Classroom Artifacts: Tools to Assess the Use of Active, Innovative, and Engaging Pedagogies Among Engineering Faculty

Michael R. Tomlinson and Norman Fortenberry
Center for the Advancement of Scholarship on Engineering Education
National Academy of Engineering
m.tomlinson@alumni.ncsu.edu, nfortenb@nae.edu

Abstract – This paper examines artifacts of the classroom as a method of assessing whether or not engineering faculty are using best-practice teaching methods in class. Past research has achieved success in using artifacts to give similar results as direct observation. Artifacts assessment alone or in combination with surveys would be more accurate and less biased than surveys alone. Success of past portfolio assessment and the rise of college e-portfolio use among faculty make it highly likely that e-portfolios could provide the artifact information necessary for large scale assessment. Additionally, e-portfolios could enhance learning and teaching thereby increasing the willingness of faculty and students to participate nationally.

INTRODUCTION
Previous work done at The Center for Advancement of Scholarship on Engineering Education (CASEE) reviewed several methods that exist to assess instructional methods in the classroom [1]. This review pointed to three assessment options: (1) Self Assessment Survey, (2) External Direct Assessment, and (3) External Indirect Assessment.

This paper looks more extensively at indirect assessment, specifically the method of artifact analysis. In this method, researchers attempt to reassemble the teaching behavior by post-analysis of material (e.g. homework, exams, syllabi, teacher logs) created during the span of a course. This material, usually in the form of an artifact portfolio, is then analyzed by a specialized third party. There are several advantages and disadvantages of implementing this method to assess classroom behavior on a large scale. Artifacts analysis would be much more objective than traditional surveys and ideally the data would be collected on a real-time basis, thereby minimizing the time-lag between behavior and assessment. The time required to assemble the portfolio could be from a few days to the entire span of the course and all data could then be analyzed remotely by a small group of highly trained staff without the necessity of having an observer present in the classroom. Major disadvantages of artifact analysis include the required time investment on the part of the instructor and the inability of artifacts to capture the more intimate parts of classroom behavior (e.g. classroom discourse, instructor temperament). Still, it is very likely that incorporating artifact analysis as a stand-alone method or in conjunction with traditional surveys would yield invaluable information in objectively assessing instructional classroom behavior. It is, therefore, our intention to review the past research literature on the use of artifacts in determining instructional classroom behavior and report on the following questions:

- What types of artifacts would be available in engineering classes?
- What types of instructional behaviors can be assessed via artifacts?
- What would be the relative faculty time commitment in collecting each type of artifact?
- How might this type of assessment be implemented on a large scale?

PAST ARTIFACT STUDIES: WHAT MAY BE ADAPTED TO ENGINEERING
In order to get an idea as to what other researchers have learned from artifact research in the past, it is useful to look at the chronological evolution of artifact studies. We will primarily focus on artifact studies done following the introduction of reform-oriented teaching practices. These are practices primarily defined by such documents as Project 2061: Science for All Americans[2], Benchmarks for Science Literacy[3], and National Science Education Standards[4] which laid out the teaching “best practices” for science education. In nearly all of the studies examined some list of rubrics was developed based on these documents, which was used to assess the examined artifacts. All of the past artifact studies taken into account here have been performed at the K-12 level. Authors of a recent article in the Journal of Engineering Education argue that, although K-12-tested techniques must be adapted to higher education populations, adaptation is likely to be much easier than invention.[5].

In 1995, Burstein, et al. reported on using artifacts (e.g. textbooks, assignments, exams, and teacher logs) as “benchmarks” against survey data about teachers’ goals, activities, and class content [6]. This study suggests that artifact data could provide a much more accurate assessment of instructional goals as well as gather useful and unbiased information about content and strategies. The overall results indicated that artifacts worked well as benchmarks for surveys and were easily collected from teachers.
In 2001, Clare 1 and Aschbacher reported on a study to evaluate the validity of using teachers’ assignments as indicators of classroom practice, specifically in urban language arts classes [7]. Several useful results were obtained from this study: First, ratings of classroom assignments yielded similar estimates of quality compared with classroom observations with specific regard to elements of constructivist practice (quality of classroom discussions, level of student participation in classroom discussions, quality of instructional feedback, and cognitive challenge of the lesson activity) but not with regard to actual lesson implementation (classroom management, level of student engagement in lessons, clarity of the learning goals, alignment of goals and lesson activities). Second, it was found that a minimum of two assignments and two raters are needed to produce a consistent measurement of quality in all studies performed. Lastly, with additional research focusing on the cost of rater training, the authors believe this method would be feasible for large-scale implementation.

Since the pilot study by Clare 1 and Aschbacher in 2001, Clare 1 and several colleagues have come together to develop the Instructional Quality Assessment (IQA)[8, 9] which is a toolkit designed to assess teacher instructional quality in elementary and middle schools. Post-analysis showed that the overall quality rating based on assignments was in excellent agreement with the rating based on direct observation for mathematics classes. The authors admittedly agree that the quality rating is based on very few key artifacts/behaviors and this raises questions about the accuracy of measuring such a complex phenomenon as instruction; however, they argue that obtaining a stable estimate of practice with as few tasks/observations as possible is important for reducing the cost and burden on teachers by this type of data collection. Collecting four assignments from teachers yielded a stable estimate of instructional quality at the teacher level in mathematics. When the data were aggregated across teachers the observation rubrics yielded a stable estimate of quality at the district level and were sensitive to differences in the reform environment of the two districts. All in all, this study has shown that it may be feasible, based on the assessment of a few assignment artifacts, to produce a reliable teacher instructional quality rating which may then be compared across faculty, institutions, or even regions.

From 2003 to 2007, Borko et al. reported on a series of studies using artifact packages (a version of a teacher portfolio) to characterize classroom practice [10-14]. This study series is, perhaps, the most extensive examination to date of artifact analysis used to assess classroom practice. Many different types of classroom artifacts were gathered into a portfolio dubbed the “Scoop Notebook”. The “Scoop Notebook” was designed to contain instructional materials (e.g. lesson plans, overhead transparencies, & grading rubrics), student work with corresponding teacher reflections, photographs of the classroom, and teacher reflections based on guiding questions posed throughout the scoop period. The overall results of the study indicated that (1) the consistency among raters was more than acceptable albeit slightly varied on a few dimensions, (2) the correlation of Scoop Notebook ratings and ratings of observations were acceptable on all dimensions and (3) a study much like this one could be feasible for large-scale assessment.

**ARTIFACT PORTFOLIOS/E-PORTFOLIOS AT COLLEGES AND UNIVERSITIES**

Although K-12 studies like the “Scoop Notebook” appear to be more widespread than those that might study college-level classrooms, the use of portfolios by college faculty is on the rise. It is estimated that on the order of 1000 or so colleges and universities in the United States are actively using or experimenting with teacher portfolios[15]. In Canada, the teaching dossier (a version of a teaching portfolio) has been successfully in use since the 1970’s. One need only spend a little time on a search engine to find numerous examples of colleges and universities around the United States evaluating[16] or using faculty and graduate portfolios[17-23] and e-portfolios[24, 25]. Peter Seldin, one of the most prolific researchers into teaching portfolios, argues that this is an increasing trend marking a historic change in higher education in which teaching is taken more seriously.

An good example of an innovative project designed to help faculty accurately capture the intellectual work that they put into teaching is the Peer Review of Teaching Project (PRTP)[26] initiated at the University of Nebraska in Lincoln. Due to the increasing frequency with which faculty are expected to document, assess, and make public their teaching practices, the PRPT sought to create a way that faculty could clearly and systematically (1) show this teaching work, (2) investigate and analyze student learning, and (3) publicly communicate this work to peers or the campus in general to receive feedback. In the PRTP, faculty work in teams, helping one another in the process of documenting their teaching. No specific teaching methodologies are forced upon the faculty so the process is largely faculty-driven. Instead, the objective is focused on documenting and analyzing student learning. In this manner, faculty members are encouraged to continually assess their own teaching styles and classroom activities. Systematic and long-term documentation provide faculty with evidence as to the successes and needs of their individual teaching methodologies. Faculty members develop their own classroom portfolio based on a systematic guided online web-portfolio platform. Faculty members design their online portfolios, upload statistics and supplemental documentation and all information is kept online for long-term documentation.

Several research groups are now looking at different types of online portfolio tools to be used in assessment of engineering programs [27-29]. A group of engineering education researchers at Virginia Tech has demonstrated the potential for an open-source e-portfolio program[28, 30] to be woven into the undergraduate program.

---

1 The former Linsay Clare is presently known as Linsay Clare Matsumura.
curriculum. Still other researchers at Penn State[29] have developed and tested an online system to aid in the collection of ABET assessment data. This system makes use of embedded assessment measures and involves faculty in all areas of the ABET process. The program is still in its infancy but initial results indicate that the system has been effective in collecting assessment data, improving communication between faculty, ABET committees, and administration.

**WHAT INSTRUCTIONAL BEHAVIORS ARE RELIABLY ASSESSED WITH ARTIFACTS?**

For comprehension, it is helpful to categorize the many classroom practices into categories or dimensions as in the “Scoop Notebook Study”[12], Reformed Teaching Observational Protocol (RTOP) [31], MOSAIC II Study[32, 33], IQA[8, 9] and others. It is apparent in these studies that in artifact-only studies the rubric categories are more generalized and fewer in number (there are 10 in Borko’s study) whereas, with studies incorporating direct observation, the categories are much more specific and in larger number (there are around 25-30 in the RTOP and MOSAIC II studies). By comparing these studies (artifact-only vs. direct observation) we can extract what practices have been assessed with artifacts and what practices have been assessed with observation but not with artifacts. Also, by using studies that are based on seminal science reform works and research[2-4] we are using frameworks which incorporate the necessary elements of active, engaging, and innovative pedagogies. As a result we have divided the resulting factors/elements/dimensions into four groups:

1. **“Hands-On”**. The extent to which students participate in activities that allow them to physically engage with scientific phenomena by handling materials and scientific equipment.

2. **Scientific Discourse Community**. The extent to which the classroom social norms foster a sense of community in which students feel free to express their scientific ideas honestly and openly. Students are expected to communicate their scientific thinking clearly to their peers and teacher, both orally and in writing, using the language of science.

3. **Connections/Applications**. The extent to which the series of lessons helps students connect science to their own experience and the world around them, apply science to real world contexts, or understand the role of science in society (e.g., how science can be used to inform social policy).

4. **Grouping**. The extent to which the teacher organizes the series of lessons to use groups to work on scientific tasks that are directly related to the scientific goals of the lesson, and to enable students to work together to complete these activities.

**Group 2: The four dimensions of classroom practice that were reasonably assessed with artifacts.[13]**

1. **Inquiry**. The extent to which the series of lessons involves the students actively engaged in posing scientifically oriented questions, designing investigations, collecting evidence, analyzing data, and answering questions based on evidence.

2. **Use of Scientific Resources**. The extent to which a variety of scientific resources (e.g., computer software, internet resources, video materials, laboratory equipment and supplies, scientific tools, print materials) permeate the learning environment and are integral to the series of lessons.

3. **Explanation/Justification**. The extent to which the teacher expects, and students provide, explanations/justifications, both orally and on written assignments.

4. **Assessment**. The extent to which the series of lessons includes a variety of formal and informal assessment strategies that measure student understanding of important scientific ideas and furnish useful information to both teachers and students.

**Group 3: The two dimensions of classroom practice moderately well assessed with artifacts.[13]**

1. **Cognitive Depth**. Cognitive depth refers to a focus on the central concepts or “big ideas” of the discipline, generalization from specific instances to larger concepts, and making connections and relationships among science concepts. This dimension includes two aspects of cognitive depth: lesson design and teacher enactment.

2. **Structure of Lessons**. The extent to which the series of lessons is organized to be conceptually coherent, such that activities are related scientifically and build on one another in a logical manner.

**Group 4: Additional elements/factors of classroom practice not proven to be easily assessed by artifacts (Table 3) can be extracted from the following studies and regrouped into two additional dimensions. (Assessed in studies using classroom observation [31-33])**

1. **Teacher Knowledge.**
i. Content knowledge. The teacher has a deep connected understanding of scientific facts and concepts and the ways in which they are used in the real world and
ii. Pedagogical content knowledge. The teacher has extensive strategies for communicating information and developing conceptual understanding in alternative ways.

2. Classroom Culture.
   i. Student/teacher relationship. The teacher acted as a resource person & facilitator, encouraging active participation of students in discussion and investigation and asking questions to help students focus and understand rather than simply presenting knowledge. The teacher was patient and encouraged students to generate alternative solution strategies and ways of interpreting evidence. Teacher made important and challenging scientific concepts accessible to all students not just to select groups. There was a climate of respect for what others had to say.
   ii. Student-centered classroom attitude. The focus and direction of lesson was often determined from ideas originating from students. Students’ questions/comments often determined direction/focus of classroom discourse. There was a high proportion of student talk & a significant amount occurred between/among students.

**WHAT TYPES OF ARTIFACTS MAY BE COLLECTED FROM CLASSROOMS? WHAT MAY BE LEARNED & WHAT IS THE TIME COMMITMENT FOR EACH TYPE?**

Combining the several artifact studies examined in this work, we are able to compile a list of the most common types of artifacts that would be available in a classroom. Table I represents all of the forms of artifacts included in the studies referenced along with comments as to the advantages and disadvantages of each artifact form. The most time-consuming artifact for teachers—teacher logs—tend also to contain the most useful information for assessing overall classroom practice. Artifacts like video and audio tapes were considered in some of these studies, however, they tend to be much more time consuming to process than other artifacts forms. Also, these types of artifacts are essentially another form of direct observation and so, for the purposes of this review (as well as most of the other artifact studies), are not considered. Other artifacts like photographs, board writing, and projects involve a small amount of classroom disruption or time-commitment for teachers. Many artifacts, however, such as assignments, handouts, and quizzes can be simply collected and grouped into an ordered portfolio format with very little effort on the part of the teacher. Since minimization of faculty time-commitment is key for faculty buy in, it is useful to ask whether enough information about classroom practice may be obtained with just a few minimum-time-commitment artifacts.

### Table I: The Different Forms of Artifacts That May Be Collected in a Classroom.

<table>
<thead>
<tr>
<th>Artifacts</th>
<th>Refs</th>
<th>Advantages / disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Textbooks</td>
<td>[6, 11]</td>
<td>Easily collected &amp; processed</td>
</tr>
<tr>
<td>Homework</td>
<td>[6, 11]</td>
<td>Easily collected &amp; processed</td>
</tr>
<tr>
<td>Quizzes</td>
<td>[6]</td>
<td>Easily collected &amp; processed</td>
</tr>
<tr>
<td>Tests</td>
<td>[6]</td>
<td>Easily collected &amp; processed</td>
</tr>
<tr>
<td>Student Work</td>
<td>[6, 7, 11]</td>
<td>Easily collected &amp; processed</td>
</tr>
<tr>
<td>Projects</td>
<td>[6, 7, 11]</td>
<td>Moderately easily collected &amp; processed</td>
</tr>
<tr>
<td>Calendar</td>
<td>[10-13]</td>
<td>Moderately easily collected &amp; processed</td>
</tr>
<tr>
<td>Guided Daily log /</td>
<td>[6, 11, 33]</td>
<td>Time-consuming for teacher -Highly informative</td>
</tr>
<tr>
<td>teacher reflections</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**LARGE-SCALE SAMPLING STRATEGIES**

In considering a study designed to assess engineering instructional practices we will also need to look at what sampling methods are best to use. Lattuca, et al.[34, 35] performed a pre- and post-analysis study of the implementation of EC2000 on engineering curricula across the USA. This sampling methodology of this study could be successfully adapted to a nationwide artifact study. This three year study sought to quantitatively measure change in the graduates as a function of the enacted curricula. Engineering programs were selected from a pool of programs across the nation based on a 2-stage, disproportionate, stratified, random sampling. This ensured adequate sample size and program representation. 1024 programs offering degrees in the seven (7) main engineering disciplines (aerospace, chemical, civil, computer, electrical, mechanical, and industrial) were selected to be represented while stratification into subgroups was based on each institution’s prior participation in formal engineering education improvement programs/assessments (e.g., participation in an NSF engineering coalition in the 1990’s). Randomization ensured equal probability that any institution would have its programs selected. Selection was disproportionate in that the smaller disciplines were over-sampled to ensure an adequate number of responses for sampling. Several Historically Black Colleges and Universities (HBCU’s) and Hispanic Serving Institutions (HSIs) were added to ensure their representation. Using this methodology, this study ensured adequate sampling of each engineering discipline, school region, and institution type.

**CONCLUSIONS AND LARGE SCALE IMPLEMENTATION STRATEGIES**

In this work we have sought to answer the major questions which come to mind if we wish to assess engineering faculty behavior via artifacts of the classroom. Specifically, we have addressed the types of artifacts one might find in a typical situation related to engineering classes and how the artifacts are collected and processed.

October 22 – 25, 2008, Saratoga Springs, NY

38th ASEE/IEEE Frontiers in Education Conference

S4D-4
or choosing artifacts that are less burdensome to collect, the or (2) convince administration, faculty, and students that the

of artifact. In order to implement a successful large-scale

can/cannot be assessed well using artifacts, and what is the

engineering classroom, what instructional behaviors can/cannot be assessed well using artifacts, and what is the

relative time commitment of collecting/processing each type

of artifact. In order to implement a successful large-scale

assessment using classroom artifacts, we must bring all of

the available information together.

The types of classroom practices that may be assessed with good success via artifacts are as follows, in order of accuracy of assessment: “hands-on” activities, scientific discourse community, connections/applications, collaborative grouping, inquiry-based learning, use of scientific resources, explanation/justification, student assessment, cognitive depth of lessons, and structure of lessons. The remaining two classroom practices that are less likely to be accurately assessed via artifacts are: teacher knowledge and classroom culture. A portfolio could likely be developed that incorporated the first ten practices and used alternate methods, i.e. student surveys, to supplement the assessment data.

One major obstacle foreseen in using artifact analysis on a large scale is faculty/student time commitment. Depending on the thoroughness of the artifact collection, the burden on faculty could be quite high. To reduce this perceived burden two options are proposed: either (1) minimize the number/burden of artifacts collected in a study or (2) convince administration, faculty, and students that the extra work is beneficial to their own goals (e.g., learning and teaching). By minimizing the number of artifacts collected or choosing artifacts that are less burdensome to collect, the original goal of assessing behavior could be met with little faculty resistance. Such artifacts as homework, tests, syllabi, projects, etc. along with teacher comments must be prepared anyway and could simply be collected and displayed in a portfolio much like the scoop-notebook. Faculty might be asked either to collect only successful examples of student work or a few examples of successful, mediocre, and unsuccessful work. This could give, not only an idea of the assignment but also of the criteria used to judge the assignments. Possibly the most desirable scenario would be that both faculty and students perceive the collection of artifacts as highly advantageous to learning and teaching. If this situation is achieved, there will be minimal institutional resistance to implementing a large scale artifact study.

It is highly likely that future large-scale assessments at the college level will take place via internet and no assessor will need to set foot in a classroom. An online e-portfolio version of the scoop notebook could prove to be even less burdensome and more professionally valuable to faculty and students and therefore, it is advisable to examine successful models of online portfolio programs already in place. In light of this, several things have made past e-portfolio studies successful:

- An e-portfolio program that is geared towards meeting the goals of faculty, department, and institution, is more likely to be accepted at all levels (student, faculty, department, institution) and be able to be interwoven with the curriculum [15, 26, 28, 36]. It is a small step to be able to externally assess a successful e-portfolio program already in place.
- A system that proves to lessen (not heighten) the burden of teaching will likely increase in use among faculty. E-portfolios can do this by serving as a template for future classes, and a resource for students.

Additionally, Several things are suggested by Seldin [36] to improve chances of faculty acceptance and “buy-in”.

- Administrators at all levels should collaborate with faculty to set clear goals for both outstanding teaching and acceptable teaching
- Portfolio evaluation criteria should be publicized so that faculty members are clear about the expectations
- Portfolios should be prepared with a mentor for faculty new to the process
- The program should be thoroughly pilot tested on volunteers, including some of the most prestigious teachers and researchers on the faculty before any attempts at institutionalization
- Institutions should support portfolio development workshops and mentorships, e.g., by providing release time or compensation for workshop leaders
- Administration should take the portfolio process seriously by rewarding faculty with strong teaching and research portfolios without favoring one or the other.

In order to begin implementation on a large scale, one might standardize the types of artifacts required in an e-portfolio while at the same time allowing a great amount of freedom to the instructor to teach innovatively. By modeling an e-portfolio system on some successful previous pilot studies [26, 28] and incorporating a few guiding standards of artifact types to collect and display [9, 12] it is feasible that information on classroom practices could begin to be gathered on a national level. Taking Peter Seldin’s words into consideration, we could imagine using collaborative focus groups composed of professors, students, administration, and educational researchers to design a nation-wide e-portfolio program. The feeling of partial ownership of the process would ensure that the parties at the grass-root level would be invested in the process. Clear goals and grading criteria as well as an informative e-portfolio mentoring process could be designed into the system by the group. Faculty and students must see this program as being simple, efficient, and a benefit to teaching and learning. The potential for a standardized widespread online portfolio for college teachers has never been more possible than now. The momentum is already present within thousands of universities nationwide. We need only bring together academic institutions, teachers, students, administration, and all of the interest groups present and begin pilot testing online e-portfolio studies.

ACKNOWLEDGMENTS

We’re grateful for NSF support via award DUE-0618125. The views expressed are those of the authors alone. Dr. Tomlinson was supported as a Christine Mirzayan Science and Technology Policy Graduate Fellow in CASEE.
REFERENCES


[26] CWRU, *Case School of Medicine: Faculty eportfolio.* 2008 27 Feb; Available from: http://cerebrum.case.edu/ePortfoliofac/.


