2006-1790: VANTH OBSERVATION SYSTEM COMPONENT ASSESSMENT

Monica Cox, Purdue University

Monica Farmer Cox is an Assistant Professor of Engineering Education at Purdue University. She received her Ph.D. in Higher Education Administration at Peabody College of Vanderbilt University, her M.S. in Industrial Engineering at the University of Alabama, and her B.S. in Mathematics at Spelman College. Her research interests include teaching and learning in engineering education; engineering faculty and student development; and assessment and evaluation of engineering curricula, faculty pedagogy, student learning, student retention, and student engagement within engineering courses.

Alene Harris, Purdue University

VaNTH Observation System Component Assessment

Abstract- Since 1999, the VaNTH Observation System (VOS), a direct classroom observation system, has been used to collect data about classroom activities within bioengineering courses. Two components of the VOS, the Classroom Interaction Observation and the Global Ratings, specifically collect data about whether observed courses contain elements of the "How People Learn" (HPL) framework, as set forth in the National Research Council publication *How People Learn: Mind, Brain, Experience, and School*.[1] VOS observers use the Classroom Interaction Observation to collect information about the types of interactions that occur between faculty and students and among students within a course, and observers use the Global Ratings to evaluate summatively the elements of a course. Although several semesters of data have been collected at two of the universities, the validity of the VOS has not been assessed. To evaluate the validity of the VOS, five validity studies were conducted. Two content validity studies examined the extent to which eleven education content experts judged the elements of the HPL framework to be present within the Classroom Interaction Observation and Global Ratings components of the VOS, respectively. A convergent validity study noted the extent to which sampled Classroom Interaction Observation data collected in live classes correlated with full-class period Classroom Interaction Observation data collected in videotapes of those same classes. A second convergent validity study reported correlations between two different Classroom Interaction Observation assessment methods. Finally, a criterion validity study evaluated how well a newly-developed HPL Index classified Classroom Interaction Observation data within bioengineering courses that were designated as either traditional or nontraditional courses. This paper provides overviews of each validity study.

Introduction

Since 1999, the VaNTH Observation System (VOS), a direct classroom observation system, has been used to collect data about classroom activities within bioengineering courses.[2] Developed from the Stallings Observation System, which registers the presence and absence of over 600 in-class student and teacher behaviors and activities within K-12 classrooms,[3,4,5] the VOS data has been used to assess curricular changes that are based upon the "How People Learn" (HPL) framework within postsecondary engineering classrooms in the VaNTH Engineering Research Center (ERC) for Bioengineering Educational Technology.

In an effort to improve instruction and learning within bioengineering courses, faculty worked to implement effective classroom learning practices as demonstrated within the HPL framework. This framework is comprised of four dimensions that, when used together, enhance students' academic experiences and optimize learning.[1] The four dimensions represent activities that are *learner-centered* (i.e., students' prior experiences and misconceptions are factored into how course content is presented), *assessment-centered* (i.e., formative and summative assessment techniques are used to provide opportunities for students and faculty to receive feedback), *knowledge-centered* (i.e., lecture material is organized and presented so that students develop deep understanding of course concepts) and *community-centered* (e.g., students engage in collaborative learning within the classroom).

Trained classroom observers collect VOS classroom data using four components—(1) the Classroom Interaction Observation (CIO), (2) the Student Engagement Observation (SEO), (3) the Narrative Notes (NN), and (4) the Global Ratings (GR).[2] The CIO is a professor-focused portion of the VOS that uses code strings to capture student and faculty interactions and the presence of HPL dimensions within classrooms. Within each repeating three-minute CIO coding session, VOS observers record approximately thirty to forty-five code strings at the speed of speech. Immediately following each CIO, SEO data are collected. The SEO is a student-focused portion of the VOS that takes a thirty- to sixty- second "snapshot" of students to capture their engagement in five undesirable classroom behavior categories and six desirable classroom behavior categories. After the SEO, an observer uses the Narrative Notes portion of the VOS to input qualitative information about the lesson. Throughout the remainder of the observed session, the collection cycle of CIO, SEO, and NN data continues. At the end of the session, an observer uses the GR portion of the VOS to summatively evaluate an instructor's teaching and learning patterns. Each GR item represents either signaling with cognitive organizers, assessing students' understanding, or maintaining lesson engagement.

Although the VOS presents valuable information about engineering faculty's usage of nontraditional classroom instruction within the HPL framework, the validity of the VOS had not been examined. For this reason, the first author developed a five-part validity study to examine the validity of the Classroom Interaction Observation and Global Ratings portions of the VOS. This paper provides overviews of each validity study.

Validity Study Overviews

This paper describes the research methodology for each of the five validity studies conducted on the CIO and Global Rating portions of the VaNTH Observation System (VOS).[7] Studies 1 and 2 examine the content validity of the CIO and GR portions of the VOS, respectively. Study 3 examines the convergent validity of alternative indices of the amount of HPL-based pedagogy that is present within the classes in the sample. Study 4 examines the extent to which results converge when they are derived from alternative data gathering methods (i.e., sample of real-time coding vs. videotaped class sessions). Finally, Study 5 examines whether an index of "HPLness" discriminates between courses that are known to use HPL-based versus traditional pedagogy.

Validity Study Descriptions

Study 1- Content Validity of the Classroom Interaction Observation Portion of the VOS

Content validity examines "the extent to which a measurement reflects a certain intended domain of content."[8] The purpose of this study was to assess the extent to which eleven content experts familiar with the HPL framework agree with current classifications of the four dimensions of the HPL framework (*knowledge-centered* [K], *learner-centered* [L], *assessment-centered* [A], and *community-centered* [C]) within the CIO portion of the VOS. (The extent to which experts agreed with the current classification of classroom organization was also noted since organization is an item of interest in the CIO). Specific interest was taken in how VOS observers' classification of the presence of the four HPL dimensions (and organization) agreed

with those of selected content experts, in which classroom vignettes or types of classroom vignettes were easy or difficult for experts to agree upon, and in variances in agreement across different kinds of experts (e.g., professors, postdocs, and graduate students).

To assess the extent to which elements of the current CIO portion of the VOS accurately captured classroom organization and the four dimensions of the HPL framework as defined by HPL framework authors, a survey containing twenty vignettes was distributed to each expert. On this survey, each expert rated the extent to which the four HPL dimensions (and non-HPL category of *organization*) were present in the appropriate portion of twenty vignettes of actual bioengineering classrooms. To assure that the experts were using the same definitions for the HPL framework dimensions, they were given a summary sheet that provided brief definitions of the major elements of the HPL framework and examples of the four dimensions, as well as organization. After verbally discussing the given definitions and classroom examples with the first author, content experts rated three orienting vignettes that were similar to the twenty vignettes on the survey. They then rated the extent to which a highlighted portion of twenty one- to two-minute vignettes did contain or did not contain HPL dimensions and organization as defined on the HPL summary sheet.

Using a four-point Likert scale, content experts rated the extent to which organization, knowledge-centeredness, learner-centeredness, assessment-centeredness, and community-centeredness were present within the highlighted portion of each of the twenty vignettes. Figure 1 shows one of the actual vignettes given to content experts. Content expert agreement with the current VOS observer rating standards were calculated across the eleven content experts and across the twenty vignettes.[7]

8) (The professor is talking about pressures in the heart.)
PROF: I want to focus today in particular on what's happening with the pressure.
PROF: So that's this middle, yellow band (refers to projected graphic).
PROF: But of course it's very closely- in fact, inescapably linked to what's happening in the very top band, and that's the electrocardiogram.
PROF: But again, and down here, what goes on down here (gestures to lower part of image) is also closely related to the volume issues here (gestures to middle part of image).
PROF: I want to look here at the pressures.

To what extent are the following HPL elements present in the highlighted segment?

	Not at All	Only a Little	Some	A Great Deal
Knowledge-Centered	1	2	3	4
Learner-Centered	1	2	3	4
Assessment-Centered	1	2	3	4
Community-Centered	1	2	3	4
Organization	1	2	3	4

Comment(s)_____

Figure 1. Sample Assessment for the CIO Content Validity Study[7]

The percent agreement between HPL content experts' ratings and trained VOS observers' ratings for each vignette across vignettes and across experts were analyzed at three levels. The most liberal criterion for agreement counted three values (2= "a little," 3= "some," 4= "a great deal") as agreement. By regarding agreement as including ratings of 2, 3, or 4, the most liberal degree of agreement implies that *at least* a little of the dimension was present. Constraining the agreement to ratings of 3 or higher meant that at least some (3= "some" or 4= "a great deal") dimension was present. The most stringent degree of agreement only counted responses of 4 ("a great deal"). Naturally, as the inclusion criteria moved from liberal (2, 3 or 4) to stringent (only 4), the degree of agreement on dimensions declined.

Study 2- Content Validity of the Global Ratings Portion of the VOS

The purpose of this study was to assess the extent to which the same eleven content experts in the CIO content validity agreed that the seventeen Global Ratings items used within the VOS represented the four dimensions of the HPL framework. Because the Global Ratings indicators were designed to represent effective classroom pedagogical behaviors, it was hypothesized that the majority of the indicators reflected the principles of the HPL framework.

Five research questions were of interest within Study 2. First, to what extent did the Global Ratings (GR) portion of the VOS accurately capture the four dimensions of the HPL framework as defined by HPL framework authors and as represented by the percent agreement across experts and across indicators? Second, could GR indicators be grouped into subscales that represented HPL dimensions? Third, could GR subscales be used to create a Global Ratings HPL index to analyze current VOS data across semesters? Fourth, were certain GR indicators within this study problematic or difficult for experts to rate along the HPL dimensions and/or organization, and if so, why? Finally, were there differences in ratings of indicators across the eleven experts (e.g., professors, postdocs, graduate students, etc.), and if so, what were these differences?

To assess the extent to which the Global Ratings portion of the VOS accurately captured the four dimensions of the HPL framework as defined by HPL framework authors, HPL content experts completed a protocol similar to the one used in Study 1. Content experts rated the extent to which knowledge-, learner-, assessment-, and community-centeredness (along with organization) were represented within each of the seventeen Global Ratings indicators. Similar to the CIO content validity study, experts used a "1" to note whether HPL dimensions and organization are present "not at all," a "2" to note whether HPL dimensions and organization are present "only a little," a "3" to note whether HPL dimensions and organization are present "some," and "4" to note whether HPL dimensions and organization were present "a great deal." Space was provided for experts to write comments about each indicator and the applicability of the HPL dimensions (Figure 2).

(1.) **"The professor provides a chronological outline of the steps of the lesson."**

Given what you know about HPL, which label(s) best categorizes item #1?

	Not at All	Only a Little	Some	A Great Deal
Knowledge-Centered	1	2	3	4
Learner-Centered	1	2	3	4
Assessment-Centered	1	2	3	4
Community-Centered	1	2	3	4
Organization	1	2	3	4

Comment(s)_____

Figure 2. Example of a Global Indicator and Rating Scheme for the HPL Dimensions and Organization[7]

Study 3- Convergent Validity of the Classroom Interaction Observation Portion of the VOS

Convergent validity is an "overlap between alternative measures that are intended to tap the same construct but have different sources of irrelevant, undesired variation"[9] The purpose of this study was to examine the convergent validity of two assessment methods that were created to analyze CIO data via correlations between the methods. Both methods were used to analyze data collected within twenty-eight biomedical engineering-related courses (182 observations) during five semesters. The first CIO assessment method analyzed the percent of *individual* HPL dimensions (i.e., the amount of knowledge-, learner-, assessment-, and community-centeredness) that were present within each observed class session.[10] The sum of the observed instances across dimensions was greater than 100% because some dimensions were not mutually exclusive. The second method, a newly-created HPL Index, used entire CIO codes strings (all CIO categories) to analyze the percent of HPL-oriented instruction that occurs within each class session. (More information about the creation of the HPL Index will be presented in forthcoming papers.) The HPL-oriented instruction percent within the HPL Index represented the *interdependencies* of the CIO and represented a portion of an overall sum that equals 100%,. The primary research question for Study 3 examined the extent to which the HPL Index accurately reflected the prevalence of HPL-based pedagogy in a class or course as correlated to results obtained using the previous CIO assessment method.

Study 4- Convergent Validity of Alternative Data Gathering Methods for the Classroom Interaction Observation Portion of the VOS

The purpose of this study was to examine the convergent validity of two alternative gathering data schemes for the Classroom Interaction Observation portion of the VOS and to determine if the current CIO/VOS data collection method (cycles of time-sampled behaviors) accurately represented what occurs during an entire class. Twenty (20) class sessions were observed and recorded during the spring 2004 semester, and CIO data were collected within these sessions using both collection schemes. The first scheme gathered CIO data live using a time-sampled scheme, within the cycle of a three-minute Classroom Interaction Observation, followed by a thirty- to sixty-second Student Engagement Observation, followed by one- to two-minute Narrative Notes. The second scheme gathered CIO data via videotape continuously throughout the observation. The HPL Index was used to contrast the results from the two schemes (sample of real-time coding versus videotape), and it was hypothesized that there would be positive correspondences between CIO data coded using the current VOS scheme within "live" classrooms and CIO data coded using the alternate VOS scheme within videotaped versions of the same classrooms.

For the twenty classes within the sample, the HPL Index was used to calculate the total percent of class time spent in organizational activities, HPL-oriented instruction, and traditional instruction. Values for both the sample of real-time CIO data and for the videotaped CIO data were analyzed, and correlations and profiles for two coding schemes were found. These results showed whether CIO data collected using the current, sampled VOS data collection method accurately represented what occurred during an entire class, along with the extent to which an entire class session was not captured by time-sampled CIO data.

Study 5- Criterion Validity of a Newly-Developed "How People Learn" Index

Criterion validity is defined as "the degree to which an instrument's scores are related to external criteria believed to measure the attribute of interest"[11] This study examined the criterion validity of the HPL Index (see Study 3) derived from the CIO portion of the VOS. CIO data from twenty-eight courses (eighteen HPL-oriented courses and eleven traditionally taught courses) were analyzed within this study. The main question of interest within this study addressed whether the HPL Index derived in Study 3 was sensitive enough to capture HPL-related differences in courses that are known to employ HPL or traditional pedagogy. Being able to distinguish between two contrasting groups yielded evidence of criterion validity of the HPL Index. It was hypothesized that higher HPL instructional scores (on average) would be found within HPL courses and that lower HPL instructional scores (on average) would be found within traditional courses.

Data were grouped based upon the classification of courses as either HPL-oriented or traditional courses. The HPL Index (based on the Classroom Interaction Observation) was used to calculate the percent of HPL instruction, traditional instruction, and organization for all courses. Comparisons for seventeen HPL-oriented courses versus thirteen courses that entail traditional pedagogy were made.

Future Work

This five-part validity study advances knowledge in several ways. First, since the HPL Index developed within this study was found to have criterion validity, the HPL Index will be

used to analyze additional semesters of data within bioengineering courses at Vanderbilt University and Northwestern University and within first-year engineering courses at Purdue University. Data from the HPL Index will be also be used to supplement current data that is being collected about faculty and students. Second, data about the summative indicators of effective teaching will be used as a foundation for revising the current Global Ratings portion of the VOS. In the future, this revised instrument may be used as a student-based classroom assessment tool. Third, this study provides additional information about current ways of operationalizing the HPL framework within engineering courses. Through dissemination of these findings, engineering faculty may become more aware of their own teaching practices and of ways to improve their teaching strategies and effectiveness.

Bibliography

1. Bransford, J.D., Brown, A.L., & Cocking, R.R. (1999). How people learn: Brain, mind, experience, and school. Washington, DC: National Academy Press.

2. Harris, A.H. and Cox, M.F. (2003). Developing an Observation System to Capture Instructional Differences in Engineering Classrooms. *Journal of Engineering Education* 92: 329-336.

3. Adolf, J.W. (1983). An examination of Stallings' implementation and child effects of teaching practices in Follow Through classrooms. ERIC Ascension Number ED234907.

4. Stallings, J. (1977). Learning to look: A handbook on classroom observation and teaching models. Wadsworth, Belmont, CA.

5.---------- (1978). The development of the contextual observation system. Paper presented at the American Education Research Association Conference. Toronto: Canada.

6. ----------(1980). Allocated academic learning time revisited, or beyond time on task. Educational Researcher, 40, 11-16.

7. Cox, M.F. (2005). An Examination of the Validity of the VaNTH Observation System (VOS). Unpublished doctoral dissertation. Nashville, TN: Vanderbilt University.

8. Carmines, E.G., & Zeller, R.A. (1991). Reliability and validity assessment. Newbury Park: Sage Publications.

9. Judd, C.M., Smith, E.R., & Kidder, L.H. (1991). Research methods in social relations. (Sixth Edition). Fort Worth, TX: Harcourt Brace Javanovich College Publishers.

10. Norris, P.R., Harris, A.H., and Washington, C. (2004). The VOS data manager: Providing immediate feedback on teaching effectiveness. Proceedings of the American Society for Engineering Education Session 1793: 8 pages.

11. Doherty, R.W., Hillberg, R.S., Epaloose, G., & Tharp, R.G. (2002). Standards performance continuum: Development and validation of measure of effective pedagogy. The Journal of Educational Research, 96, 2, 78-91.

Glossary

HPL- How People Learn
CIO- Classroom Interaction Observation
SEO- Student Engagement Observation
NN- Narrative Notes
GR- Global Ratings