalizing the HPL model (Schwartz et al.<sup>5</sup>).

Having established that the VaNTH modules and course enhancements are educationally beneficial by scientific standards, the purpose of this paper is to describe how we are examining how these educational benefits were attained. That is, we focus on methods of determining if aspects of the Legacy Cycle make a difference in learning and other outcomes.

## **II. The Legacy Cycle**

Every Legacy Cycle (Schwartz et al.<sup>5</sup>) begins with a challenge. Following this challenge, learners journey through five steps of the Legacy Cycle. These include: (1) generating preliminary ideas about the challenge, (2) obtaining information from experts representing multiple perspectives on the challenge, (3) conducting additional research and revising their preliminary ideas about the challenge, (4) testing their understanding of key principles by solving problems (receiving formative feedback), and (5) completion of a transfer task and sharing their results with their peers.

Our brief description of the steps within the Legacy Cycle represents an idealized version of the model. As VaNTH faculty implemented their modules/courses based on the Legacy Cycle, sometimes they deviated in various ways from this ideal. Because several modules were implemented, refined, and tested over time, these deviations in the way the Legacy Cycle was implemented provide a unique opportunity to assess the effects of components of the Legacy Cycle, and more generally, the features of the HPL framework. In examining the effects of isolated deviations (e.g., not providing formative feedback, no explicit review of preconceptions and misconceptions), we rely on standardized estimates of effects (Effect Sizes) derived from experimental, quasi-experimental and non-experimental (e.g., pre-post change) studies conducted within and across VaNTH partner institutions.

## **III. Methods of Assessment: An Expanded Base for Evaluation**

At the heart of our approach to assessing educational innovations is experimentation. Experimentation entails three features: (1) the active manipulation of conditions (i.e., the specification of an innovation and a counterfactual condition), (2) systematic assessments of the effects of these manipulations on learning and other outcomes, and (3) statistical or design-based control over extraneous factors that could bias the results.

Across 19 studies (with 28 distinct effects), about 40% of the studies that were conducted within VaNTH employed true experimental designs, with randomization (See Cordray, Harris and Gilbert<sup>4</sup>). The other (60%) comparative studies used quasi-experimental methods (with non-random assignment) to assess the effectiveness of modules and course augmentations.

Although an experimental approach to improving engineering education is demanding, it provides the only path to deriving unbiased estimates of the effects of learning on students. On the other hand, practical and ethical issues must be addressed. One particularly difficult practical constraint in using true experiments in assessing

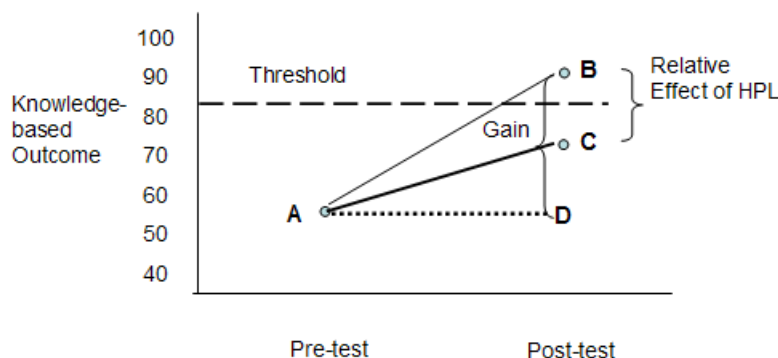
VaNTH materials entailed small class size, making the construction of a counterfactual condition impossible. When randomization was not feasible, however, we endeavored to help faculty invent design strategies that matched the circumstances and their willingness to undertake the research. One particularly effective strategy was to build a comparative design by adding conditions (control groups, modified treatments) incrementally. We refer to this as an emergent quasi-experiment (see Cordray, Pion, A. Harris, & Norris<sup>3</sup>). Although these are more practical designs, their inferential clarity is not as great as that of the randomized case (Cordray, Harris and Gilbert<sup>4</sup>). In this paper, we argue that these emergent designs can provide the basis for further assessments that are relative rigorous.

On the ethical side, questions about how much experimentation was needed arose. For example, at what point is an innovation sufficiently well tested to warrant adopting those practices and dropping any further experimentation? Once an approach is deemed effective, it could be argued that it is unethical not to adopt those practices and it is surely unethical to deny individuals access to these improved methods of instruction just for the purpose of constructing a control condition. These issues that were faced by VaNTH investigators over the past 7 years as they developed, tested and adopted (disseminated) their innovative approaches. Over time, we developed a continuous improvement model of assessment that attempts to accommodate these practical and ethical concerns.

As noted above, about 60 modules have been developed across the VaNTH institutions and partners. To date, only a fraction of the innovations were actually tested using experimental methods. In practice, VaNTH investigators used additional methods to test, at least partially, the effects of their innovations. As such, the amount of evidence available about VaNTH innovations is greater than depicted in our prior analysis of experiments and quasi-experiments.

A common form of assessment that has been increasingly utilized is a pre-post gain analysis. The differences and similarities between the experimental and gain-based analyses are depicted in Figure 1. From Figure 1 we can see that three evaluative criteria can be used to assess the effects of innovations: (1) relative effects; (2) differential gain; and (3) achievement in the innovation group that exceeds a pre-specified threshold for success.

Figure 1. Three Evaluative Criteria



These three criteria are embedded in the basic experimental design and can be seen in Figure 1. Because of randomization to conditions, prior to the delivery of the innovation (pre-test), the average performance across groups is expected to be equivalent (see A in Figure 1). Moreover, the gain observed between pretest and post-test for the innovation group ( $\text{Gain}^{\text{Tx}} = B-A$  in Figure 1) can be judged against what would have happened had the innovation not been used – that is, the performance of the participants in the control condition ( $\text{Gain}^{\text{C}} = C-A$  in Figure 1). In the logic of the experiment, the relative effect of the innovation is the difference between the performance of each group at the post-test (Relative Effect =  $B-C$ ). The RE is related to the effect size (ES) in the sense that the latter is expressed in standard deviation units. The RE is considered an unbiased effect when groups are composed at random and the result of randomization (group equivalence) is maintained through the collection of the post-test results. When non-randomized designs are used, there is usually an initial difference between the groups at the pretest (A). As such, the post-test differences can be biased by this initial group non-equivalence.

The relative effect of the intervention versus the control can also be thought of as differential gain from pre- to post-test, across the conditions. That is, differential gain is given as  $[(B-A) - (C-A)]$ . If an experimental assessment shows that the innovation group outperformed the control participants (e.g., their learning increased at a faster rate), we can use a standardized index of gain to determine if subsequent cohorts of participants achieve similar gain, without the need for a control condition (see below).

The third criteria, meeting or exceeding a pre-specified threshold, can be used to judge whether the innovation group performed well-enough to declare that participants mastered the material. Once it is demonstrated that innovations routinely meet or exceed this level of performance, the performance of subsequent cohorts of participants can simply be monitored to assure that performance remains at the threshold level. At this stage of the research it is not necessary to use control conditions to determine if the innovation is effective.

Within VaNTH, a common or standardized outcome could not be used to assess educational outcomes across modules or courses. Rather, deep understanding of basic bioengineering knowledge was commonly assessed through a combination of problem solving and performance on transfer tasks. These outcomes are usually scored with rubrics designed to assess the depth of knowledge and understanding through knowledge-based questions (KBQ). Because of the study-to-study variation in the manner with which outcomes were measured, it is necessary to use effect sizes as a common metric for assessing gain, relative effects and threshold attainment. Three effect size (ES) indicators can be used to index these criteria:

$$\begin{aligned} \text{ES}_{\text{Gain}} &= (M_{\text{post}} - M_{\text{pre}})_{\text{Trial}} / \text{SD}_{\text{pooled}} \\ \text{ES}_{\text{Differential Gain}} &= [(M_{\text{post}} - M_{\text{pre}}) / \text{SD}_{\text{pooled}}]_{\text{Trial}} - [(M_{\text{post}} - M_{\text{pre}}) / \text{SD}_{\text{pooled}}]_{\text{Control}} \\ \text{ES}_{\text{Threshold}} &= (M_{\text{Post}})_{\text{Trial}} - \text{Threshold} \end{aligned}$$

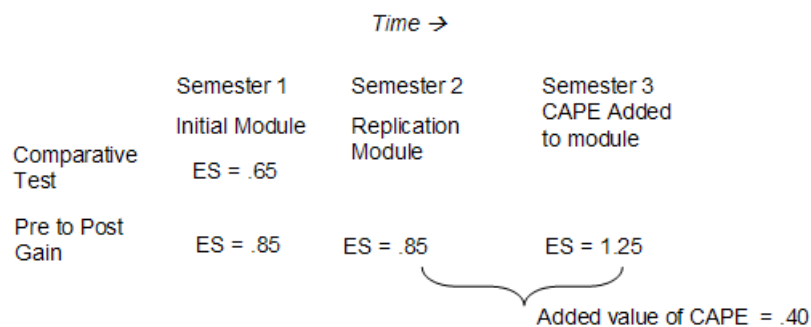
The use of these effect size indices is illustrated in the next section.

## IV. Illustrations

In this section we present a somewhat stylized example of the method of assessment described above. We also provide summaries of actual results from several prior assessments that illustrate aspects of the method.

Figure 2 provides a hypothetical illustration of how evaluations of modules have been commonly structured by VaNTH investigators. In this example, a module was assessed three times by the same investigator. Over these assessments the methods change from a comparative, experimental design (Semester 1) to a simple assessment of pre-post gain (Semesters 2 and 3). In addition, the module is modified during the third round of assessment (Semester 3). The effects sizes estimates show that in Semester 1, the relative effect of the model was 0.65 and the pre-to-post gain for the innovation group was 0.85. The difference between these estimates ( $0.20 = 0.85 - 0.65$ ) is the effect of rival explanations (e.g., testing, maturation) that are controlled by the counterfactual condition. In the second semester, the module is repeated. Rather than conducting another experiment, the investigator measures the pre-to-post gain for the innovation (only) group. In our illustration the gain for this second group is equal to the gain seen in the initial experiment ( $0.85 = 0.85$ ). In other words the replication was just as effective as the initial experimental group, which was more effective (0.65) than the traditional instruction in the counterfactual condition. For our example, the investigator added the Courseware Authoring and Packaging Environment (or CAPE, Howard<sup>6</sup>) as a means of delivering the educational material in Semester 3. Again, pre-to-post gain was assessed for the intervention group only. But, the effects of the addition of CAPE can be assessed by contrasting the gain observed under the two different versions of the module (CAPE and No-Cape in Semester 3 versus Semester 2, respectively). The example in Figure 2 shows that the addition of CAPE results in an increment in the effect size (ES) of 0.40. In meta-analytic terms, an ES of 0.40 is considered a moderate size effect. This sequence of assessments shows how evaluation designs can be reformulated to test replications and modifications without the need to use randomize experiments for each evaluation study.

**Figure 2. Assessing the Effects of Module Modifications: An Illustration**



## Some Real Examples

- (1) *The effects of teamwork, a critical component of the HPL model.* Barr et al.<sup>7,8,9</sup> tested a series of modules over three consecutive years. During the first two years, the participants in the innovation condition worked alone. In the third year, they worked in small teams. When asked about the extent to which the module enhanced their ability to work with others (an ABET criteria), data from participants in the Year 1 and 2 studies showed that there was no difference between the innovation and control groups (ES=0.06). Further, the pre-to-post gain in reported ability to work with others was negative (ES= - 0.20) for participants in the HPL condition. On the other hand, when participants in the Year 3 study were asked about their ability to work in teams, they showed a positive pre-to-post gain (ES=0.40). Here, there was no control condition but using the data from Years 1 and 2 to gauge the meaning of an ES of 0.40 provides a meaningful way of assessing the effects of teamwork. Replication of these findings from other studies should increase our confidence in the validity of the comparison.
- (2) *How adequate is the performance?* Martin et al.<sup>10</sup> used a high quality quasi-experiment to examine the effects of an HPL-inspired biotransport course on adaptive expertise. The experimental results revealed large effects for innovativeness (ES=1.32) and efficiency (ES=1.56). Closer inspection of the performance of students in the HPL-based course revealed that they achieved at a level equivalent to 80% of the total score on the rubric for innovation but only 38% on the rubric for efficiency. Although no threshold had been established for performance by Martin et al.<sup>10</sup>, these descriptive statistics are informative. That is, even though the experimental results suggest that the course was effective relative to performance in a traditional course, the modest level of mastery on the efficiency scale suggests a greater dose of the innovation might be needed to bring performance up to a level that could be regarded as mastery (e.g., 80%).
- (3) *How to optimize the “generate ideas” component of the Legacy Cycle.* Linsenmeier et al.<sup>11</sup> tested the effects of an HPL-inspired metabolism laboratory over three consecutive years using a randomized experiment (with randomization to conditions within years). The comparative results show that the lab is effective (ESs = 0.57 and 0.30 in Years 1 and 2, respectively) relative to a traditional lab format. Prior to Year 3, they enhanced an aspect of the innovative lab to encourage thinking about the material before the lab. This enhancement resulted in greater performance (ES=0.70,  $p < 0.01$ ) relative to that of the participants in the prior innovation conditions. Clearly, the new web-based material assisted students in understanding factors contributing to metabolic rate. Accuracy and completeness of their descriptions of these factors were also enhanced (ES=0.81) relative to their counterparts in prior versions of the metabolism laboratory.

## V. Discussion and Future Directions

The research and the methodology underlying this paper is ongoing and currently under development. We regard the results presented here as preliminary. Most of the questions that are to be address as part of our comprehensive assessment of the impact of VaNTH and its materials focus on differences between studies in certain elements of the Legacy Cycle. The questions pertaining to differences include, for example, How different are the effects when the identification of preconceptions and misconceptions are and are not formally assessed? How different are learning outcomes when feedback is or is not provided? How different are learning outcomes when solutions to challenges involve team work, rather than individual work? We also use the method to examine whether variations in the level of fidelity of implementation of the model affect learning outcomes. Such results, if they can be trusted and are replicable, provide evidence of the robustness of the Legacy Cycle approach to enhancing learning. In general, the approach we are using resembles a continuous improvement model of research and development (R&D). This model seems ideal for ongoing assessment of bioengineering programs, allowing managers to test and improve educational innovations by the use of systematic research procedures.

### Acknowledgements:

This work was supported primarily by the Engineering Research centers Program of the National Science Foundation under Award Number EEC-9876363.

## References

- <sup>1</sup>Harris, T.R., Bransford, J.D. & Brophy, S.P. (2002). Roles for learning sciences and learning technologies in biomedical engineering education: A review of recent advances. *Annual Review of Biomedical Engineering*, 4, 29-48.
- <sup>2</sup>Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). How people learn: Brain, mind, experience, and school. Washington, DC: National Academy Press.
- <sup>3</sup>Cordray, D.S., Pion, G.M., Harris, A. & Norris, P. (2003). The value of the VaNTH Engineering Center. *IEEE Engineering in Medicine and Biology Magazine*, May/June, pp. 47-54.
- <sup>4</sup>Cordray, D.S., Harris, T.R. & Gilbert, J. (2007). "What Works" in Engineering Education? A Meta-analysis of VaNTH/ERC Biomedical Engineering Modules. Paper available from [www.vanderbilt.vanth.org](http://www.vanderbilt.vanth.org).
- <sup>5</sup>Schwartz, D.L. Lin, X., Brophy, S., & Bransford, J.D. (1999). Toward the development of flexibly adaptive instructional designs. In C,M, Reigelut (Ed.), *Instructional design theories and models: Volume 11*. Hillsdale, NJ: Erlbaum.
- <sup>6</sup>Howard, L. (2003). Adaptive learning technologies for bioengineering education. *IEEE Engineering in Medicine and Biology Magazine*. May/June, pp. 58-65.



- <sup>7</sup>Barr, R., Pandy, M., Petrosino, A., Abraham, L., Karnade, T. & Patel, B. (2003). Classroom testing of virtual biomechanics laboratory (VBL) learning modules. Proceedings of the 2003 ASEE Gulf-Southwest Annual Conference, University of Texas at Arlington.
- <sup>8</sup>Barr, R., Pandy, M., Petrosino, A., Austin, B. & Goldberg, E. (2004). Classroom testing of VaNTH Biomechanics learning modules. Proceedings of the 2004 American Society for Engineering Education Annual Conference & Exposition.
- <sup>9</sup>Barr, R., Pandy, M., Petrosino, A. & Svihla, V. (2005). Challenge-based instruction: The VaNTH Biomechanics learning modules. 35<sup>th</sup> ASEE/IEEE Frontiers in Education Conference, October 19-22, 2005, Indianapolis, IN, USA
- <sup>10</sup>Martin, T., Rivale, S., & Diller, K.R. (2006). Comparison of student learning in challenge-based and traditional instruction in biomedical engineering. The University of Texas, Austin. Under review.
- <sup>11</sup>Linsenmeier, R., Kanter, D., Smith, D., Linsenmeier, K., & McKenna, A. (2006). Evaluation of challenge-based metabolism laboratory for undergraduates. Northwestern University, Evanston, Illinois. Under review.