2013

How engineers learn: a study of problem-based learning in the engineering classroom and implications for course design

Robert Mayer
Iowa State University

How engineers learn: a study of problem-based learning in the engineering classroom and implications for course design

by

Robert Roger Mayer

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Industrial Engineering

Program of Study Committee:
Richard T. Stone, Major Professor
Stephen B. Gilbert
William N. Dilla

Iowa State University

Ames, Iowa

2013

TABLE OF CONTENTS

ABSTRACT

This paper presents two studies in the methods of problem-based learning (PBL), where students work in small groups to explore specific problems under the guidance of an instructor. PBL has proven to be highly-effective in engineering education, but there is still room to improve. PBL opponents charge that the unstructured learning environment characteristic of PBL is contrary to cognitive theories on how people learn. Proponents of PBL argue that it is designed to offer just enough structure that students will succeed at learning the material. Studies by Schmidt et al. (2007) suggest that the use of "scaffolds," structures that support the conceptual learning process early-on, but are gradually removed later, can greatly help when students first engage in PBL.

In this paper, the design and use of scaffolds was tested using two scientific studies in an entry-level engineering course. A total of 94 students participated in the research. In the first study, two different worksheets (hard scaffolds) were evaluated; one provided far more structure than the other in conducting the lab procedure and calculating results. After preparing a full lab report, students were given a post-lab examination. The results of this examination indicate that the highly-structured scaffolding was significantly more effective at facilitating the learning outcome. In the second study, the use of lectures and supplementary text (soft scaffolds) was evaluated alongside the improved hard scaffold developed for study one. Students were either given (a) no lecture or supplementary text, (b) lecture/text before lab, or (c) lecture/text following lab. The second study found no significant difference across groups.

When taken together, these results indicate that the process of PBL is effective at teaching students difficult engineering concepts. Specifically, they show that it is the actual process of studying a problem where students learn the most, in contrast to being "fed" information from a lecture or textbook. The results further indicate that students must be provided with a highly-structure scaffold to achieve the highest learning outcome. While further study is needed, the implication for engineering course design is that lectures should be reduced or eliminated in favor of more hands-on problem solving encounters.

1. INTRODUCTION

What distinguishes man as a species is not only his capacity for learning, but for teaching as well (Wood, Bruner, & Ross, 1976).

At today's modern academic institutions, teachers have proven capable of carrying out a remarkable level of instruction. Even so, the question "are we teaching well enough" is often heard, particularly in engineering programs. Haag, et al. (2011) observe that "the current work environment requires engineers to be global citizens, as well as aspirational, ethical leaders." Are our educational programs able to meet this requirement? Is it even possible that these skills can be "taught" in the fundamental sense? Others question the educational process as a whole, blaming it for increasing attrition rates from engineering programs (Lord & Camacho, 2007). One might even ask what defines a "good" educational process.

For years, a significant decline in the number of students graduating with degrees in STEM (Science, Technology, Engineering, and Math) fields has been observed (Besterfield-Sacre et al., 1997; Beaufait, 1991; Astin, 1993). Despite numerous attempts to understand the problem (Heckel, 1996; Hermond, 1995; Besterfield-Sacre et al., 1997, and many others), we still do not have a reasonable understanding as to why this trend persists. Perhaps our primary schools are not placing enough emphasis on science and its practice (Lee & Burkam, 1996). Perhaps students are not receiving high-quality instruction from their teachers and are arriving at college unprepared (Smith, 2007). Perhaps our society has come to value ease of passage over hard work (Hewitt & Seymour, 1992). Regardless of the cause, it cannot be denied that first- and second-year engineering students leave their respective programs at an alarming rate, and that this trend persists across all engineering disciplines at all institutions. (Besterfield-Sacre et al., 2013). It has even been suggested that this trend poses a serious risk to the United States' global standing and security (Nusca, 2010).

Felder (1993) argues that more effective teaching methods in introductory courses will result in a higher retention rate of early students. One of these methods is known as Problem-Based Learning (PBL), where students work in small groups to explore a specially-designed problem under the guidance of an instructor. They observe that such collaborative learning environments hold promise but "most experiments with these methods have been carried out on a one-shot basis: a professor tries a new method in a course…many students respond well to the new method; and most of them never see anything like it again" (Felder, 1993).

The observations by Felder and others, even though made two decades prior, are still applicable to the education research scene today. While it is easy to find published papers on a myriad of education-related topics (e.g., the effect of instructor facial expressions on learning outcomes, Theonas, et al. 2007), controlled, repeatable experiments on the major cognitive theories of learning and their applications are difficult to dig up (Pease & Kuhn, 2010). As these (and other) authors indicate, most published articles focus on the broad effects of a method specific to a particular course, but without a precisely-controlled experiment, it is impossible to discern what underlying cognitive process made the method effective. Without knowing why a method is effective, using it to establish a general set of guidelines for how courses should be designed or taught is a futile endeavor.

While this paper does not seek to address all of the fundamental issues related to engineering education research, it does hope to shed some light on specific processes and tools– namely, PBL– that have been proven to be effective when used correctly. It does this with the hope that education as a whole, and engineering education in particular, can continue to improve and produce students who, like their teachers, will contribute to positive change to the world.

Objective

This research evaluates one specific aspect of PBL: the use of scaffolds. Studies by Schmidt et al. (2007) explain that scaffolds are structures that support the conceptual learning process early-on, but are gradually removed later. Research suggest that scaffolds can greatly help when students first engage in PBL. PBL opponents charge that the unstructured learning environment characteristic of PBL is contrary to cognitive theories on how people learn; however, proponents argue that it is designed to offer just enough structure that students will succeed at learning the material. Scaffolds may hold the key to this debate.

The objective of this research is to demonstrate the effectiveness of scaffolds when using PBL in a real engineering classroom. A secondary objective of this research is to offer guidance on how scaffolds should be designed and presented to students during the PBL process. To achieve this, two separate studies are conducted. The first explores the effect of problem-solving structure provided through worksheets (hard scaffolds). The second evaluates the effectiveness of supplementary text and lecture (soft scaffolds), given an effective level of structure via hard scaffolds. The effects of these method are evaluated on each student's overall subject-related knowledge performance on a post-lab examination.

Genesis of this Research

IE (Industrial Engineering) 248 (*Engineering System Design, Manufacturing Processes & Specifications*) is a sophomore-level core course in the IE curriculum at Iowa State University. The class serves as an introduction to basic concepts on metrology, engineering drawings, specifications, quality issues, and the design and improvement of systems. It is taught in a PBL format, where students attend (3) one-hour lectures and (1) two-hour PBL-based lab per week. In lab, they reinforce, practice, and apply the concepts taught in lecture. It should be noted that there is not a 1:1 correspondence between lecture and lab material; therefore, there are cases where students are "on their own" regarding the laboratory. Following each lab session, students are expected to work in groups to prepare lab write-ups, which strongly follow the format and content of a research article (introduction, objective, methods, results, discussion, and references).

The first PBL session in the course is on *measurement variation*. This particular topic integrates concepts of probability and statistics, while making those concepts concrete through students' exploration of their own data collected during the lab. Because it is the first PBL session in the first PBL course that students encounter in this curriculum, students seem to have great difficulty with it. Nine years' results of this lab were unsatisfactory, and students and teachers were both frustrated. Thus the hypothesis emerged that the lab should be changed, and more instructional support (i.e. a scaffold) was needed. This study originated out of a desire to quantitatively evaluate the effect of this change to the instructional materials; however, it has since evolved into a general evaluation in the methods of PBL-based instruction.

As an additional factor, the topic taught by this course module is a particularly challenging one. As research (and practical experience) by Konold has shown (Konold, 1989; Konold et al. 1993; Konold, 1995; Konold and Pollatsek, 2002), students have a particularly difficult time grasping the fundamental concepts of variability and incorporating them into "statistical thinking" (Feder, 1984; Wild & Pfannkuch, 1999) about the real world. At the same time, Vidic (2011) and Mateo-Sanz and colleagues (2010) observed that proper understanding of statistical principles is critical for engineers to be successful in their field. Thus, PBL methods are particularly interesting in this case, as the complex reasoning fostered by PBL is theorized to assist in the development of a working knowledge of statistical concepts essential to the field of engineering.

2. REVIEW OF THE LITERATURE

How People Learn

A reasonable discussion about educational methods should begin with mention of relevant theories of cognition in the learning process. As Kirschner and colleagues (2006) pointedly observed, "any instructional procedure that ignores the structures that constitute human cognitive architecture is not likely to be effective." It should be added that addressing cognitive structures is important to understanding *how* and *why* a particular method is effective, and this information is far more useful than simply know *what* works. It is therefore important for this discussion to give a general overview of the present understanding of learning, so that a reasonable interpretation may be made as to how the methods described within this paper achieve their effect. It should be noted that this discussion serves as a spotlight on specific concepts to be addressed, rather than a comprehensive explanation of how the brain works. Indeed, the process of learning is not well understood (Koedinger, et al. 2012), and future research will likely reveal new insights not foreseeable here nor compatible with the methods adopted by this research.

Cognitive Constructivism

First, this paper accepts the fundamental theory that individuals learn by actively integrating new knowledge into their existing cognitive framework. A smorgasbord of familiar research, from Jean Piaget to Donald Norman and others use the concept of mental models or schema into which knowledge is integrated within the learner. This knowledge includes both declarative and procedural types stored into long-term memory and accessible to the learner in short-term working memory via a central attention executive (Anderson, 1983). According to Anderson, the degree to which knowledge in working memory is accessible depends on the number and strength of links (relationships) that an individual is able to form among various concepts (Anderson 1983). Therefore, the process of learning causes new knowledge to be linked to existing knowledge, and students "learn" when the strength or number of these connections is great enough to facilitate spontaneous recall. Theories which rely upon this fundamental understanding are collectively known as "constructivist" theories (Derry, 1996).

Next, since learning integrates new knowledge with existing knowledge, students' existing conceptions are very important to the learning process. It has been established, for example, that students whose existing mental models are incompatible with new knowledge ultimately end up with distorted conceptions of the subject to be learned (e.g. Vosniadou and Brewer, 1989, where students imagining a spherical earth from the prior perspective of a "pancake" earth might see a pancake sitting on top of a spherical earth; the pancake idea is never eradicated). A critical review by Streveler and colleagues (2008) describes an increasing importance placed on understanding why students' misconceptions in science and engineering are difficult to correct. Thus, any attempt to teach new material, and especially understand a particular teaching method, must account for the pre-conceived notions of the students to be taught.

Cognitive Load Theory

In addition to the above theories, Cognitive Load Theory (CLT; Paas & van Merriënboer, 1994; Sweller, et al., 1998) should be mentioned. This theory explains how learners acquire information under various conditions of task and context, and it originates due to the limited capacity of working memory. When students acquire information, their working memory holds information from both the new source and their existing long-term memory base. The working memory space allows the new concepts to be integrated into the old; thus, if it becomes overloaded (due to too many disorganized "chunks" of information), the student will not be able to integrate new knowledge into their existing mental schema (Sweller et al., 1998). This implies that additional tasks which are irrelevant to the learning process (such as finding and sorting through journal articles, for example) only serve to reduce the amount of knowledge gained by the learner (Sweller, 1994). This is known as *extraneous cognitive load*, which stands in contrast to *intrinsic cognitive load*, a characteristic of the material being learned (mostly related to the complexity of the concepts in the material) (Sweller, 1994; Sweller & Chandler, 1994).

CLT has been a source of debate about whether or not so-called discovery-based methodologies (i.e. PBL and similar) can achieve their intended effect. Opponents of PBL say that these methods require too much extraneous load on the part of the learner (Kirschner, et al. 2006). PBL proponents argue that the very struggle to find and document information is what defines these methods and makes them effective (which causes the discovery process to be categorized as germane cognitive load; Schimdt et al. 2007). As will be explained in greater detail later, the present paper accepts as a basic hypothesis that students learn best when they are able to focus on relevant problem-solving tasks rather than handle unnecessary and irrelevant searches for information (as exemplified by the use of scaffolds; Hmelo-Silver, et al. 2007). This does not mean that CLT is any more or less valid in the current context, but the basic premise that reducing unnecessary cognitive load allows better learning is accepted to be true.

Critical Thinking in Engineering

It can be said that the most important quality in an engineer is the ability to think critically—to analyze a given proposition from a variety of angles and determine its strengths and weaknesses. Critical thinking is generally accepted as a primary goal of higher education (Mason 2007, Abrami et al. 2008, Bailin and Siegel 2003, Sheffler 1973). As educators, it is our responsibility to build these skills in our engineering students, or we will produce engineers who can pass courses but who cannot function in the workplace. Therefore, both the teaching and evaluation of critical thinking skills are very important in an engineering curriculum (Douglas, 2012). It stands to reason that the goals of an engineering education process might best be met through methods which can be proven to build critical thinking skills, in addition to conveying subject-specific knowledge. This implies a need to reliably evaluate both students' contextual knowledge and their critical thinking abilities simultaneously.

As recently as 2008, Abrami and his colleagues performed an exhaustive analysis of literature concerning the teaching of critical thinking skills. They mainly found that methods which attempt to teach critical thinking skills are generally successful at doing so. In particular, they found that students learn critical thinking most reliably when CT skills acquisition is a stated objective of the course; indirect approaches, where students are expected to learn CT as a by-product, are generally not as effective.

There are, however, some important limitations to the current literature on critical thinking. As Abrami et al. (2008) observe, "[critical thinking] is a complex and controversial notion that is difficult to define and, consequently, to study." Most definitions take a pornographic ("I know it when I see it") approach to defining critical thinking, which is not terribly useful from a research perspective. On the other hand, the current scientific definition of critical thinking (as defined by the American Philosophical Association – Facione, 1990) comprises the better part of a paragraph, which is nice, but intractable (Anderson et al., 2001). Therefore, it can be argued that there needs to be an engineering-specific definition of what comprises critical thinking, and that this specific skill can be taught in the classroom. Such an undertaking is far beyond the scope of this paper.

Finally, as Douglas (2007) and others point out, there are few critical thinking studies in the literature specifically addressing engineering education. It remains to be studied if engineering instruction methods are more effective than others at teaching critical thinking skills. In fact, much of the recent literature on teaching methods seems to forego any mention of it at all. Others seem to be divided on the cognitive processes that underlie critical thinking skills and the techniques that should be used to teach them. Therefore, this paper does not hope to offer any new theories or provide evidence supporting one specific theory over another. The results presented later will only serve as a starting point for later discussion on the issue of teaching critical thinking in engineering education.

Problem-Based Learning

One major undertaking has been the implementation and assessment of Problem-Based Learning (PBL) in a wide variety of educational settings (Duschl, 2008; Lehrer & Schauble, 2006), including engineering education. According to Gijbels et al. (2005), PBL is a discovery-based learning method "intended to guide students to become experts in a field of study, capable of identifying the problems of the discipline and analyzing and contributing to the solutions." To achieve this, PBL-based courses are typically taught in a split fashion (lecture plus lab), where students have the opportunity to work in groups of three to five and solve problems under the guidance of seasoned instructors (Schmidt et al., 2007).

How PBL Works

While the precise mechanism behind the effectiveness of PBL has not been determined (Pease and Kuhn 2010), several theories serve to explain what is going on inside the brains of the students. The most germane (and contentious) of these is Cognitive Load Theory (CLT) (described in greater detail above), which holds that students learn best when their cognitive efforts are mostly spent internalizing new information rather than trying to figure out what to learn or where to find the material. For difficult concepts, students are going to spend a large amount of effort simply integrating new knowledge into their existing mental schema (Kirschner, et al., 2006). In layman's terms, CLT implies that students require a large amount of guidance during the PBL process in order to achieve a successful learning outcome, particularly if they have not been exposed to the material previously. Some hypothesize (e.g. Hmelo-Silver, 2004; Pea, 1993; Salomon, 1993) that PBL works because the cognitive load is shared among the various group members, "taking advantage of…distributed expertise by allowing the whole group to tackle problems that would normally be too difficult for each student alone" (Hmelo-Silver, 2004). In theory, by the time they have finished the process, students have integrated new knowledge into their existing cognitive framework and have formed sufficient connections so as to make that new knowledge accessible (Schmidt et al., 2011 and many others).

PBL Controversies

While many of the benefits of PBL, such as increased abilities in problem-solving, are apparent and well-proven (Gijbels et al., 2005), there are many questions about PBL that remain unanswered. In particular, it is not entirely clear what factors or underlying cognitive processes make PBL successful (Schmidt et al. 2007) and to what degree PBL-based curricula are effective at conveying conceptual knowledge (Gijbels et al., 2005; Streveler, et al., 2008). Meta-analysis of literature by Gijbels et al. (2005) concluded that PBL is better at increasing students' understanding of the principles that link various concepts, as well as helping students apply those principles and concepts during problem solving. However, when addressing the transfer of conceptual knowledge, Gijbels et al. (2005) found that PBL and conventional approaches (lecture only) were not significantly different.

One place where this debate is particularly poignant is in a recent literary interchange regarding the use of minimally-guided teaching versus direct instruction techniques. In a series of articles by Sweller, Kirschner, and Clark (2006, 2007), the authors steadfastly defend the use of direct instruction over problem-based learning and other techniques that place emphasis on self-guided learning. The responses to these papers (and in fact, the raising of the issues in the first place) appears to be more rooted in opinion than in the results of controlled experiments (a fact which Sweller et al., 2007 readily admit). Sweller and his colleagues specifically call out the minimally-guided nature of PBL (and other similar approaches), a nature that is disputed by the proponents of PBL.

At the crux of this debate is whether direct instruction or PBL is better at fostering critical thinking skills. Is critical thinking a result of specific practice in problem-solving techniques, as PBL proponents claim, or must students have robust conceptual knowledge before critical thinking is possible? Sweller, Kirschner, and colleagues use the cognitive theories previously mentioned in the present paper to simultaneously support their position and rebut those who tout PBL. However, it is not obvious by reading their papers that they specifically address critical thinking skills or the role that such direct-instruction techniques play in building them in students. Their logic dictates that direct instruction builds robust conceptual knowledge in students, such that critical thinking capacity eventually emerges on its own; however, it is not obvious that this happens, and a meta-analysis by Abrami and colleagues (2008) found that the opposite is true: CT skills are best enhanced when the explicit goal of instruction is to build them. Furthermore, the "emergent CT skills" hypothesis assumes that the ability to think critically about a subject is entirely dependent upon an individual's existing expertise in that subject (i.e. CT skills are non-generalizable), an assumption which is somewhat dubious (Abrami et al., 2008).

Regardless of their intent, the discourse by Sweller and colleagues leaves the impression that it is better to give students all the answers up front than to train them expressly to think for themselves. Many engineering instructors would find this approach objectionable (Mina, et al. 2003). At the same time, nobody has yet decisively proven whether either approach (PBL vs. direct instruction) is effective at teaching critical thinking skills. This question must be answered before meaningful scientific discussion on the issue is possible; hence, there remains a considerable amount of work to be done. The purpose of the present paper is not to enter into either side of this debate, but instead to offer some objective scientific evidence for the effectiveness of one specific version of PBL.

PBL Defined

Part of the reason for these contradictory and confusing reports in the literature is that PBL has been widely adopted across disciplines and educational settings. In a critical review of the literature, Taylor and Miflin (2008) highlight the various ways that PBL has been applied. Throughout this process, they argue, the very definition of PBL is "as elusive [at present] as it has been since the concept was first expressed over forty years ago." They observe that the very process of disseminating the methods of PBL have resulted in countless variants and methodologies applied in panacea-like fashion, not all of which are compatible with the original intentions for (or theories behind) PBL.

With this in mind, it is important to identify what PBL means in the context of this paper. The definition of PBL used here is largely consistent with the original ideas of Barrows and Tamblyn (1980), as well as the findings of Taylor and Miflin (2008); however, small details (such as the size of the groups) have been reduced in accordance with more recent research into group sizes. In general, however, PBL is defined to consist of a multi-phased collaborative approach to education where students gain knowledge as they work in small groups (3-5 students) and attempt to solve a problem carefully-designed by the instructor. Throughout the problem-solving process, students work together, integrating existing knowledge and seeking out new knowledge, all with the help of the instructor. The key to PBL is that learning, for the most part, is pull-based (students seek the necessary knowledge) rather than push-based (students are fed knowledge by an instructor) (Schmidt, et al. 2011).

Controversial Definition and Use of Scaffolds in PBL

As mentioned previously, a major criticism of PBL is its unstructured nature. Proponents argue, however, that PBL does have structure where it is necessary. One way to provide structure is through the use of scaffolds in the PBL process (Reiser, 2004). Scaffolds are any item or tool that provides additional structure to the PBL process (Schmidt et al., 2007; Simons & Klein, 2007; Saye & Brush, 2002), allowing students to achieve success in learning where they otherwise would not (Wood et al., 1976). Scaffolds work by reducing the amount of cognitive effort that students must expend to learn the material; by providing students with concepts beforehand, students' attentional processes can be focused on the problem rather than on knowledge acquisition (Schmidt et al., 2007). Unfortunately, the overly-broad definition of *scaffold* offers little guidance as to how they should be implemented (Pea, 2004; Simons & Klein, 2007).

There seems to be general consensus that scaffolds can either take the form of a lab handout or worksheet (a hard scaffold) or a tutor or instructor (a soft scaffold) (Simons & Klein, 2007; Saye & Brush, 2002). Authors also agree that, in general, as students become more accustomed to the PBL process, the use of scaffolds can be gradually reduced until students are primarily responsible for their own learning (Schmidt et al., 2011). What constitutes effective scaffolds, to what degree the scaffolds should be reduced, and how much time the reduction should span are the topics of current and future research.

Recently, there has been extensive debate on the use and effect of scaffolds in PBL. Some argue that scaffolds are ineffective because PBL methods rely on unstructured interactions to effect learning (Kirschner et al., 2006; Choo et al., 2010). Others say this argument is illogical because all instruction must at least contain some form of structure, or it would not be effective (Schmidt, et al. 2007; Simons & Klein 2007; Hmelo-Silver et al. 2007). This has led to differing opinions on how PBL sessions should be conducted and what materials should be provided to assist students in the learning process (Taylor & Miflin, 2008).

At the same time, research into the efficacy of scaffolds (and other aspects of PBL) has been fraught with difficulty (Pease & Kuhn, 2010). Experiments on PBL are problematic to conduct because extraneous factors (like motivation, content, instruction, and assessment methods) are difficult to hold constant over an extended study. Furthermore, studies are nearly impossible to compare because of inconsistent assessment methods (Belland, 2008), inconsistent definitions of PBL, and inconsistent ideas as to how learning sessions should be conducted (Taylor & Miflin, 2008). Furthermore, the effect of time spent learning in-class versus out-of-class has not been effectively evaluated (Pease & Kuhn, 2010). This has made the debate over PBL less about the cognitive science behind it and more about procedures and research methods used to evaluate it (Colliver, 2000; Belland et al., 2008). As a result, the literature cannot be said to contain a reliable set of facts, much less a theory, regarding PBL or the use of scaffolds.

This paper contributes to this research by evaluating the effect of scaffolds in an actual PBL setting over a short amount of time, where the effects of confounding variables are minimized. By taking this approach, the intention is to provide evidence supporting the use of scaffolds, as well as qualitative methods on how they might be designed, presented, and used.

3. METHODS

Two separate, related studies were conducted. Study 1 evaluated the difference between different types of scaffolding. Study 2 evaluated the difference between instruction methods. Thus, this section will first describe the methods common to both studies, then follow up with specific procedures for each study.

Common Components

Both studies presented students with a real problem-based laboratory session which they would experience during the normal progression of their coursework. The study environment was designed to simulate the original course environment as much as possible, using the same instructors, contact time, format and organization, physical location and setup, and work content as the original course. It should be noted that there are aspects to the present course design (e.g. group size and group assignment process) that may not be ideal, but those aspects are not the subject of the study and are kept consistent for simplicity.

Study sessions were targeted for groups of up to 10 students, and were held on several afternoon/evening times within one week of each other. All participants completed a pre-test, a one-hour laboratory session, a group homework assignment, and had optional group work time, during which the instructors were present if they had questions on the homework assignment. Students were expected to complete the homework assignment within a week of the lab. All students returned approximately 2-3 weeks later (following a mid-semester break) for a post-test. The laboratory session, homework assignment, and post-test were substantially identical between the two studies. The studies varied the lab procedure and lecture content independently, as described below.

Study participants were chosen from among various engineering courses at Iowa State; as motivation, participants were offered extra credit in those courses (the percentage varied slightly by course, but was targeted to be about 1/3 of a letter grade), and were told that their extra credit points would depend on their performance on the homework and final exam. Participants in both studies were either "experienced" or "inexperienced" based on the number of courses they had taken. A very small minority of the students (7%) participated in both studies; this was determined not to present an issue because the studies, while similar, were not identical, and they were conducted far enough apart that participants would remember little of the first study procedure (it should also be noted that a t-test was run on those students, and found no significant difference in their scores).

Study One

This study evaluated the effect of schema-congruent worksheet scaffolding on students' learning performance. Students were randomly assigned to either a control group or experiment group. The control group presented students with a weakly-congruent worksheet scaffold, while the experiment group presented students with a highly schema-congruent worksheet which had been designed in accordance with Donald Norman's User-Centered Design principles (Norman, 1990). The schema-congruent worksheet presented information in a more congruent sequence, including requiring students to perform the calculations and answer the discussion questions in a specific order. These materials are available in Appendix A. The differences in the control and experiment sessions are summarized in Table 1 below.

TABLE 1: COMPARISON OF CONTROL SETUP AND EXPERIMENT SETUP FOR STUDY ONE (DIFFERENCES IN SCAFFOLD)

Item	Control Setup	Experiment Setup
Amount of concept knowledge provided in lab handout	Much less	Much more
Correlation of lab procedure with concept definitions	Not correlated	Correlated
Structure and explanation of methods in lab procedure	Less structure	More structure
Number and complexity of calculations	(Equivalent)	(Equivalent)
Applicability of Calculations to Learning Objectives	Unclear	Clear
Number and type of concepts/applications addressed in lab write-up	(Equivalent)	(Equivalent)
Time spent conducting lab experiment	(Equivalent)	(Equivalent)

Study One Data Collection

Data in this study was collected from written short-answer exams, which is the same format as presented in the real class. The pre-test was substantially shorter than the post-test and evaluated students' pre-existing conceptual knowledge. The post-test was a 30-minute exam which had greater depth of conceptual evaluation as well as application questions (requiring students to both understand the concept and apply it in a specific situation). For the text of the pre- and post-tests, see Appendix B.

The exams were administered in a quiet testing environment. Students were informed of the time remaining, and were encouraged (but not required) to finish the exam within the allotted time slot. Questions were graded by two independent graders following the conclusion of the study, and discrepancies were resolved by discussion between graders. The questions were designed to elicit precise responses; as a result, grading was generally binary (either entirely correct or entirely incorrect); however, an occasional answer earned partial credit for being correct but incomplete. The grading philosophy was as follows:

- Correct answer, correct justification = 100%
- Correct answer, partially-correct justification = 30-60%
- Incorrect justification OR incorrect answer = 0%

Because the pre- and post-tests were different, scores could not be directly compared between them to determine whether students learned; the rationale for this is that a direct comparison would have taken much more time on the students' part, and would have reduced their overall motivation to learn the material, nor would such a comparison help achieve the study objective.

Study Two

This study evaluated the effect of general instruction on the topics to be learned during the lab. Students were randomly placed into three sessions: control, experiment I and experiment II (these sessions are summarized in Table 2 below). All sessions used the schema-congruent scaffold design presented in *study one*. In addition, a brief (20-minute) lecture and four pages of supplementary text were provided to students in the experiment sessions, while control students received no lecture or supplementary materials. The supplementary text and lecture slides (available in Appendix C) highlighted the major concepts to be learned from the lab, and the lecture in particular made links between the various concepts clear to the students. Students were encouraged to ask questions during the lecture.

TABLE 2: LIST OF RUNS FOR STUDY TWO

Item	Lecture	Lab Handout Contents
Control	None	Procedure Only
Experiment I	Before Lab	Lecture Material + Procedure
Experiment II	After Lab	Procedure + Lecture Material

Study Two Data Collection

The Blackboard course management system was used to manage the bulk of data collection in the course. This data consisted of (1) personal data questionnaire, (2) pre-lab quiz, (3) group homework scores, (4) post-lab quiz, and (5) post-study questionnaire. Quiz questions were generally weighted based upon the importance of the concept to be conveyed (more important or higher-level concepts received greater weight). Administration and grading of the exams was entirely consistent with the procedure outlined for *study one* above.

The personal data questionnaire consisted of 13 multiple-choice questions. There were no correct or incorrect answers. It was administered and scored by the Blackboard course management system. The test questions are available in Appendix D. In general, the test asked questions on the following subjects:

- Gender, major, year in school
- Prior coursework
- Prior hands-on lab experiences
- Academic performance

The pre-lab quiz consisted of 22 multiple-choice questions. It was administered in a controlled test environment by the Blackboard course management system. The test questions are available in Appendix E. The test asked questions related to the following subjects:

- General critical thinking/problem-solving ability (5 questions)
- General probabilistic/statistical reasoning (6 questions)
- Specific measurement variability knowledge (10 questions)

The homework assignment consisted of 7 questions designed to engage students in group discussion of the material covered during the lab. The questions were primarily two-part (describe-explain) questions, and they were arranged in such an order that the students would first link concepts, then go in-depth using the concepts to explain the outcome of their experiment. The homework assignment text is available in Appendix C. The homework assignment was graded and returned to the students by the instructor within a week from the time it was turned in. Grades were compiled for each question.

The post-lab quiz consisted of two parts: a multiple choice portion (11 questions, presented in Blackboard) and a short-answer portion (11 questions, on paper). The questions are available in Appendix E. The quiz was administered in a 30-minute session in a controlled test environment; four session times on two consecutive nights were available for students to fit it into their schedule. The written questions were graded by two separate instructors, and any discrepancies were resolved through discussion. Questions were asked on the following subjects:

- General critical thinking/problem-solving ability (5 multiple choice, 1 written)
- General probabilistic/statistical reasoning (5 multiple choice)

- Specific subject conceptual/application knowledge (10 written, 1 multiple choice)

The exit survey was optional and consisted of six Likert-scale questions. The questions are available in Appendix D. Questions were asked on the general topics of:

- Group participation on homework
- Value obtained by participating
- Self-evaluation on exam results (results had not been released at the time)

Assessment of Statistical Reasoning and Critical Thinking

As additional data points, critical thinking and statistical reasoning skills were assessed independently for each student. This was accomplished via a series of multiple-choice questions related to each skill respectively. Students were asked an equivalent and parallel number of questions on pre- and post-test. Statistics questions were obtained by browsing relevant literature and extracting questions that would clearly demonstrate proper versus improper statistical and probabilistic reasoning (Konold, 1995; Kahneman & Tversky, 1972). Critical thinking questions were sourced from existing practice exams for standard critical thinking tests (for example, the Watson-Glaser, Cornell, and California critical skills tests) and were chosen to elicit students' logical thinking and critical reasoning skills. None of the questions required any subject-specific knowledge; each question had a single correct answer that could be determined precisely from the information provided in the question. Only questions which had been tested and proven not to be defective were used (Polanowski, 2013).

Regarding critical thinking assessment, there is widespread disagreement as to the appropriateness of using multiple-choice methods to assess critical thinking. Indeed, there is controversy over the very nature of critical thinking, causing some to argue that its evaluation is a futile endeavor (Lai, 2011; Norris, 1989; Douglas, 2012). While the authors reject this line of thinking, it must be admitted that the evaluation methods used in the present study are far from ideal. Many of the multiple-choice critical thinking skills assessment products present several pages of questions to students, whereas the present study used five questions per exam (for a total of 10 multiple-choice questions total). The rationale for this is that the objective of this research is not to establish rigid evidence for a change in critical thinking, but rather to identify trends that may exist which could possibly be the subject of future research.

Statistical Methods

Response data was first evaluated for normality and equality of variances. All response variables were checked against a chi-squared goodness-of-fit test before further analysis was conducted. Response variables were then screened against the independent variables for statistically-significant effects using a standard multi-factor analysis of variance (ANOVA) in the JMP program. Following effect screening, the ANOVA was then re-run with significant variables only to generate the final result tables.

4. RESULTS

Study One

This study had a total of 24 participants. One participant was excluded following completion of the study due to significantly below-average pre- and post-test scores (near zero for both). The control group contained 9 students and the experiment group contained 15 students. Due to the small sample sizes, an alpha level of 0.10 was considered to be significant.

As a note on the lower alpha level, Cohen observes that 80% confidence intervals make more sense for psychologically-based studies (Cohen, 1990); therefore, this study is justified in using any alpha less than 0.20. According to Rice & Harris (2005), studies of an exploratory nature (which this study admittedly is) have fewer controls in place, and standards for acceptance are appropriately lower. Their observations show that future, better-controlled studies will yield even stronger significance and effect sizes (*d*) (Rice & Harris, 2005).

Pre- to Post-Lab Results

Subjects completed both a pre- and post-lab examination. A paired-samples t-test (see Figure 1 and Table 3 below) showed a significant increase of 27.3% on average from pre-test to post-test. It should be noted that the tests were not identical; however, they were similar in wording, content, and format. The pre-test was shorter than the post-test and tested concept knowledge only; the post-test evaluated both concept and application knowledge. These results indicate that students generally achieved a positive learning outcome from the lab experiment.

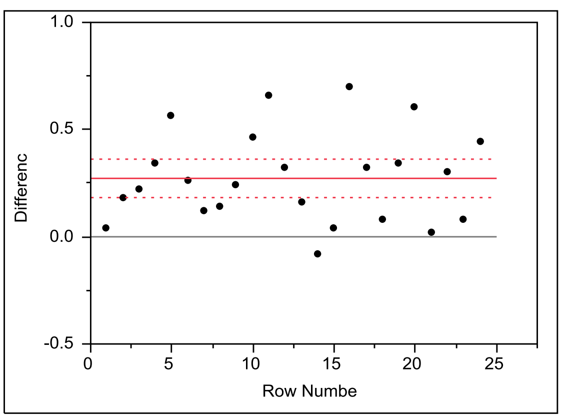


FIGURE 1: STUDY ONE PAIRED-SAMPLES T-TEST (PRE- TO POST-TEST)

TABLE 3: JMP PAIRED-SAMPLES RESULTS (PRE- TO POST-TEST)

Post-test Percent	**0.6025**	**t-Ratio**	**6.288108**
Pretest Percent	0.33	DF	23
Mean Difference	0.2725	Prob > |t|	<.0001*
Std Error	0.04334	Prob > t	<.0001*
Upper 95%	0.36215	Prob < t	1.0000
Lower 95%	0.18285		
N	24		
Correlation	0.38955		

Effect of Experiment vs. Control

Groups were tested for equality of variance. The O'Brien and Levene tests for variance equality were chosen due to their use of mean values and robustness to violation of normality (Levene, 1960; O'Brien 1978) due to small sample sizes. These returned inconclusive (p-Values ~0.50) on the pre-test data, thus it could not be shown that the equal variances was violated between the two groups for the pre-test scores. Subsequently, no significant difference in pre-test scores was found between groups.

For the post-test, a preliminary multi-factor ANOVA analysis showed an interaction between students who had taken the IE 248 class and the training method, likely due to the fact that they were already familiar with the control method and the experiment method was entirely new. Therefore, it was determined to exclude that group of students (5 participants) from this portion of the analysis. The post-test equal variances test were conclusive (p-Values of .03 for both the Levene and O'Brien tests). This indicates a significant difference in *standard deviation* between the control group and the experiment group, with the experiment group having a significantly lower standard deviation (see Table 4 below). Note that we assume equality of variance for the post-test score analysis because the underlying population is the same in both groups.

TABLE 4: STUDY ONE EQUALITY OF VARIANCES TESTS, POST-TEST SCORES

Level	Count	Std Dev
N	7	0.2110
Y	12	0.1090

Test	F Ratio	p-Value
O'Brien[.5]	5.5409	0.0309*
Levene	5.2660	0.0347*

A three-factor analysis of variance was performed on the post-test data. Controlling factors were pre-test score and interaction between treatment and pre-test score. Assuming equal variances (see explanation above), a one-tailed t-test showed a significant difference in post-test score between the two groups at the α-0.05 level ($p = .0417$). From a purely descriptive standpoint, the experiment group performed better than the control group by about 12.5% on average after controlling for other factors. These results are shown in Figure 2 and Table 5 below.

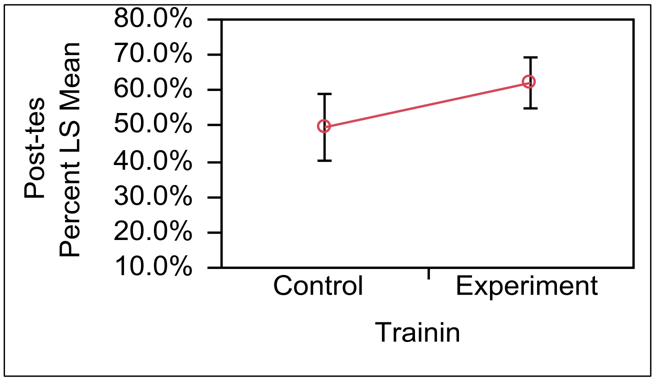


Figure 2: Study One, Comparison of Control and Experiment Groups

Table 5: Study One – t-Test of Difference between Control vs. Experiment

Item	Value	Item	Value
Difference	0.1253	t Ratio	1.8546
Std Err Dif	0.0676	DF	15
Upper CL Dif	0.2438	Prob > |t|	0.0834
Lower CL Dif	0.0069	Prob > t	0.0417
Confidence	0.90	Prob < t	0.9583

Table 6: ANOVA Full Factorial Analysis of Post-Test Scores

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0.4624	0.0706	6.55	<.0001*
Pretest Percent	0.3814	0.2431	1.57	0.1375
Training[Control]	**-0.0627**	0.0338	-1.85	**0.0834**
(Pretest Percent-0.25263)*Training[Control]	**0.4903**	0.2431	2.02	**0.0620**

Interaction

Figure 3 below displays the statistically-significant interaction between group and pre-test scores. As the chart shows, students in the control group performed roughy proportional to their pre-test score on the post-test (i.e. a low score on the pre-test implied a low score on the post-test). This was not the case for the experiment group, however; their scores did not depend at all on the pre-test score (i.e. no correlation between the two).

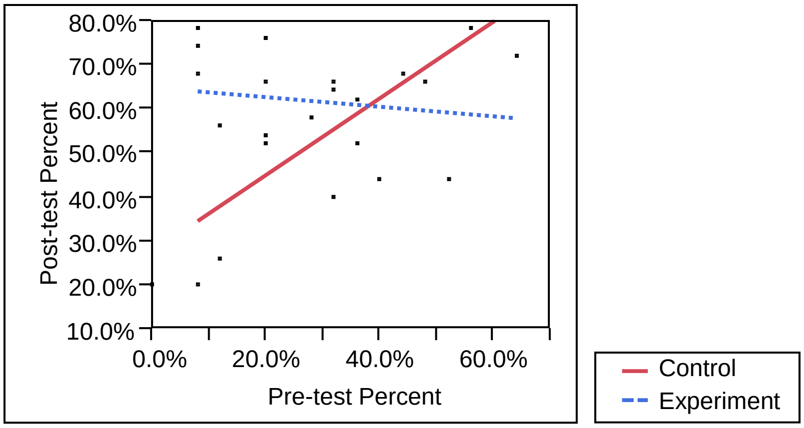


FIGURE 3: STUDY ONE - INTERACTION BETWEEN GROUP AND PRE-TEST SCORE

Effect Size

There was a large effect size, shown by Cohen's *d*, which focuses on the practical significance of results rather than statistical significance (Cohen, 1988 and 1990). Jacob Cohen, one of the leading experts in the fields of statistics and psychology, states that it is generally better to understand "how big" rather than "how statistically significant" the experimental effects are (Cohen, 1990). Since this study involves human cognitive abilities, and since related meta-analyses (Springer et al., 1999) use Cohen's *d* as a practical measure of significance across studies, its use here serves as demonstration of the overall impact from the results achieved.

In this study, the effect size (*d*) was computed to be 2.887 (see Equation 1 below). Another way to view this is that the participants in the experimental group scored nearly three standard deviations higher than the control group, on average, after controlling for the factors mentioned above. According to Cohen, a *d*-value greater than 0.8 indicates a large effect, "about as high as they come" (Cohen, 1988). Note that Cohen's value requires the use of pooled standard deviation, which is appropriate for the interpretation of this value (Cohen, 1988).

EQUATION 1: FORMULA AND CALCULATION FOR COHEN'S D (COHEN, 1988)

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s} \; where \; s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2}}$$

$$d = \frac{.6214 - .4961}{0.0434} \; ; d = 2.887$$

Study Two

This study had a total of 68 participants. One participant was excluded following the grading of the post-test, where the subject's answers indicated an inability to comprehend the exam questions (likely due to a language barrier). Participants were divided into two general groups based upon their prior coursework. Students who indicated that they had taken certain courses were placed into the *experienced* group, while others were considered *inexperienced*. There were 32 inexperienced students and 36 experienced students.

Participants were further divided into one of three treatment types: *control*, *experiment I, and experiment II*. Table 7 below contains a summary of the numbers of participants in each group.

TABLE 7: STUDY TWO - SUMMARY OF PARTICIPANT TYPES

	Control (0)	Experiment I (1)	Experiment II (2)
Experienced	9	13	14
Inexperienced	12	6	14

Results by Treatment Type

Figure 4 below shows the graphical depiction of the specific subject results by treatment type across both experienced and inexperienced students. As the 95% confidence interval error bars show, there were no statistically significant differences in average score among any of the treatments for either group before controlling for additional factors.

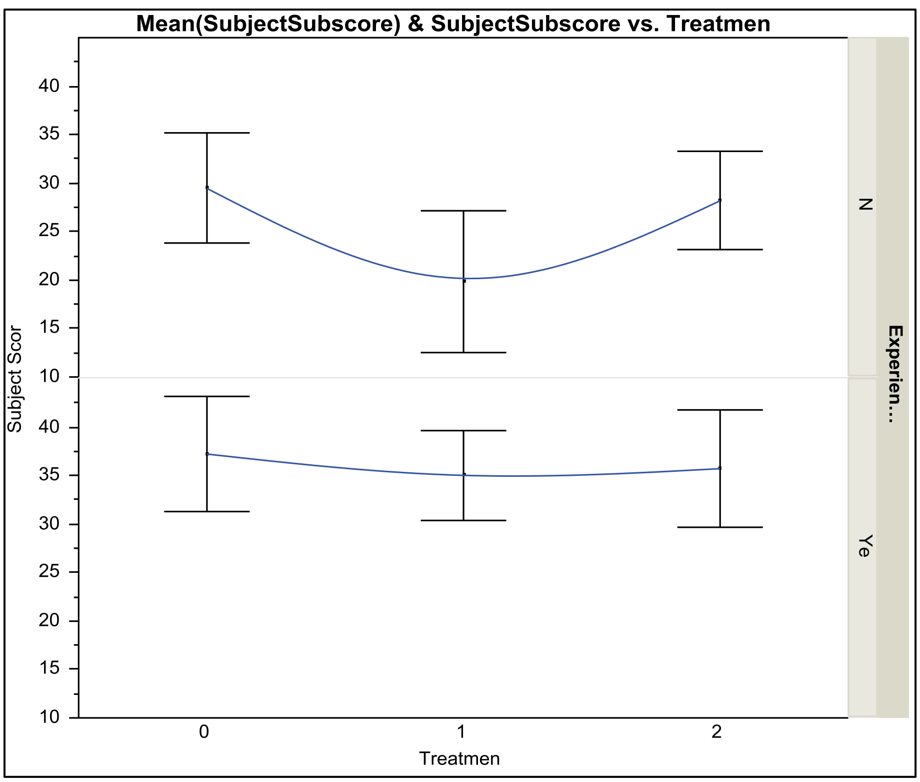


FIGURE 4: STUDY TWO - RESULTS BY TREATMENT AND EXPERIENCE LEVEL

A two-way ANOVA was performed on the factors believed to be correlated with the final exam score, including GPA (self-reported), ACT score (self-reported), pre-test score, CT composite score, and others. If a student did not report a GPA or ACT, the column average was used in place of the missing value. The analysis was performed independently for each experience level, and the results appear below in Table 8 and Table 9. No significant interactions were present.

As these data show, *GPA* and *Experiment I treatment* significantly affected the outcome for inexperienced students, while *critical thinking score* significantly affected outcome for experienced students. The model for inexperienced students predicted approximately 44% of the variation in scores (R^2 of .443) while the experienced student model predicted approximately 40% of the variation (R^2 of .403). In the tables below, the treatment factor estimate indicates the difference that particular treatment would add to (or subtract from) the overall score if the student experienced that particular treatment. Note that the treatment scores were computed from the reference score of Experiment II.

TABLE 8: STUDY TWO - ANOVA FACTORIAL ANALYSIS FOR INEXPERIENCED STUDENTS

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	2.144	6.550	0.33	0.7460
Treatment[Control]	2.077	1.892	1.10	0.2819
Treatment[Exp I]	**-5.851**	2.182	-2.68	**0.0123**
CT Composite	0.609	0.342	1.78	0.0861
GPA	**5.060**	1.818	2.78	**0.0097**

TABLE 9: STUDY TWO - ANOVA FACTORIAL ANALYSIS FOR EXPERIENCED STUDENTS

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	10.691	8.187	1.31	0.2012
Treatment[Control]	1.874	1.858	1.01	0.3210
Treatment[Exp I]	-1.048	1.681	-0.62	0.5374
CT Composite	**1.377**	0.346	3.98	**0.0004**
GPA	1.476	2.745	0.54	0.5947

5. DISCUSSION

Two separate but related studies were conducted to evaluate theories related to PBL in a real engineering classroom. Looking at both studies together, an interesting picture begins to emerge about how students learn in the PBL environment, as well as what factors are important to ensure a successful, consistent learning outcome.

Study One

The objective of *study one* was to evaluate the effect of a hard scaffold on student learning performance. *Study one* presented a relatively homogenous group of students with one of two learning setups. The first was the original course material, which did a rather poor job of linking the learning objectives for the lab session with the procedure followed in the lab. The second was a re-designed lab handout and procedure that (a) included additional abstract concepts and

(b) provided a much more structured procedure under which to conduct the lab experiment and analyze the data. It was hypothesized that students who had completed the experimental lab setup would have a higher performance on the post-test than their control counterparts because of the (ostensibly) better scaffold.

While the study had a relatively small sample size, there were two significant differences that emerged between the *control* group and the *experiment* group. The first was that students in the *experiment* group had a significantly lower standard deviation of scores on the post-test than their control counterparts (see Table 4 on page 15). Next, a significant increase in average score was seen (by about 12.5 percentage points) for the experiment group over the control group (Table 5 on page 16).

This result supports the original hypothesis that students would perform better when provided with a more-structured laboratory handout. The improved scaffold gave students in the *experiment* group a better ability to form connections between the abstract concepts presented in the lab and the concrete lab procedure. The lower standard deviation, while originally unexpected, could be explained by lower-ability students being helped more by the improved lab setup than higher-ability students. This explanation is consistent with Chi & VanLehn (2008) and Bloom (1984), who evaluated learning outcomes between weak and strong students. With targeted tutoring, strong students' ability did not increase significantly; however, weaker students' ability increased dramatically, which reduced the standard deviation between overall scores.

It is further apparent by examining the interaction between *training method* and *pre-test score* (see Figure 3 on page 17) that this was happening. In the *control* setup, students' score on the post-test was roughly proportional to their pre-test score; however, there was no correlation of pre-test score to post-test score with the *experiment* group. It is easy to see that students who did poorly on the pre-test did just as well on the post-test as higher-performing students. Furthermore, it is clear that the high-performers stayed high; the low-performers rose disproportionally to meet them. When taken in context, this result indicates that the scaffold in the lab is actually functioning in a similar way as a tutor, probably because of the structured way the questions are asked (as a tutor would ask them); students who cannot answer the questions will naturally be prompted to seek personal help from the instructor (or their group members), and will seek help as long as they are unable to understand the concepts.

Limitations of Study One

The results of *study one*, while promising, should be put into the context of their limitations. First, there was a relatively small sample size composed of both experienced and inexperienced students (with experience defined by the student having taken the original course that taught the material). The inclusion of these students into groups of inexperienced students may have affected the outcome of the learning process for all students. Secondly, and more importantly, there were two main differences between the scaffold presented to the experiment group and that presented to the control group: (a) the lab procedure was modified to be more coherent and (b) there was additional abstract background information provided within the packet. It is not apparent which of these changes (or the combination thereof) resulted in the ultimate outcome.

Study Two

The objective of this study was to evaluate the effect of providing students with abstract background information as part of a scaffold. It was conducted in such a way as to address the major limitations of *study one*, having a higher sample size, making a priori distinctions between experienced and inexperienced students, and changing only one variable (the amount of background information provided to students).

The results of this study were somewhat more interesting than the results of *study one*. First, it was expected that students in the control group would underperform both experimental groups. This was not the case at all – overall, students in all three groups performed equally well on the post-test. When the results are examined across experience levels, however, the picture changes somewhat. Inexperienced learners performed equally well in the *control* and *experiment II* condition (lecture after lab), but students in the *experiment I* condition (lecture before lab) performed significantly worse than their counterparts (-4.44 percent on average, p=.0131) on the specific subject-matter portion of the final exam. Experienced students performed equally-well regardless of treatment group.

The fact that inexperienced students performed significantly worse than students who received no lecture or had lecture after the lab is very intriguing. One possible explanation for this is that students who were not familiar with this material became confused about it during the lecture; since they had not performed the hands-on portion of the lab yet, there were no prior concrete experiences to which they could relate the information. It is not apparent from searching through relevant literature whether other studies have demonstrated a detrimental effect of a lecture, so this definitely should be a topic for future research.

Additional factors that explained variation in students' scores were *University GPA* and *critical thinking composite score*; however, this was not consistent across experienced vs. inexperienced students. For experienced students, the strongest predictor of their score was *critical thinking composite*, and *GPA* was not significant. For inexperienced students, the opposite was true. It's possible that older, more experienced students have stronger critical thinking skills, which were required on this test for nearly all of the questions; once again, however, it would be unwise to hypothesize too much about the causes of this without further study.

Remarks

Proponents of PBL tout its ability to foster strong learning outcomes even in the face of robust misconceptions in mental models or schema (Vidic, 2011; Riskowski et al., 2009; Mills & Treagust, 2003; Perrenet et al., 2000). The studies presented in the present paper support those claims. There are several published papers regarding the difficult nature of the materials presented to students in these studies (Konold, 1989; Konold et al. 1993; Konold, 1995; Konold and Pollatsek, 2002), and evidence exists that students arrive with pre-conceived mental models and dispositions which make it difficult for them to learn the material via traditional lecture methods (Konold, 1989; Konold et al. 1993). According to Dr. Stephen Vardeman, professor of Statistics and Industrial Engineering at Iowa State University, who has authored textbooks and instructed on this topic for more than three decades, students require the hands-on experience afforded by the laboratory; their participation in the collection and analysis of data is essential to achieve the learning outcome (personal communication, 2011; Vardeman, 1996). To address student misconceptions, Perrenet and colleagues (2000) say that "generally speaking, a subtle process of guided co-operative learning is necessary, one which shares some of the characteristics of PBL, but requires smaller groups and more structuring by the teacher." It should be noted that (a) it is uncertain which definition of PBL Perrenet is referring to and that (b) as previously described, the present studies were designed *a priori* exactly as described by Perrenet.

Structure of PBL Session is Important

What can be seen from the results of these studies is that the specific structure and content of the hands-on laboratory sessions is very important to achieving the end goal. Small, seemingly trivial changes to the PBL procedure (in this case, data collection and analysis procedures) result in large, significant performance gains for students. Going back to *study one*, there were several key modifications to the PBL scaffold that likely contributed to the successful learning outcome. First, students were given a logical process to follow when collecting data. It was made obvious, both through the data collection paperwork and the physical procedure, what data was being collected. Data collection was segmented into phases, and there was a clear purpose to each phase of the data collection process; this is in contrast to the original lab procedure, where students collected all of the data at the same time. Furthermore, while the calculations in the modified procedure were more complex, they used familiar terms like *mean* and *standard deviation*, whereas the original lab used control charting constants and other "magic numbers." The result was a greater ability for students to connect the analysis phase to their raw data. Finally, the questions students were asked in their homework assignment had a logical progression to them; the first question asked them to explain their learning objectives, while the last question asked them to draw logical conclusions from the results of their analysis. This is in contrast to the original lab procedure, which required students to answer a series of "discussion points," but those points were in no particular order nor was it obvious how they connected to the actual lab procedure.

Effectiveness of PBL

It was originally hypothesized that changes to the lab procedure would be important, but not crucial, to achieving the learning outcome. The lecture and supplementary text were believed to be equally important. Other studies have shown that students can learn significantly more when they receive abstract learning content form a lecture or text source (Schwerdt & Wuppermann, 2010). In fact, editorial comments by Kirschner et al. (2006) maintain that students need to be given all the information they might need up-front. The present studies, when taken together, cast some doubt onto these assertions; from the results, it is obvious that a properly-designed PBL session can and will trigger most, if not all, of the learning that takes place. Furthermore, the results show that it is possible for a lecture to be detrimental under some (not fully-understood) conditions, which directly contradicts an assertion made by Schwerdt and Wuppermann (2010) ("No support for detrimental effects of lecture-style teaching can be found …."). The results of this study could help explain why other studies (ref. Schmidt et al., 2011) found PBL methods to be superior to lecture-based methods, and why PBL proponents favor discovery methods as opposed to providing students with information *a priori*.

As to why students in the *experienced* group did not experience a similar effect, the literature offers a few clues. First, two studies by Herbert and Burt (2001, 2004) explored the change in knowledge structures that takes place as students become more familiar with a particular topic (i.e. *knowledge schematization*). In particular, they reference the idea of "knowing" something versus "remembering" something, implying a distinction between semantic memory and episodic memory, respectively (Herbert & Burt, 2004; Tulving, 1985). In general, it is believed that as students are exposed to a particular topic repeatedly, the knowledge structures become more organized and well-represented in semantic memory, such that the details of the episodic memories no longer are clear. At this point, students have "internalized" the knowledge (Herbert & Burt, 2004).

This theory of learning appears to be consistent with what was seen in these two studies, particularly in *study two*. Students in the *experienced* group not only outperformed students in the *inexperienced* group overall, they showed robustness to whatever deleterious effect caused the *inexperienced* students to falter when presented with a lecture before their lab. It also helps explain why PBL works, especially among students who are just being introduced to the material. According to Korthagen and Lagerwerf (1995), it is important for students to experience new, difficult-to-learn concepts in real-world situations. If episodic memory holds the key to learning these difficult topics, then it stands to reason that students learn most effectively when totally immersed in the problem, rather than being positioned as a casual bystander (i.e. the lecture hall).

Finally, the results here do not mean that providing students with optional lectures or supplementary information is futile. Some students felt that the lecture was beneficial to helping them understand the concepts. In addition, it is speculated that optional lectures can be an important tool in fostering understanding in the PBL classroom. Further research should be done into the reasons why students would choose to attend an optional lecture (i.e. what motivates their attendance), and whether or not this correlates with a true gain in learning outcome. On the other hand, neither instructor nor student time is "free" – meaning that if a lab can be designed to convey the topic adequately or superiorly, the instructor's time might best be spent teaching in other ways (perhaps by eliminating lectures altogether).

6. CONCLUSION

In this paper, the results of two studies are presented and analyzed. The picture that emerges when both studies are taken into account is that it is the PBL process itself, not the lecture or supplementary information provided, that determines the learning outcome. Specifically, a properly-designed PBL session, including relevant scaffolds suitable to the learner's experience and desired learning outcomes, can provide for an equally- or more-successful learning outcome than with lecture or lecture-plus-PBL. This is first established in *study one*, where an improved laboratory procedure gave students some structured guidance to their problem-solving steps, and confirmed in *study two*, where no significant difference was found in the performance of students who had received a lecture versus those who had not.

It has been suggested that students who learn science and engineering concepts experience a higher workload because this knowledge has a richer, more complex structure (Perrenet et al., 2000). One reason for this is that, in some cases, students must simultaneously learn and apply the material (Jong & Ferguson-Hessler, 1986). The material presented to students in this research is both wide in scope and deep in terms of understanding required to successfully apply it in new situations. Students require several "passes" at it before they get it right; this is typical of the type of material presented in engineering classes of all disciplines. Therefore, it stands to reason that classes that teach engineering and science topics are not the same as classes that teach other topics, and that methods which may work well for other classes may not work as well in engineering classes.

In terms of PBL, it is generally accepted that some level of structure is needed to support the process; it is the type and quantity of that structure that is debated (Kirschner, et al. 2006; Schmidt, et al. 2007; Hmelo-Silver, et al. 2004). The results of this study support that view, and further provide that a specific type of supporting structure is needed (i.e. not just any scaffold will do).

If the results of this study can be replicated in similar courses, the implications for engineering education are immense. The present culture of the University is to teach students primarily via the lecture (push-based) format. Courses like IE 248, where students spend a considerable amount of time in a hands-on laboratory, are relatively rare in the curriculum, and are even still accompanied by a lecture component. The present studies suggest that it is possible to transition a greater percentage of engineering education into PBL-based laboratories, and that this method has the potential to transform learning and improve it far beyond where it stands today.

7. REFERENCES

Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research*, *78*(4), 1102-1134.

Albert, J. H. (2003). College students' conceptions of probability. *The American Statistician*, *57*(1), 37-45.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of verbal learning and verbal behavior*, *22*(3), 261-295.

Anderson, T., Howe, C., Soden, R., Halliday, J., & Low, J. (2001). Peer interaction and the learning of critical thinking skills in further education students. *Instructional Science*, *29*(1), 1-32.

Astin, A. W. (1993). Engineering outcomes. *ASEE Prism*, *3*(1), 27-30.

Bailin, S., & Siegel, H. (2003). Critical thinking. In N. Blake, P. Smeyers, R. Smith, & P. Standish (Eds.), *The Blackwell guide to the philosophy of education* (pp. 181–193). Oxford, UK: Blackwell.

Bao, L., & Redish, E. F. (2002). Understanding probabilistic interpretations of physical systems: A prerequisite to learning quantum physics. *American Journal of Physics*, *70*, 210.

Barrows HS, Tamblyn RM. 1980. *Problem-based learning: An approach to Medical Education*. New York: Springer.

Beaufait, F. W. (1991, September). Engineering education needs surgery. In *Frontiers in Education Conference, 1991. Twenty-First Annual Conference.'Engineering Education in a New World Order.'Proceedings.* (pp. 519-522). IEEE.

Belland, B. R., Glazewski, K. D., & Richardson, J. C. (2007). A scaffolding framework to support the construction of evidence-based arguments among middle school students. *Association for Educational Communications and Technology, 56,* 401-422.

Belland, B., French, B., & Ertmer, P. (2008).Validity and problem-based learning research: A review of instruments used to assess intended learning outcomes. Paper presented at the Annual Meeting of the American Educational Research Association, New York, March 24 – 28, 2008.

Besterfield-Sacre, M., Atman, C. J., & Shuman, L. J. (1997). Characteristics of freshman engineering students: Models for determining student attrition in engineering. *Journal of Engineering Education*, *86*(2), 139-149.

Besterfield-Sacre, M.E., Atman, C.J., and Shuman, L.J. (1995, June). Freshman Attitudes: Changes during the First Year. *1995 ASEE Annual Meeting Proceedings*, ASEE, pp. 157-162.

Bloom (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 13, 4-16.

Board of Engineering Education - National Research Council (1992). Improving Retention in Undergraduate Engineering Education. *Issues in Engineering Education: A Bulletin Addressing Culture Change in Engineering Education*, vol. 1, no. 1

Brown, N. W., & Cross Jr, E. J. (1993). Retention in Engineering and Personality. *Educational and psychological measurement*, *53*(3), 661-71.

Butler, A. C. (2010). Repeated Testing Produces Superior Transfer of Learning Relative to Repeated Studying, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1118-1133.

Chi, M. T. H. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *The Journal of the Learning Sciences*, *14*(2), 161-199.

Chi, M.T.H., & VanLehn, K. (2008). Eliminating the gap between the high and low students through meta-cognitive strategy instruction. *Learning Research and Development Center, University of Pittsburgh, PA*, 603-618.

Choo, S. S. Y., Rotgans, J. I., & Yew, E. H. J. (2010). Effect of worksheet scaffolds on student learning in problem-based learning. *Advancements in Health Sciences Education*, *16*, 517-528.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* (Second edition). Hillsdale, NJ: Erlbaum.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*(12), 1304-1312.

Colliver, J. A. (2000). Effectiveness of problem-based learning curricula: Research and theory. *Academic Medicine*, *75*(3).

Corter, J. E., Esche, S. K., Chassapis, C., Ma, J., & Nickerson, J. V. (2011). Process and learning outcomes from remotely-operated, simulated, and hands-on student laboratories. *Computers & Education*, *57*(3), 2054-2067.

D'Argembeau, A., & Mathy, A. (2011). Tracking the construction of episodic future thoughts. *Journal of Experimental Psychology-General*, *140*(2), 258.

Derry, S. J. (1996). Cognitive schema theory in the constructivist debate. *Educational Psychologist*, *31*(3-4), 163-174.

diSessa, A. A. (1982). Unlearning Aristotelian physics: A study of knowledge-based learning. *Cognitive Science*, *6*, 37-75.

Douglas, E. P. (2012). Defining and Measuring Critical Thinking in Engineering. *Procedia-Social and Behavioral Sciences*, *56*, 153-159.

Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education*, 32, 268 – 291.

Ertmer, P. A., & Simons, K. D. (2006). Jumping the PBL implementation hurdle: Supporting the efforts of K–12 teachers. *Interdisciplinary Journal of Problem-based Learning*, *1*(1), 5.

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Research findings and recommendations*. Newark, DE: American Philosophical Association. (ERIC Document Reproduction Service No. ED315423)

Feder, P. I. (1984). Glossary and Tables for Statistical Quality Control.*Technometrics*, *26*(1), 64-64.

Felder, R. M. (1995). A longitudinal study of engineering student performance and retention. IV. Instructional Methods. *Journal of Engineering Education*,*84*(4), 361-367.

Felder, R. M., Felder, G. N., & Dietz, E. J. (1998). A Longitudinal Study of Engineering Student Performance and Retention. V. Comparisons with Traditionally-Taught Students. *Journal of Engineering Education*, *87*(4), 469-480.

Felder, R. M., Forrest, K. D., Baker-Ward, L., Dietz, E. J., & Mohr, P. H. (1993). A longitudinal study of engineering student performance and retention: I. Success and failure in the introductory course. *Journal of Engineering Education*, *82*(1), 15-21.

Frick, R. W. (1998). Interpreting statistical testing: Process and propensity, not population and random sampling. *Behavior Research Methods, Instruments, & Computers*, *30*(3), 527-535.

Gijbels, D., Dochy, F., Van den Bossche, P., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research*, *75*(27).

Gürbüz, R., & Birgin, O. (2012). The effect of computer-assisted teaching on remedying misconceptions: The case of the subject "probability". *Computers & Education*, *58*(3), 931-941.

Gürbüz, R., Çatlıoğlu, H., Birgin, O., & Erdem, E. (2010). An investigation of fifth grade students' conceptual development of probability concepts based on activity based instruction: A quasi-experimental study. *Educational Sciences: Theory & Practice*, *10*(2), 1053-1068.

Haag, S., Hubele, N., Garcia, A., & McBeath, K. (2007). Engineering undergraduate attrition and contributing factors. *International Journal of Engineering Education*, *23*(5), 929-940.

Heckel, R. W. (1996). Engineering Freshman Enrollments: Critical and Non-critical Factors. *Journal of Engineering Education*, *85*(1), 15-21.

Herbert, D. M. B., & Burt, J. S. (2004). Problem-based learning: Where are we now? *Applied Cognitive Psychology*, *18*(1), 77-88.

Herbert, D., & Burt, J. S. (2001). Memory awareness and schematization: Learning in the university context. *Applied cognitive psychology*, *15*(6), 617-637.

Hermond, D. (1995). Measuring the retention strategies of a minority engineering program: A service quality perspective. *Journal of Engineering Education*, *84*(4), 395-400.

Hewitt, N. M., & Seymour, E. (1992). A long, discouraging climb. ASEE Prism, 1(6), 24-28.

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn?. *Educational Psychology Review*, *16*(3), 235-266.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 99-107.

Hundhausen, C., Agarwal, P., Zollars, R., & Carter, A. (2011). The Design and Experimental Evaluation of a Scaffolded Software Environment to Improve Engineering Students' Disciplinary Problem-Solving Skills. *Journal of Engineering Education*, *100*(3), 574-603.

Jong, T., & Ferguson-Hessler, M. G. (1986). Cognitive structures of good and poor novice problem solvers in physics. *Journal of educational psychology*,*78*(4), 279-288.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive psychology*, *3*(3), 430-454.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75-86.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning. *Cognitive science*, *36*(5), 757-798.

Konold, C. (1989). Informal conceptions of probability. *Cognition and instruction*,*6*(1), 59-98.

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education*, *3*(1), 1-9.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, 259-289.

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics education*, 392-414.

Koretsky, M. & Brooks, B. (2011). Comparison of Student Responses to Easy and Difficult Thermodynamics Conceptual Questions during Peer Instruction. *International Journal of Engineering Education* Vol. 27, No. 4, pp. 897–908.

Korthagen, F., & Lagerwerf, B. (1995). Levels in learning. Journal of Research in Science Teaching, 32, 1011–1038.

Lai, E. R. (2011). Critical thinking: A literature review. *Pearson's Research Reports*, *6*.

Land, S. M., & Zembal-Saul, C. (2003). Scaffolding reflection and articulation of scientific explanations in a data-rich, project-based learning environment: An investigation of progress portfolio. *Educational Technology Research and Development*, *51*(4), 65-84.

Lee, H. S., & Anderson, J. R. (2013). Student Learning: What Has Instruction Got to Do With It?. *Annual Review of Psychology*, *64*, 445-469.

Lee, V. E., & Burkam, D. T. (1996). Gender differences in middle grade science achievement: Subject domain, ability level, and course emphasis. *Science Education*, *80*(6), 613-650.

Lehrer, R., & Schauble, L. (2006). Scientific thinking and scientific literacy: Supporting development in learning contexts. In K. A. Renninger & I. Sigel (Eds.), *Handbook of child psychology* (Vol. 4, 6th ed.). Hoboken, NJ:Wiley.

Levene, H. (1960). Robust testes for equality of variances. In *Contributions to Probability and Statistics* (I. Olkin, ed.) 278–292. Stanford Univ. Press, Palo Alto, CA. MR0120709

Liu, T. C., Lin, Y. C., & Tsai, C. C. (2009). Identifying Senior High School Students' Misconceptions About Statistical Correlation, And Their Possible Causes: An Exploratory Study Using Concept Mapping With Interviews. *International Journal of Science and Mathematics Education*, *7*(4), 791-820.

Lord, S., & Camacho, M. M. (2007, October). *Effective teaching practices: Preliminary analysis of engineering educators* . Paper presented at 37th ASEE/IEEE frontiers in education conference, Milwaukee, WI.

Mason, M. (2007). Critical thinking and learning. *Educational Philosophy and Theory*, *39*(4), 339-349.

Mateo-Sanz, J., Solanas, A., Puigjaner, D., & Olivé, C. (2010). Refining Statistical Problems: A Hybrid Problem-Based Learning Methodology to Improve Students' Motivation. *Int. J. Eng Ed.* Vol. 26, No. 3, pp. 667–680, 2010

Mayer, R., Moeller, B., Kaliwata, V., Zweber, B., Stone, R., & Frank, M. (2012, September). Educating Engineering Undergraduates: Effects of Scaffolding in a Problem-Based Learning Environment. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, No. 1, pp. 2507-2511). SAGE Publications.

McInerney, M. J., & Fink, L. D. (2003). Team-based learning enhances long-term retention and critical thinking in an undergraduate microbial physiology course. *Journal of Microbiology & Biology Education*, *4*(1).

Meade, J. (1991, Sept.). The Missing Piece. *ASEE Prism*, pp. 19-22.

Mills, J. E., & Treagust, D. F. (2003). Engineering education—Is problem-based or project-based learning the answer?. *Australasian Journal of Engineering Education*, *3*, 2-16.

Mina, M., Omidvar, I., & Knott, K. (2003). Learning to think critically to solve engineering problems: Revisiting John Dewey's ideas for evaluating engineering education. *Retrieved April*, 2, 2013.

Niederhauser, D.S., Reynolds, R.E., Salmen, D.J., & Skolmoski, P. (2000). The Influence of Cognitive Load on Learning from Hypertext. *J. Educational Computing Research*, Vol. 23(3) 237–255.

Norman, Donald A. (2002). *The Design of Everyday Things*. (Reprint) New York: Basic Books.

Norris, S. P. (1989). Can we test validly for critical thinking?. *Educational Researcher*, *18*(9), 21-26.

Nusca, A. (2010). DARPA: 'Significant decline' in U.S. science, tech degrees 'harming national security'. Retrieved 11 April 2013 from http://www.smartplanet.com/blog/smart-takes/darpa-8216significant-decline-in-us-science-tech-degrees-8216harming-national-security/3412

O'Brien, R.G. (1979). A general ANOVA method for robust test of additive models for variance. *Journal of the American Statistical Association*, 74, 877–880.

O'Sullivan, C. S., & Durso, F. T. (1984). Effect of schema-incongruent information on memory for stereotypical attributes. *Journal of Personality and Social Psychology*, *47*(1), 55-70.

Paas, F. G., & Van Merriënboer, J. J. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*,*6*(4), 351-371.

Paz Penagos, H. (2011). How can metacognition be developed through problem-solving in higher education?. *Ingeniería e Investigación*, *31*(1), 213-223.

Pea, R. D. (1993). Practices of distributed intelligence and designs for education. In Salomon, G., and Perkins, D. (eds.), *Distributed Cognitions: Psychological and Educational Considerations*, Cambridge University Press, New York, pp. 47–87.

Pea, R.D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. Journal of the Learning Sciences 13: 423–451.

Pease, M. A., & Kuhn, D. (2010). Experimental analysis of the effective components of problem-based learning. *Wiley Periodicals, Inc.*

Perrenet, J. C., Bouhuijs, P. A. J., & Smits, J. G. M. M. (2000). The suitability of problem-based learning for engineering education: theory and practice.*Teaching in higher education*, *5*(3), 345-358.

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *The Journal of the Learning Sciences*, *13*(3), 273-304.

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's *d*, and r. *Law and Human Behavior*, *29*(5).

Riskowski, J. L., Todd, C. D., Wee, B., Dark, M., & Harbor, J. (2009). Exploring the effectiveness of an interdisciplinary water resources engineering module in an eighth grade science course. *International journal of engineering education*,*25*(1), 181.

Rogers, A. & Spitzmueller, C. (2009). Individualism–collectivism and the role of goal orientation in organizational training. *International Journal of Training and Development* 13:3

Roschelle, J. (1996). *Designing for cognitive communication: Epistemic fidelity or mediating collaborative inquiry* (pp. 13-25). Taylor & Francis, London.

Salomon, G. (1993). No distribution without individual cognition: A dynamic interactional view. In Salomon, G., and Perkins, D. (eds.), *Distributed Cognitions: Psychological and Educational Considerations*, Cambridge University Press, New York, pp. 111–138.

Savery, J. R., & Duffy, T. M. (1995). Problem based learning: An instructional model and its constructivist framework. *EDUCATIONAL TECHNOLOGY-SADDLE BROOK NJ-*, *35*, 31-31.

Saye, J. W., & Brush, T. (2002). Scaffolding critical reasoning about history and social issues in multimedia-supported learning environments. *Educational Technology Research and Development*, *50*(3), 77-96.

Schmidt, H. G., Loyens, S. M., Van Gog, T., & Paas, F. (2007). Problem-based learning is compatible with human cognitive architecture: Commentary on Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*(2), 91-97.

Schmidt, H. G., Rotgans, J. I., & Yew, E. H. J. (2011). The process of problem-based learning: what works and why. *Medical Education*, *45*, 792-806.

Schwerdt, G., & Wuppermann, A. C. (2011). Is traditional teaching really all that bad? A within-student between-subject approach. *Economics of Education Review*, *30*(2), 365-379.

Seymour, E., & Hewitt, N. M. (1994). *Talking about leaving: factors contributing to high attrition rates among science, mathematics & engineering undergraduate majors: final report to the Alfred P. Sloan Foundation on an ethnographic inquiry at seven institutions*. Ethnography and Assessment Research, Bureau of Sociological Research, University of Colorado.

Sheffler, I. (1973). *Reason and teaching.* Indianapolis, IN: Hackett.

Simons, K. D., & Klein, J. D. (2007). The impact of scaffolding on student achievement levels in a problem-based learning environment. *Instructional Science*, *35*, 41-72.

Smith, Lauren (2007). ACT scores edge up in 2007 but suggest that many students are unprepared for college-level work. *The Chronicle of Higher Education*. Retrieved 11 April 2013 from http://www.csun.edu/pubrels/clips/Aug07/08-15-07A.pdf

Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, *69*(21).

Streveler, R. A., Litzinger, T. A., Miller, R. L., & Steif, P. S. (2008). Learning conceptual knowledge in the engineering sciences: Overview and future research directions. *Journal of Engineering Education*.

Sutcliffe, A. G., & Maiden, N. A. M. (1992). Analysing the novice analyst: cognitive models in software engineering. *International Journal of Man-Machine Studies*, *36*(5), 719-740.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, *4*(4), 295-312.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn.*Cognition and instruction*, *12*(3), 185-233.

Sweller, J., Kirschner, P. A., & Clark, R. E. (2007). Why minimally guided teaching techniques do not work: A reply to commentaries. *Educational Psychologist*, *42*(2), 115-121.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, *10*(3), 251-296.

Taylor, D., & Miflin, B. (2008). Problem-based learning: Where are we now? *Medical Teacher*, *30*(8), 742-763.

Theonas, G., Hobbs, D., & Rigas, D. (2008). The Effect of Facial Expressions on Students in Virtual Educational Environments. *International Journal of Social Sciences*, *2*(42-49).

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology* 26:1-12.

Vidic, A. (2011). Impact of Problem-based Statistics Course in Engineering on Students' Problem-Solving Skills. *International Journal of Engineering Education* Vol. 27, No. 4, pp. 885–896.

Vosniadou, S., & Brewer, W. F. (1989). *The concept of the earth's shape: A study of conceptual change in childhood*. University of Illinois at Urbana-Champaign.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223-248.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*, 89-100.

Yadav, A., Subedi, D., Lundeberg, M., & Bunting, C. (2011). Problem-based Learning: Influence on Students' Learning in an Electrical Engineering Course. *Journal of Engineering Education* Vol. 100, No. 2, pp. 253-280.

APPENDIX A

Study One Materials

This appendix contains the lab materials presented to students in Study One.

Pages 33 through 38 contain the original lab handout, data collection tables, and formulas given to students in the control group.

Pages 39 through 45 contain the modified lab handout and data collection tables provided to students in the experiment group. Note that students used a total of four copies of each table.

Pages 46 and 47 contain the laboratory slides presented to students in both groups as a lecture before the laboratory began.

Measurement Error Experiment

Process Description: There is error associated with every measurement which is taken. If the magnitude of the error is small, relative to the variability of the entity being measured, then the measurement system is acceptable. Measurement error less than 10% is preferred, but values as large as 30% are acceptable for special circumstances.

Lab Objectives: To investigate the sources of measurement errors associated with commonly used measurement instruments.

Lab Procedure: Choose a feature to be measured, determine the measurement technique, and review the technique with all of the inspectors. Ten pieces of the same object are required, and should be numbered so they can be identified. The test requires two or three inspectors, and one person to administer the test. The administrator should randomly choose the order in which the parts are to be measured, and hand the part to the inspector. Each inspector needs to measure each part two or three times. The repeated measurements cannot be done immediately following each other. The administrator will record the measurements as the inspector announces the value. The administrator must strive to eliminate any bias in the test.

Repeat the procedure with a different feature, inspectors, measurement instrument, and administrator.

For this application, assume that the measurement error needs to be less than 20% to be considered acceptable.

Develop a spreadsheet in Excel to make the calculations. Include an electronic copy of the spreadsheet, as instructed by the TA. Make sure that the spreadsheet will be easy for someone else to understand and use.

Partial List of Items to Include in Your Report:

1. Comment on the magnitude of the components of the measurement error (equipment and appraiser).

2. What is the measurement error as a percentage of total variability? What is the part variation as a percentage of total variability? Comment on the use of these two different measures of error.

3. How could the measurement error be reduced?

4. What did the administrator do to eliminate any bias? What else could have been done? Why is this important?

Table 1: Data collection sheet

Gauge Repeatability Data Collection Sheet

Appraiser /Trial #	Aluminum Part (measured with caliper)										
	1	2	3	4	5	6	7	8	9	10	
A/1											
A/2											
Average											$\overline{X}_a =$
Range											$\overline{R}_a =$

Appraiser /Trial #	Aluminum Part (measured with caliper)										
	1	2	3	4	5	6	7	8	9	10	
B/1											
B/2											
Average											$\overline{X}_b =$
Range											$\overline{R}_b =$

$R_p = Max\overline{R} - Min\overline{R} =$
$\overline{\overline{R}} = \left(\overline{R}_a + \overline{R}_b\right)/2 =$
$\overline{X}_{DIFF} = Max\overline{X} - Min\overline{X} =$
$*UCL_R = \overline{\overline{R}} \times 3.27 =$

*UCL_R represents the limit of the individual R's. Circle those that are beyond this limit. Identify the cause and correct. Repeat these readings using the same appraiser and tool as originally used or discard values and re-average and recompute $\overline{\overline{R}}$ and the limiting value from the remaining observations.

Gage Repeatability Data Collection Sheet

Appraiser /Trial #	Aluminum Part (measured with micrometer)										
	1	2	3	4	5	6	7	8	9	10	
A/1											
A/2											
Average											$\overline{X}_a =$
Range											$\overline{R}_a =$

Appraiser /Trial #	Aluminum Part (measured with micrometer)										
	1	2	3	4	5	6	7	8	9	10	
B/1											
B/2											
Average											$\overline{X}_b =$
Range											$\overline{R}_b =$

$R_p = Max\overline{R} - Min\overline{R} =$
$\overline{\overline{R}} = \left(\overline{R}_a + \overline{R}_b\right)/2 =$
$\overline{X}_{DIFF} = Max\overline{X} - Min\overline{X} =$
$*UCL_R = \overline{\overline{R}} \times 3.27 =$

*UCL_R represents the limit of the individual R's. Circle those that are beyond this limit. Identify the cause and correct. Repeat these readings using the same appraiser and tool as originally used or discard values and re-average and recompute $\overline{\overline{R}}$ and the limiting value from the remaining observations.

Gage Repeatability Data Collection Sheet

Appraiser /Trial #	Plastic Part (measured with caliper)										
	1	2	3	4	5	6	7	8	9	10	
A/1											
A/2											
Average											$\overline{X}_a =$
Range											$\overline{R}_a =$

Appraiser /Trial #	Plastic Part (measured with caliper)										
	1	2	3	4	5	6	7	8	9	10	
B/1											
B/2											
Average											$\overline{X}_b =$
Range											$\overline{R}_b =$

$R_p = Max\overline{R} - Min\overline{R} =$
$\overline{\overline{R}} = \left(\overline{R}_a + \overline{R}_b\right)/2 =$
$\overline{X}_{DIFF} = Max\overline{X} - Min\overline{X} =$
$*UCL_R = \overline{\overline{R}} \times 3.27 =$

*UCL_R represents the limit of the individual R's. Circle those that are beyond this limit. Identify the cause and correct. Repeat these readings using the same appraiser and tool as originally used or discard values and re-average and recompute $\overline{\overline{R}}$ and the limiting value from the remaining observations.

Gage Repeatability Data Collection Sheet

Appraiser /Trial #	Plastic Part (measured with micrometer)										
	1	2	3	4	5	6	7	8	9	10	
A/1											
A/2											
Average											$\overline{X}_a =$
Range											$\overline{R}_a =$

Appraiser /Trial #	Plastic Part (measured with micrometer)										
	1	2	3	4	5	6	7	8	9	10	
B/1											
B/2											
Average											$\overline{X}_b =$
Range											$\overline{R}_b =$

$R_p = Max\overline{R} - Min\overline{R} =$
$\overline{\overline{R}} = \left(\overline{R}_a + \overline{R}_b\right)/2 =$
$\overline{X}_{DIFF} = Max\overline{X} - Min\overline{X} =$
$*UCL_R = \overline{\overline{R}} \times 3.27 =$

*UCL_R represents the limit of the individual R's. Circle those that are beyond this limit. Identify the cause and correct. Repeat these readings using the same appraiser and tool as originally used or discard values and re-average and recompute $\overline{\overline{R}}$ and the limiting value from the remaining observations.

Table 2: Gage Repeatability and Reproducibility Calculation Sheet

From Table 1: $\bar{\bar{R}}$ = $\bar{X}_{DIFF}$ = R_p=

Measurement Unit Analysis	%Total Variation (TV)
Repeatability – Equipment Variation (EV) EV $= \bar{\bar{R}} \times K_1$ = _____x_____ = ________	% EV = 100 [EV/TV] = 100 [______/______] = ________%
Reproducibility – Appraiser Variation (AV) AV $= \sqrt{\left(\bar{X}_{DIFF} \times K_2\right)^2 - \left(EV^2 / nr\right)}$ $= \sqrt{(\_\_\_\_ \times \_\_\_\_)^2 - (\_\_\_\_^2 / \_\_\_\_ \times \_\_\_\_)}$ = _________	% AV = 100 [AV/TV] = 100 [____/_____] = ________% n = number of parts r = number of trials
Repeatability & Reproducibility (R&R) R&R $= \sqrt{\left(EV^2 + AV^2\right)}$ $= \sqrt{\left(\_\_\_\_\_^2 + \_\_\_\_\_^2\right)}$ = _________	% R&R = 100 [R&R/TV] = 100 [_____/_____] = ________%
Part Variation (PV) PV $= R_p \times K_3$ = ____ x ____ = ______	% PV = 100 [PV/TV] = 100 [_____/_____] = ________%
Total Variation (TV) TV $= \sqrt{R\&R^2 + PV^2}$ $= \sqrt{(\_\_\_\_)^2 + (\_\_\_\_\_)^2}$ = _______	

# of Trials	K_1
2	4.56
3	3.05

Appraisers	2	3
K_2	3.65	2.70

Parts	K_3
2	3.65
3	2.70
4	2.30
5	2.08
6	1.93
7	1.82
8	1.74
9	1.67
10	1.62

All calculations are based upon predicting 5.15 Sigma (99.0% of the area under the normal distribution curve).
K_1 is 5.15/d^*_2 where d^*_2 is dependent on the number of trials (m) and the number of parts times the number of appraisers (g) which is assumed to be greater than 15.
AV – If a negative value is calculated under the square root sign, the appraiser variation (AV) defaults to zero (0)
K_2 is 5.15/ d^*_2 where d^*_2 is dependent on the number of appraisers (m) and (g) is 1, since there is only one range calculation.
K_3 is 5.15/d^*_2 where d^*_2 is dependent on the number of parts (m) and (g) is 1, since there is only one range calculation. d^*_2 is obtained from Table D_3 "Quality Control and Industrial Statistics," A.J. Duncan.

Sources of Variability

IE 248 Lab Experiment 1

Lab Objectives

To investigate sources of variability and learn how to separate different components of variability using statistics and proper experiment design.

Introduction

Variability is defined as "the quality of being uneven and lacking uniformity[1]." In other words, variability is an inevitable part of life. It is impossible to achieve perfect repetition, whether you are throwing a football or producing a precision part for the International Space Station. There is always some degree of variability from one try to the next; it may be very small, as with the precision part for the ISS, or very large, as with your football throw. Variability in the world is the main reason that engineers exist.

Everything that can physically affect a production process introduces a portion of variability. As engineers, it is our job to understand and attempt to control that variability. Proper control of variability allows us to reduce uncertainty, so we can produce cars that start, roads that are smooth, and planes that don't crash. Too much variability leads to great uncertainty, and when it is left unchecked, you often end up hearing about it in the news. The question then is "How do we account for all of the sources of variability?"

Understanding Variability – Mean and Standard Deviation

To control variability, we must first understand what it looks like. In statistics, there are two essential ways of describing a set of data. The first is mean ($\bar{x}$), which tells us where the center of the data is. The second is standard deviation (s), which indicates the spread of the data. Each physical source of variability contributes to variability in our data sets.

Figure 1 illustrates the concepts of mean and standard deviation using several normal distribution plots. Notice how the curves get wider with increasing standard deviation, and how they shift with a change in the mean. Mean and deviation are two different concepts, and they both must be considered when accounting for variability in our data.

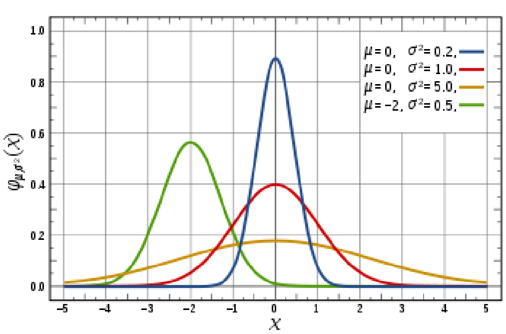


Figure 1: Standard Normal Distribution with Various Means and Standard Deviations

[1] http://www.thefreedictionary.com/variability

In a measurement system, another way of looking at mean and standard deviation is by understanding accuracy and precision. A good way of thinking about accuracy and precision is to look at a dart board, as in Figure 2. Accuracy refers to the direction that one is "aimed," while precision refers to the degree with which one is able to hit the target consistently. In this figure, it is clear that accuracy and precision are two completely separate ideas.

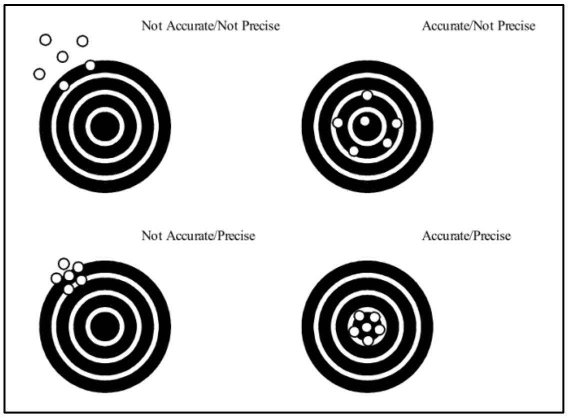


FIGURE 2: ACCURACY VS. PRECISION USING TARGETS

SOURCES OF VARIABILITY

As mentioned previously, every physical component of a process adds its own portion of variability. Some of these sources of variability are small and can be ignored, while others are quite large and must be reduced or eliminated. As an example, let's examine a standard milling process for a block of aluminum. The aluminum must be clamped into a horizontal mill, where a face mill cutter then shaves off approximately .03" from the top of the block to cut it down to the correct height. The process is supposed to produce blocks that are 1.45" +/- .005", and this dimension is verified at the end of the process by using a digital micrometer.

Where is the variability in this process? First, not all of the aluminum blocks start out the exact same size. Differences in initial height cause slightly different amounts of material to be shaved off with the face mill. When more metal is cut, the blocks heat up and expand slightly, causing the dimensions to be just a little off. Next, the mill has to be set to the correct depth each time a block is placed on it. The height is reported by a digital gage which has a small (probably negligible) amount of variability associated with each reading. The block is clamped into a vice each time, but there is still a small amount of variability in its lateral and vertical position. Changes in temperature, humidity, and pressure from day to day can cause the material to change shape or to react differently to the cutting process. Finally, there is a small amount of variability in the measuring process at the end.

All of this variability can be summed up into two main components: process variability and measurement variability. Process variability describes the differences from part to part in the output of a process, while measurement variability applies to the techniques used to measure those parts. Equation 1 illustrates this concept.

$$Total\ Variability = \sigma_{total} = \sqrt{\sigma_{process}^2 + \sigma_{measurement}^2}$$

EQUATION 1: COMPONENTS OF TOTAL VARIABILITY

Measurement Variability

When variability is introduced in the measuring process, it can be particularly problematic because we can only "see" process variability through measurements. It is very important that measurement variability be kept low (less than 10% of the overall tolerance window); otherwise, engineers will not be able to tell the difference between process variability and measurement variability. This is one of the main goals of the field of *metrology* – the science of measuring.

In this lab, we are going to take an in-depth look at measurement variability. As with process variability, there are multiple sources of measurement variability. These sources include the device itself, the operator, the environmental conditions (if they change over time while measurements are taken), and the measurement technique. Some of those, such as environmental conditions, can be ignored, especially for measurements taken on the same day at the same time. Others, such as operator variability, can be substantial enough to ruin an otherwise-good measurement process.

We classify the variability of a measurement system as falling under two main components:

- **Repeatability Variation**, which is the variability associated with the measurements from a *single operator* using a *single device* to measure a *single part* multiple times
- **Reproducibility Variation**, which is the variability associated in the measurements from *multiple operators* using a *single device* to measure a *single part* multiple times.

It helps to think about *repeatability variation* as something that is associated with a particular device and operator combination, while *reproducibility variation* is associated with the differences in measurement techniques between operators[2]. These two components add together to form measurement variability (see Equation 2).

$$Measurement\ Variability = \sigma_{measurement} = \sqrt{\sigma^2_{repeatability} + \sigma^2_{reproducibility}}$$

Equation 2: Components of Measurement Variability

Statistics and Variability

The field of statistics exists because variability is an unavoidable component of everything. Statistics is concerned with the proper methods of collecting, summarizing, and interpreting data, all within the context of real-world variability. Using statistics, it is possible to separate out the components of variability. While more in-depth statistical concepts will be presented in later courses (particularly IE 361), it is useful to gain a basic understanding of the statistical concepts that will be used in the collection and analysis of data. Those concepts will be highlighted as the lab procedure unfolds.

[2] Vardeman and Morris, IE 361 Module 3, 2011

Procedure

You will be conducting this procedure in groups of no more than five members. The procedure for this lab is divided into two sections. The first part allows you to evaluate the components of measurement variability, while the second allows you to evaluate process variability.

Note: in all gage readings, use a maximum of three significant figures. In all calculations, carry four significant figures unless otherwise instructed.

Part One

Your group will be given a single measurement device (either a digital caliper or micrometer) and a single (aluminum or plastic) part. This will form the basis for the measurement system. Each person in the group will use the device to measure the part. This process will be repeated five times, and the results of the measurements will be recorded in **Table 1.** Be sure to mark the part type and gage type before you begin measuring.

Once the data has been recorded into the table, use a calculator or spreadsheet to compute the mean and standard deviation of each column in the table. These values are estimates of the *repeatability variation* for each combination of operator and device. (Is there any evidence to show that one operator is less accurate or precise in his or her measurements?) Note: all units of mean and standard deviation are in inches.

Next, compute the grand mean (the mean of the means) and the standard deviation of the means, which you will use to estimate the *reproducibility variation*. Finally, compute the pooled standard deviation, which is an estimate of the overall *repeatability variation* for the measurement system (an "average" repeatability variation, but you must compute it by using variance in each column - s^2). Together, repeatability and reproducibility form an *upper bound* (a worst-case scenario) for measurement variability in this measurement system. (Keep in mind that these are estimates only; you will learn how to apply the proper formulas to achieve better estimates in IE 361.)

Your group will then be given the other measurement device. Repeat the same process for the other device. Since the device is different, we will treat it as a different measurement system. (Is there any evidence to show that one measurement system is more accurate or precise than the other?)

Finally, the process of measuring with both caliper and micrometer will be repeated for the other part (aluminum or plastic). In total, you will take four sets of measurements: aluminum and plastic part measured with both caliper and micrometer.

At the end of part one, you will have an estimate of the measurement variability associated with each of the two measurement systems for both the aluminum and plastic part.

Part Two

Your group will be given additional measurement devices (either calipers or micrometers) along with all ten parts (aluminum or plastic). Three group members should select a part at random and take a measurement. The other two group members should record the measurement into **Table 2** (you are making duplicate records so that data can be shared with the other group). Be sure to mark the part type and gage type before you begin measuring.

Repeat this process until all of the parts have been measured exactly three times. Recording and measurement duties should be shared among all group members. When you have finished recording measurements, your measurement devices will be switched with the other group. Repeat the measurement and recording process for the parts again.

After you have taken measurements on all 10 parts, exchange one set of your data tables with the other group. You now have two data sets, one for each part measured with both devices. Proceed through the calculations to determine the process (part-to-part) and measurement variability. (How are these numbers affected by the precision of the measurement system?)

Discussion Items

1. Comment on the magnitude of the components of measurement variability (repeatability and reproducibility) within each measurement system.
    a. Should it be expected that one component will be greater than another?
    b. Compare the measurement variability computed in part one with the one in part two.
    c. Why are there differences between the plastic part and the aluminum part?
    d. How would you reduce measurement variability in this experiment?

2. Comment on the differences between the caliper measurement system and the micrometer measurement system in terms of overall measurement variability.
    a. For each tool and part, calculate the percentage of total variability that is attributable to measurement variability (you will need to use variance, σ^2, for this calculation).
    b. What does (b) indicate about either system of measurement (e.g. are they acceptable)? Is there any indication that one measurement system is more accurate or precise than the other?
    c. What are your recommendations regarding the measurement of the plastic and aluminum parts (i.e. how should they be measured)?

TABLE 1: MEASUREMENT VARIABILITY (REPEATABILITY AND REPRODUCIBILITY)

Part (circle one): Aluminum / Plastic **Tool (circle one)**: Caliper / Micrometer **Part #** ____

		Student 1	Student 2	Student 3	Student 4	Student 5		
	Measurement 1							
	Measurement 2							
	Measurement 3							
	Measurement 4							
	Measurement 5							
Repeatability	Mean ($\bar{x}$)	$\overline{x_1}$= ________	$\overline{x_2}$= ________	$\overline{x_3}$= ________	$\overline{x_4}$= ________	$\overline{x_5}$= ________	Grand Mean* Standard Deviation*	$\bar{x}.$= ____________ $s_{\bar{x}}$= ____________
	Standard Deviation (s)	s_1= ________	s_2= ________	s_3= ________	s_4= ________	s_5= ________	Pooled Standard Deviation**	s_p = ____________

*Grand Mean = average of $\bar{x}$'s; standard deviation of $\bar{x}$'s => in Excel, use STDEV.S

**Pooled Standard Deviation = SQRT(average of s^2) = $\hat{\boldsymbol{\sigma}}_{\boldsymbol{repeatability}}$

$\hat{\boldsymbol{\sigma}}_{\boldsymbol{reproducibility}} \approx \sqrt{s_{\bar{x}}^2 - \frac{1}{m} * s_{pooled}^2}$ where m is #of measurements (5) and $s_{\bar{x}}^2$ is the variance of the grand mean (not the pooled standard deviation).

$\hat{\boldsymbol{\sigma}}_{\boldsymbol{repeatability}} =$ ______________ $\hat{\boldsymbol{\sigma}}_{\boldsymbol{reproducibility}} =$ ______________ $\hat{\boldsymbol{\sigma}}_{\boldsymbol{measurement}} =$ ______________

TABLE 2: TOTAL VARIABILITY

Part (**circle one**): Aluminum / Plastic Tool (**circle one**): Caliper / Micrometer

	Measurement 1	Measurement 2	Measurement 3	**Mean ($\bar{x}$)**	**Standard Deviation (s)**
Part # 1					
Part # 2					
Part # 3					
Part # 4					
Part # 5					
Part # 6					
Part # 7					
Part # 8					
Part # 9					
Part # 10					
			Grand Mean* Std. Deviation*	$\bar{x}_{..}$= ___________ $s_{\bar{x}}$= __________	s_p^{**} = ____________

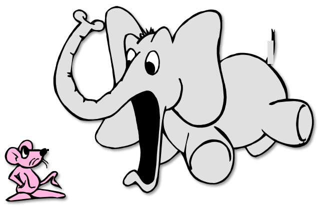
*Grand Mean = average of $\bar{x}$'s; standard deviation of $\bar{x}$'s => in Excel, use STDEV.S
**Pooled Standard Deviation = SQRT(average of s^2) = $\hat{\boldsymbol{\sigma}}_{\boldsymbol{measurement}}$

$\hat{\boldsymbol{\sigma}}_{\boldsymbol{process}} \approx \sqrt{s_{\bar{x}}^2 - \frac{1}{m} * s_{pooled}^2}$ where m is #of measurements (3) and $s_{\bar{x}}^2$ is the variance of the grand mean (not the pooled standard deviation).

$\hat{\boldsymbol{\sigma}}_{\boldsymbol{process}} =$ ______________ $\hat{\boldsymbol{\sigma}}_{\boldsymbol{measurement}} =$ _____________ $\hat{\boldsymbol{\sigma}}_{\boldsymbol{total}} =$ _____________

Sources of Variability

An engineers guide to understanding and attempting to control the inevitable- variation

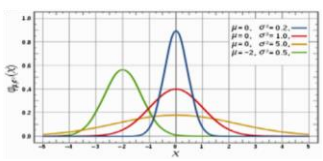


Nature of Variability

- Impossible to achieve perfect repetition
  - Throwing a football
  - Producing a precision part
- Always variability from one try to the next
  - May be very small, but it exists

Understanding Variability

- Mean (μ) = Center of the data



- Standard Deviation (σ) = Spread of the data
  - Imagine if everyone in the class lined up with the tallest person in the center and decreasing height on either side

Precision vs. Accuracy

- Another description of a data set
  - Related to mean (μ) and standard deviation ϕ)
- Accuracy = consistently hitting the center of a target
- Precision = consistently hitting the same area of a target

Precision vs. Accuracy

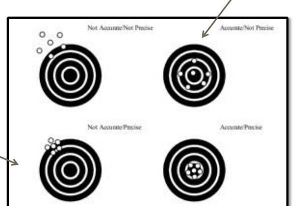


Figure: Measurement/Target Shooting Analogy

Accuracy is performing fairly well but with variation between attempts.

Precision is performing very consistently between attempts but not necessarily well.

Sources of Variability

- $Total\ Variability =$

$$\sigma_{total} = \sqrt{\sigma_{process}^2 + \sigma_{measurement}^2}$$

- Two components make up Total Variability
  - 1) Process Variability - ($\sigma_{process}^2$)
  - 2) Measurement Variability – ($\sigma_{measurement}^2$)

From Dr. Vardeman:

- "Basically, I have a sandwich here with 50 slices of bread and 49 pieces of cheese on it."

Measurement Variability

- From techniques used to measure parts
  - Device or Tool
  - Operator
  - Environment
- Environment will be ignored in this lab

Components of Measurement Variability

- 1) Repeatability Variation – ($\sigma^2_{repeatability}$)
  - ***Single*** Operator
  - Single Device and Part
    - Multiple measurements made
- 2) Reproducibility variation – ($\sigma^2_{reproducibility}$)
  - ***Multiple*** operators
  - Single Device and Part
    - Multiple measurements made

Equation of Measurement Variability

- $Measurement\ Variability =$

$$\sigma_{measurement} = \sqrt{\sigma^2_{repeatability} + \sigma^2_{reproducibility}}$$

- This experiment will show how Process Variability can be determined by collecting data on Measurement Variation and Total Variation

APPENDIX B

Study One Exams

This appendix contains the exams administered to students in Study One.

Pages 49 and 50 contain the pre-lab exam (written).

Pages 51 through 54 contain the post-lab exam (written).

IE 248 Measurement Lab

Pre-Lab Examinations

Instructions: Please write your name on the back of this exam. There are five questions on two pages for a total of 25 points. Answer each question to the best of your ability in the space provided. Any writing outside of the box will be ignored. You may use words or pictures as necessary. You have 10 minutes to complete the exam.

1. A particular model of widget is specified to be exactly 3.0" wide. A machining process has been set up to produce these widgets. Is it possible for this machining process to produce widgets that meet the specification? Explain your answer. (5 points)

2. Variability is introduced in both manufacturing processes and in measurement processes.
   (a) Explain the difference between process variability and measurement variability. (4 points)

   (b) Does measurement variability affect process variability? Explain your answer. (3 points)

3. Explain the difference between accuracy and precision in the context of a measurement system. (3 points)

4. A machinist uses a digital micrometer to measure the thickness of the first metal part that was produced by a new machining process. The reading on the micrometer was 1.7498".
   (a) If he takes nine additional measurements of the same part in the same way, will all of the measurements have the same reading? Why or why not? (3 points)

   (b) If four other machinists measure the same part once each, will they obtain the same reading as the machinist in part (a)? Why or why not? (2 points)

5. In question #3, the part was measured to be 1.7498". The part has a specification of 1.750 +/- .001".
   (a) Assuming that measurement variation is negligible, has this part been produced to specification? (2 points)

   (b) What does your answer to (a) say about future parts produced by this process? (3 points)

IE 248 Measurement Lab

Final Examination

Instructions:

1. Please write your name on the back of this exam and nowhere else.
2. This exam is closed book.
3. There are five questions for a total of 50 points. There are four pages.
4. Answer each item to the best of your ability in the space provided. Any writing outside of the box will be ignored. You may use words or pictures as necessary.
5. You have 30 minutes to complete the exam.

1. A digital micrometer is used to measure the thickness of strands of wire produced by a new extrusion process. List the types of variability reflected in the measurement data collected under each of the following scenarios. Your choices are (A) Repeatability Variation, (B) Reproducibility Variation, and (C) Process Variation (write only the letters in the box).

   (a) (4 points) A single inspector takes five measurements of a single wire.

   (b) (4 points) A single inspector measures the first ten wires to come out of the process. Each is measured once.

   (c) (4 points) Three different inspectors measure 30 of the next 100 wires to come out of the process.

2. A mechanical engineer prepares a drawing for the J2-X rocket engine nozzle to be used in the (now-cancelled) Ares V space vehicle at NASA. The drawing calls for the nozzle throat to be precisely 77.85mm in diameter (no tolerance for error in this engine lest it malfunction and explode).

   (a) (5 points) Is it possible for the machinists at NASA to make the part to the engineer's specification? Why or why not?

   (b) (5 points) At NASA, every finished part is verified independently by QA inspectors before being installed in the vehicle. Assuming that this part has been produced to specification, how can it be measured to verify its diameter is precisely 77.85mm?

3. Figures 1 and 2 below show the distribution of measurements taken from the same feature on a set of 70 aluminum blocks. The measurements in Figure 1 were obtained by several operators each using a digital caliper to measure the feature, while those in Figure 2 were obtained using a Coordinate Measuring Machine (CMM). (A CMM is a computer-controlled device used to obtain high-quality measurements, but it takes much longer to measure each part.) Each block was measured exactly once under each measurement system.

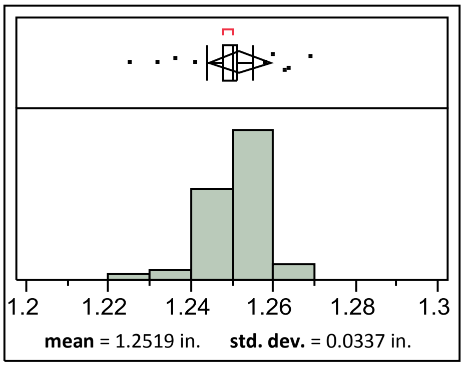


FIGURE 1: MEASUREMENTS FROM CALIPERS

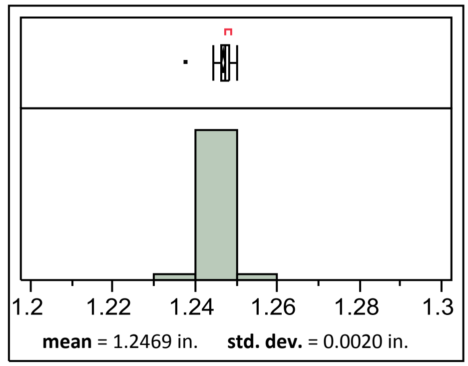


FIGURE 2: MEASUREMENTS FROM CMM

(a) (5 points) Which of the two systems shows greater variation? Why?

(b) (5 points) How do the two measurement systems compare in terms of accuracy and precision?

(c) (5 points) Choose one of the two measuring systems. What would you do to improve the accuracy and/or precision of the system?

4. (5 points) A part has a feature with a specified tolerance of +/- .025". A particular operator using a digital caliper to measure the feature has a measurement standard deviation of .03". Is this measurement system acceptable (i.e. would you use it to accept/reject parts)? Explain.

5. (8 points) Tecton Industries in Spencer, Iowa produces a variety of precision-machined components for the hydraulic and hydrostatic transmission industry. These components have tolerances of +/- .005" and are produced in a fully-automated CNC machining center. A quality inspector uses a digital caliper to verify *every part* and a CMM to verify *every 25th part* that comes off the production line. (The CMM measures multiple locations on the part, thus performing a more thorough verification).

If a part fails the caliper inspection, it is not sent to the CMM. Currently, about 1 out of every 100 parts fails inspection (counting both CMM- and inspector-failed parts).

Tecton recently tripled their production rate in response to customer demand. They still use the same CNC machining center, but they added two additional shifts (and two more quality inspectors). Each inspector uses his/her favorite tool to measure; two still use the caliper and one prefers to use a micrometer. After the rate increase, Tecton saw a large increase in the number of parts failing inspection (up to about 1 in 30).

List at least two possible causes for this issue. Given the information you have, which cause is more likely, and what would you do to address it? Explain.

APPENDIX C
Study Two Materials

This appendix contains the lab materials presented to students in Study Two.

Pages 56 through 58 contains the basic lab procedure given to students in the control group.

Pages 59 through 63 contain the detailed lab handout provided to students in both experiment groups.

Pages 64 and 65 contain data collection tables provided to students in both groups. Note that students used a total of four copies of each table.

Pages 66 and 67 contain the discussion questions that all students were assigned following the lab.

For the laboratory slides presented to students, refer to pages 46 and 47 in Appendix A.

Sources of Variability

IE 248 Lab Experiment 1

Lab Objectives

To investigate sources of variability and learn how to separate different components of variability using statistics and proper experiment design.

Introduction

Everything that can physically affect a production process introduces a portion of variability. As engineers, it is our job to understand and attempt to control that variability. Variability can be summed up into two main components: process variability and measurement variability. Process variability describes the differences from part to part in the output of a process, while measurement variability applies to the techniques used to measure those parts. Equation 3 illustrates this concept.

$$Total\ Variability = \sigma_{total} = \sqrt{\sigma^2_{process} + \sigma^2_{measurement}}$$

Equation 1: Components of Total Variability

Measurement Variability

When variability is introduced in the measuring process, it can be particularly problematic because we can only "see" process variability through measurements. It is very important that measurement variability be kept low (less than 10% of the overall tolerance window); otherwise, engineers will not be able to tell the difference between process variability and measurement variability. This is one of the main goals of the field of *metrology* – the science of measuring.

In this lab, we are going to take an in-depth look at measurement variability. As with process variability, there are multiple sources of measurement variability. These sources include the device itself, the operator, the environmental conditions (if they change over time while measurements are taken), and the measurement technique. Some of those, such as environmental conditions, can be ignored, especially for measurements taken on the same day at the same time. Others, such as operator variability, can be substantial enough to ruin an otherwise-good measurement process.

We classify the variability of a measurement system as falling under two main components:

- **Repeatability Variation**, which is the variability associated with the measurements from a *single operator* using a *single device* to measure a *single part* multiple times
- **Reproducibility Variation**, which is the variability associated in the measurements from *multiple operators* using a *single device* to measure a *single part* multiple times.

It helps to think about *repeatability variation* as something that is associated with a particular device and operator combination, while *reproducibility variation* is associated with the differences in measurement techniques between operators[1]. These two components add together to form measurement variability (see Equation 4).

$$Measurement\ Variability = \sigma_{measurement} = \sqrt{\sigma^2_{repeatability} + \sigma^2_{reproducibility}}$$

Equation 2: Components of Measurement Variability

Procedure

You will be conducting this procedure in groups of no more than five members. The procedure for this lab is divided into two sections. The first part allows you to evaluate the components of measurement variability, while the second allows you to evaluate process variability.

Note: in all gage readings, use a maximum of three significant figures. In all calculations, carry four significant figures unless otherwise instructed.

Part One

Your group will be given a single measurement device (either a digital caliper or micrometer) and a single (aluminum or plastic) part. This will form the basis for the measurement system. Each person in the group will use the device to measure the part. This process will be repeated five times, and the results of the measurements will be recorded in **Table 1. Be sure to mark the part type, part number, and gage type before you begin measuring.**

Your group will then be given the other measurement device. Repeat the process for the other device. Since the device is different, we will treat it as a different measurement system. (Is there any evidence to show that one measurement system is more accurate or precise than the other?)

Finally, the process of measuring with both caliper and micrometer will be repeated for the other part (aluminum or plastic). *In total, you will take four sets of measurements*: aluminum and plastic part measured with both caliper and micrometer.

At the end of part two, you will have an estimate of the measurement variability associated with each of the two measurement systems for both the aluminum and plastic part.

Part Two

Your group will be given three measurement devices (either calipers or micrometers) along with ten parts (aluminum or plastic). Three group members should select a part at random and take a measurement. The other group member(s) should record the measurement into **two copies of Table 2** (you are making duplicate records so that data can be shared with the other group). **Be sure to mark the part type and gage type before you begin measuring.**

Repeat this process until all of the parts have been measured exactly three times. *Recording and measurement duties should be shared among all group members*. When you have finished recording measurements, your measurement devices will be switched with the other group. Repeat the measurement and recording process for the parts again.

After you have taken measurements on all 10 parts, exchange one set of your data tables with the other group. You now have two data sets, one for each part measured with both devices. After you are finished with Part 2, proceed through the calculations to determine the process (part-to-part) and measurement variability. (How are these numbers affected by the precision of the measurement system?)

In Computer Lab

Once the data has been recorded into the table, use a calculator or spreadsheet to compute the mean and standard deviation of each column in the table. These values are estimates of the *repeatability variation* for each combination of operator and device. (Is there any evidence to show that one operator is less accurate or precise in his or her measurements?) Note: all units of mean and standard deviation are in inches.

Next, compute the grand mean (the mean of the means) and the standard deviation of the means, which you will use to estimate the *reproducibility variation*. Finally, compute the pooled standard deviation, which is an estimate of the overall *repeatability variation* for the measurement system (an "average" repeatability variation, but you must compute it by using variance in each column - s^2). Together, repeatability and reproducibility form an *upper bound* (a worst-case scenario) for measurement variability in this measurement system. (Keep in mind that these are estimates only; you will learn how to apply the proper formulas to achieve better estimates in IE 361.)

Discussion Items

On Blackboard, there are specific discussion questions that your group will need to answer. Please follow the directions on Blackboard to complete the assignment.

Sources of Variability

IE 248 Lab Experiment 1

Lab Objectives

To investigate sources of variability and learn how to separate different components of variability using statistics and proper experiment design.

Introduction

Variability is defined as "the quality of being uneven and lacking uniformity[2]." In other words, variability is an inevitable part of life. It is impossible to achieve perfect repetition, whether you are throwing a football or producing a precision part for the International Space Station. There is always some degree of variability from one try to the next; it may be very small, as with the precision part for the ISS, or very large, as with your football throw. Variability in the world is the main reason that engineers exist.

Everything that can physically affect a production process introduces a portion of variability. As engineers, it is our job to understand and attempt to control that variability. Proper control of variability allows us to reduce uncertainty, so we can produce cars that start, roads that are smooth, and planes that don't crash. Too much variability leads to great uncertainty, and when it is left unchecked, you often end up hearing about it in the news. The question then is "How do we account for all of the sources of variability?"

Understanding Variability – Mean and Standard Deviation

To control variability, we must first understand what it looks like. In statistics, there are two essential ways of describing a set of data. The first is mean ($\bar{x}$), which tells us where the center of the data is. The second is standard deviation (s), which indicates the spread of the data. Each physical source of variability contributes to variability in our data sets.

Figure 1 illustrates the concepts of mean and standard deviation using several normal distribution plots. Notice how the curves get wider with increasing standard deviation, and how they shift with a change in the mean. Mean and deviation are two different concepts, and they both must be considered when accounting for variability in our data.

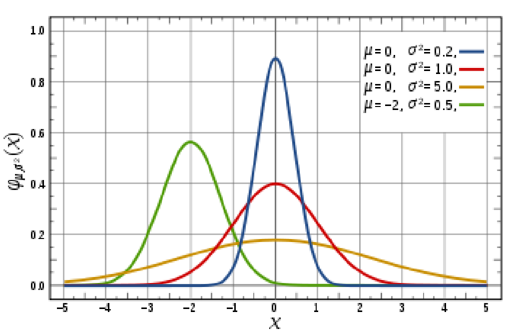


Figure 1: Standard Normal Distribution with Various Means and Standard Deviations

---

[2] http://www.thefreedictionary.com/variability

In a measurement system, another way of looking at mean and standard deviation is by understanding accuracy and precision. A good way of thinking about accuracy and precision is to look at a dart board, as in Figure 2. Accuracy refers to the direction that one is "aimed," while precision refers to the degree with which one is able to hit the target consistently. In this figure, it is clear that accuracy and precision are two completely separate ideas.

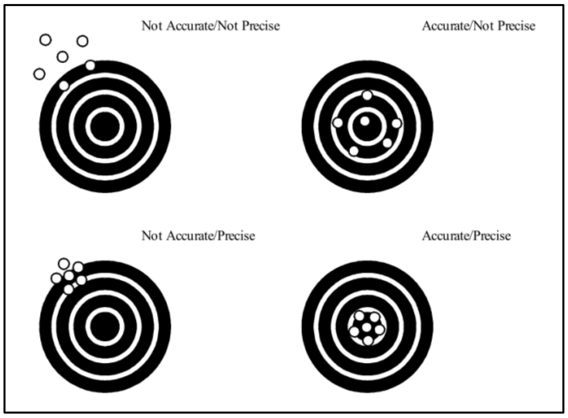


Figure 2: Accuracy vs. Precision using Targets

Sources of Variability

As mentioned previously, every physical component of a process adds its own portion of variability. Some of these sources of variability are small and can be ignored, while others are quite large and must be reduced or eliminated. As an example, let's examine a standard milling process for a block of aluminum. The aluminum must be clamped into a horizontal mill, where a face mill cutter then shaves off approximately .03" from the top of the block to cut it down to the correct height. The process is supposed to produce blocks that are 1.45" +/- .005", and this dimension is verified at the end of the process by using a digital micrometer.

Where is the variability in this process? First, not all of the aluminum blocks start out the exact same size. Differences in initial height cause slightly different amounts of material to be shaved off with the face mill. When more metal is cut, the blocks heat up and expand slightly, causing the dimensions to be just a little off. Next, the mill has to be set to the correct depth each time a block is placed on it. The height is reported by a digital gage which has a small (probably negligible) amount of variability associated with each reading. The block is clamped into a vice each time, but there is still a small amount of variability in its lateral and vertical position. Changes in temperature, humidity, and pressure from day to day can cause the material to change shape or to react differently to the cutting process. Finally, there is a small amount of variability in the measuring process at the end.

All of this variability can be summed up into two main components: process variability and measurement variability. Process variability describes the differences from part to part in the output of a process, while measurement variability applies to the techniques used to measure those parts. Equation 3 illustrates this concept.

$$Total\ Variability = \sigma_{total} = \sqrt{\sigma_{process}^2 + \sigma_{measurement}^2}$$

Equation 3: Components of Total Variability

Measurement Variability

When variability is introduced in the measuring process, it can be particularly problematic because we can only "see" process variability through measurements. It is very important that measurement variability be kept low (less than 10% of the overall tolerance window); otherwise, engineers will not be able to tell the difference between process variability and measurement variability. This is one of the main goals of the field of *metrology* – the science of measuring.

In this lab, we are going to take an in-depth look at measurement variability. As with process variability, there are multiple sources of measurement variability. These sources include the device itself, the operator, the environmental conditions (if they change over time while measurements are taken), and the measurement technique. Some of those, such as environmental conditions, can be ignored, especially for measurements taken on the same day at the same time. Others, such as operator variability, can be substantial enough to ruin an otherwise-good measurement process.

We classify the variability of a measurement system as falling under two main components:

- **Repeatability Variation**, which is the variability associated with the measurements from a *single operator* using a *single device* to measure a *single part* multiple times
- **Reproducibility Variation**, which is the variability associated in the measurements from *multiple operators* using a *single device* to measure a *single part* multiple times.

It helps to think about *repeatability variation* as something that is associated with a particular device and operator combination, while *reproducibility variation* is associated with the differences in measurement techniques between operators[3]. These two components add together to form measurement variability (see Equation 4).

$$Measurement\ Variability = \sigma_{measurement} = \sqrt{\sigma^2_{repeatability} + \sigma^2_{reproducibility}}$$

Equation 4: Components of Measurement Variability

Statistics and Variability

The field of statistics exists because variability is an unavoidable component of everything. Statistics is concerned with the proper methods of collecting, summarizing, and interpreting data, all within the context of real-world variability. Using statistics, it is possible to separate out the components of variability. While more in-depth statistical concepts will be presented in later courses (particularly IE 361), it is useful to gain a basic understanding of the statistical concepts that will be used in the collection and analysis of data. Those concepts will be highlighted as the lab procedure unfolds.

[3] Vardeman and Morris, IE 361 Module 3, 2011

Procedure

You will be conducting this procedure in groups of no more than five members. The procedure for this lab is divided into two sections. The first part allows you to evaluate the components of measurement variability, while the second allows you to evaluate process variability.

Note: in all gage readings, use a maximum of three significant figures. In all calculations, carry four significant figures unless otherwise instructed.

Part One

Your group will be given a single measurement device (either a digital caliper or micrometer) and a single (aluminum or plastic) part. This will form the basis for the measurement system. Each person in the group will use the device to measure the part. This process will be repeated five times, and the results of the measurements will be recorded in **Table 1. Be sure to mark the part type, part number, and gage type before you begin measuring.**

Your group will then be given the other measurement device. Repeat the process for the other device. Since the device is different, we will treat it as a different measurement system. (Is there any evidence to show that one measurement system is more accurate or precise than the other?)

Finally, the process of measuring with both caliper and micrometer will be repeated for the other part (aluminum or plastic). *In total, you will take four sets of measurements*: aluminum and plastic part measured with both caliper and micrometer.

At the end of part two, you will have an estimate of the measurement variability associated with each of the two measurement systems for both the aluminum and plastic part.

Part Two

Your group will be given three measurement devices (either calipers or micrometers) along with ten parts (aluminum or plastic). Three group members should select a part at random and take a measurement. The other group member(s) should record the measurement into **two copies of Table 2** (you are making duplicate records so that data can be shared with the other group). **Be sure to mark the part type and gage type before you begin measuring.**

Repeat this process until all of the parts have been measured exactly three times. *Recording and measurement duties should be shared among all group members*. When you have finished recording measurements, your measurement devices will be switched with the other group. Repeat the measurement and recording process for the parts again.

After you have taken measurements on all 10 parts, exchange one set of your data tables with the other group. You now have two data sets, one for each part measured with both devices. After you are finished with Part 2, proceed through the calculations to determine the process (part-to-part) and measurement variability. (How are these numbers affected by the precision of the measurement system?)

In Computer Lab

Once the data has been recorded into the table, use a calculator or spreadsheet to compute the mean and standard deviation of each column in the table. These values are estimates of the *repeatability* *variation* for each combination of operator and device. (Is there any evidence to show that one operator is less accurate or precise in his or her measurements?) Note: all units of mean and standard deviation are in inches.

Next, compute the grand mean (the mean of the means) and the standard deviation of the means, which you will use to estimate the *reproducibility variation*. Finally, compute the pooled standard deviation, which is an estimate of the overall *repeatability variation* for the measurement system (an "average" repeatability variation, but you must compute it by using variance in each column - s^2). Together, repeatability and reproducibility form an *upper bound* (a worst-case scenario) for measurement variability in this measurement system. (Keep in mind that these are estimates only; you will learn how to apply the proper formulas to achieve better estimates in IE 361.)

Discussion Items

On Blackboard, there are specific discussion questions that your group will need to answer. Please follow the directions on Blackboard to complete the assignment.

Table 1: Measurement Variability (Repeatability And Reproducibility)

Part (circle one): Aluminum / Plastic **Tool (circle one)**: Caliper / Micrometer **Part #** ____

		Student 1	Student 2	Student 3	Student 4	Student 5		
	Measurement 1							
	Measurement 2							
	Measurement 3							
	Measurement 4							
	Measurement 5							
Repeatability	Mean ($\bar{x}$)	$\overline{x_1}$= ________	$\overline{x_2}$= ________	$\overline{x_3}$= ________	$\overline{x_4}$= ________	$\overline{x_5}$= ________	Grand Mean* Standard Deviation*	$\bar{x}.$= ____________ $s_{\bar{x}}$= ____________
	Standard Deviation (s)	s_1= ________	s_2= ________	s_3= ________	s_4= ________	s_5= ________	Pooled Standard Deviation**	s_p = ____________

*Grand Mean = average of $\bar{x}$'s; standard deviation of $\bar{x}$'s => in Excel, use STDEV.S

**Pooled Standard Deviation = SQRT(average of s^2) = $\boldsymbol{\hat{\sigma}_{repeatability}}$

$\boldsymbol{\hat{\sigma}_{reproducibility}} \approx \sqrt{\max(0, s_{\bar{x}}^2 - \frac{1}{m} * s_{pooled}^2)}$ where m is #of measurements (5) and $s_{\bar{x}}^2$ is the variance of the grand mean (not the pooled standard deviation).

$\boldsymbol{\hat{\sigma}_{repeatability}}$ = ____________ $\boldsymbol{\hat{\sigma}_{reproducibility}}$ = ____________ $\boldsymbol{\hat{\sigma}_{measurement}}$ = ____________

Table 2: Total Variability

Part (**circle one**): Aluminum / Plastic Tool (**circle one**): Caliper / Micrometer

	Measurement 1	Measurement 2	Measurement 3	**Mean ($\bar{x}$)**	**Standard Deviation (s)**
Part # 1					
Part # 2					
Part # 3					
Part # 4					
Part # 5					
Part # 6					
Part # 7					
Part # 8					
Part # 9					
Part # 10					
			Grand Mean* Std. Deviation*	$\bar{x}_{..}$= ________ $s_{\bar{x}}$= ________	s_p^{**} = ________

*Grand Mean = average of $\bar{x}$'s; standard deviation of $\bar{x}$'s => in Excel, use STDEV.S
**Pooled Standard Deviation = SQRT(average of s^2) = $\hat{\boldsymbol{\sigma}}_{\boldsymbol{measurement}}$

$\hat{\boldsymbol{\sigma}}_{\boldsymbol{process}} \approx \sqrt{\max(0, s_{\bar{x}}^2 - \frac{1}{m} * s_{pooled}^2)}$ where m is #of measurements (3) and $s_{\bar{x}}^2$ is the variance of the grand mean (not the pooled standard deviation).

$\hat{\boldsymbol{\sigma}}_{\boldsymbol{process}} =$ ____________ $\hat{\boldsymbol{\sigma}}_{\boldsymbol{measurement}} =$ ____________ $\hat{\boldsymbol{\sigma}}_{\boldsymbol{total}} =$ ____________

Measurement Lab Discussion Questions

Directions:

Answer the questions below the heading for each question. Then save and submit this document for grading. All group member names must be listed, in alphabetical order. Submit one copy per group only.

Grading:

These questions are worth roughly half of the credit you will receive for this laboratory. They will be graded based on the following criteria:

a. 45% - answers provide correct and detailed information related to the question
b. 35% - figures are neat and tidy, and they clearly offer evidence to support the claims made in each corresponding answer
c. 20% - polish: attention to detail in writing style and construction of figures; headings are present; sources are cited; correct spelling, grammar, punctuation, flow, etc.

Group Names:

(Insert Names Here)

Questions:

1. Explain the difference between process variability and measurement variability. In a good measurement system, what does the relationship between these two components look like?

2. Perform a statistical analysis as directed in the procedure section. Report your values for Repeatability, Reproducibility, and Process variation for both part types and gage types (this will best be presented in a table of some sort). Comment on the magnitude of all components of variability in each system – explain why some variability components are higher than others.

3. Choose one of the copies of Table 1. Create a bar chart showing mean and standard deviation for each operator. What evidence do these data provide that one operator is more or less precise than the others when taking measurements?

4. Present histograms of the Table 2 data for each combination of part and measuring instrument (use the JMP program to create the histograms). Explain what the histograms show. What do they say about the best way to measure each part?

5. Comment on the differences between the caliper measurement system and the micrometer measurement system in terms of overall measurement variability.
    a. For each tool and part, calculate the percentage of total variability that is attributable to measurement variability (you will need to use variance, σ^2, for this calculation).
    b. What does (a) indicate about either system of measurement (e.g. are they acceptable)? Is there any indication that one measurement system is more accurate or precise than the other?

c. What are your recommendations regarding the measurement of the plastic and aluminum parts (i.e. how should they be measured)?

6. Using the data you have, is it possible to establish the accuracy of these measurement systems? Explain why or why not. If not, what additional information would you need?

7. Provide at least one way that measurement variability can be reduced in each setup of this experiment. Hint: do not discuss factors that did not affect your results (i.e. temperature in the room).

APPENDIX D

Study Two Surveys

This appendix contains the surveys used in Study Two.

The pre-lab survey (*Student Information Questionnaire*, pages 69 through 72) was given to students before they formally participated in the lab; one of the questions on the survey (the last question) helped sort the students into their respective lab sessions.

The exit survey (page 73) was given to students immediately after they completed the final exam.

Student Information Questionnaire

Question 1

I am a ...

- Female
- Male
- Other
- Prefer not to answer

Question 2

Please select your major.

- Industrial Engineering
- Mechanical Engineering
- Aerospace Engineering
- Civil, Environmental, or Construction Engineering
- Electrical, Computer, or Software Engineering
- Materials Engineering
- Chemical/Biological Engineering
- Agriculture/Biosystems Engineering
- Undeclared Engineering

Question 3

Please indicate the number of semesters you have attended Iowa State. If you have spent a semester studying abroad or in a co-op or internship, do not include that semester in your count.

- 1-2 semesters
- 3-4 semesters
- 5-6 semesters
- More than 6 semesters

Question 4

Please enter your ACT score. If you do not know your ACT score, you may find it on the Iowa State CMS site, under your Academic Profile, scroll down to Teacher Certification Data. Click https://ecms.eng.iastate.edu/students/

If you still do not have your score, enter 0.

Question 5

Please enter your Iowa State GPA on a scale of 0.0 to 4.0.

Question 6

Please enter your high school or community college GPA (whichever school you attended for at least 4 semesters prior to coming to Iowa State) on a scale of 0.0 to 4.0. **Note:** the maximum value is 4.0; if your high school GPA could be higher than 4.0, divide your GPA by the maximum it could have been, then multiply by 4.0 to arrive at your real GPA.

Question 7

Please describe your overall level of knowledge and/or experience dealing with statistical concepts, including probability, variability, and statistical distributions.

- Good familiarity (multiple statistics courses taken at the college level)
- Some familiarity (one course taken, either college or high school level)
- Basic working knowledge only (no dedicated statistics course taken, but you have been exposed to statistics concepts in your college coursework).
- No practical experience in probability and statistics

Question 8

Please indicate your level of comfort working in groups in hands-on laboratory classes.

- Very comfortable
- Somewhat comfortable
- Neither comfortable nor uncomfortable
- Somewhat uncomfortable
- Very uncomfortable
- Not Applicable

Question 9

Of the courses you indicated in the previous question, how many of those courses required some type of formal lab write-up following the lab session?

- No courses
- 1-2 courses
- 3-4 courses
- More than 4 courses

Question 10

Of all your past hands-on laboratory courses, how many of those courses had lecture components **which were deliberately synchronized with the lab component** (where the lecture material arrived within a week of the laboratory session)?

- No courses
- 1-2 courses
- 3-4 courses
- More than 4 courses

Question 11

The same researcher (Robert Mayer) conducted a similar study during the spring semester of 2012. Were you a participant in this previous study? Note: your answer here will not affect your ability to participate in the present study, but will be used to control for factors that may influence the results of this study.

- Yes
- No

Question 12

Have you been diagnosed with any type of learning disability? Note: it is important to answer this question honestly. Your answer here remains confidential and will not affect your ability to receive extra credit for participating in the study.

- Yes
- No

Question 13

Which class would you like the extra credit applied to?

- IE 271
- IE 148
- ENGR 160
- IE 448
- ME 324

Question 14

Please indicate which of the following courses (if any) you have taken or are currently taking at the college level. If you have taken a course that is not listed, but is substantially similar to a course that is listed, please check the "other" box, then list and describe the course on the next question. Check all courses that apply.

- [ ] ENGR 160/IE 148 or equivalent
- [ ] IE 248
- [ ] IE 348
- [ ] ME 324
- [ ] Stat 231
- [ ] Stat 101
- [ ] Stat 447
- [ ] Stat 401
- [ ] Stat 402
- [ ] Stat 226
- [ ] Stat 305
- [ ] IE/Stat 361
- [ ] Other engineering hands-on laboratory course
- [ ] Other course that addressed measurement variability as a major course topic (you may be asked to describe the course).

Exit Survey

Question 1

My group members contributed equally to the lab assignment.

- 1. Strongly Agree
- 2. Agree
- 3. Neither Agree nor Disagree
- 4. Disagree
- 5. Strongly Disagree
- 6. Not Applicable

Question 2

The lab assignment was helpful in learning the material.

- 1. Strongly Agree
- 2. Agree
- 3. Neither Agree nor Disagree
- 4. Disagree
- 5. Strongly Disagree
- 6. Not Applicable

Question 3

I did a substantial majority of the work in my group.

- 1. Strongly Agree
- 2. Agree
- 3. Neither Agree nor Disagree
- 4. Disagree
- 5. Strongly Disagree
- 6. Not Applicable

Question 4

I feel that I did well on the pre-lab exam.

- 1. Strongly Agree
- 2. Agree
- 3. Neither Agree nor Disagree
- 4. Disagree
- 5. Strongly Disagree
- 6. Not Applicable

Question 5

I feel that I did well on the post-lab exam.

- 1. Strongly Agree
- 2. Agree
- 3. Neither Agree nor Disagree
- 4. Disagree
- 5. Strongly Disagree
- 6. Not Applicable

Question 6

My group members contributed equally to the lab assignment.

- 1. Strongly Agree
- 2. Agree
- 3. Neither Agree nor Disagree
- 4. Disagree
- 5. Strongly Disagree
- 6. Not Applicable

APPENDIX E
Study Two Exams

This appendix contains the exams administered to students in Study Two.

Pages 75 through 79 contain the pre-lab exam. The exam was administered entirely on Blackboard, so the format does differ substantially from that presented in the document. Questions were presented one at a time, and students could interact with and backtrack through the exam.

Pages 80 through 82 contain part one of the post-lab exam (multiple choice). This portion of the exam was administered via Blackboard.

Pages 83 through 87 contain part two of the post-lab exam (written). This portion of the exam was handed to the students at the same time they began the multiple choice portion of the exam.

Pre-Lab Quiz

General Problem-Solving Questions

Question 1 (2 points)

Four problems arise at work simultaneously. **Specify the order in which they should be solved.**

a. A package must be shipped to your west coast office by 4:00.
b. Your boss needs a report on profit projections for a 1:00 meeting.
c. You accidentally delete the computer file containing the rough draft of the profit report.
d. The production line stops due to a part shortage, which only you can fix.

Question 2 (1 point)

Consider the following group of statements: “Dr. James Scott was the president of Harvard University. Every president of Harvard University drank vodka. Dr. Scott drank vodka in large quantities. Whoever drinks vodka in large quantities must be an alcoholic.” **Which of the following must be true if all the above are true?**

a. All Harvard University presidents were alcoholics.
b. Dr. James Scott was an alcoholic.
c. Those who drink vodka must be alcoholics.
d. Only the presidents of Harvard University drank alcohol.
e. None of the above is true.

Question 3 (1 point)

The following hypothetical conversation takes place:
James: President Bush’s decision to invade Iraq was based on fallacious information regarding Saddam Hussein’s research program into mass-destruction weapons.
William: You don’t understand that without the American invasion of Iraq and the eventual deposition of Saddam from power, he would continue to persecute the Kurds.

True or False: The two people participating in the conversation are addressing the same issue.

Question 4 (1 point)

A policewoman has been asking Mr. Wang and Ms. Vernon questions about a traffic accident. She asks Mr. Wang, who was one of the people involved in the accident, whether he had used his signal. Mr. Wang answers, "*Yes, I did use my signal*."
Ms. Vernon had been driving a car which was not involved in the accident. She tells the officer, "*Mr. Wang did not use his signal. But this didn't cause the accident*."
Choose which of the italicized statements, if either, is more credible.

a. Both statements are equally credible
b. The statement by Ms. Vernon is more credible
c. The statement by Mr. Wang is more credible
d. Both statements are equally suspect

Question 5 (2 points)

You read a story in the newspaper about salary negotiations with public transportation workers. The workers are threatening to go on strike tomorrow if their demands for higher wages and better benefits are not met. **What can you infer from this news story?**

a. Health insurance premiums are very expensive.
b. Bus fares will increase in the next few weeks.
c. People who ride the bus should look for possible alternative transportation.
d. Employers never like to meet salary demands.

Specific Subject-Matter Questions

Question 6 (3 point)

According to Merriam-Webster, a *process* is "a series of actions or steps taken to achieve an end." Using this definition of *process*, **indicate which of the following *italicized* items describes the output of a process** by placing a check next to the item.

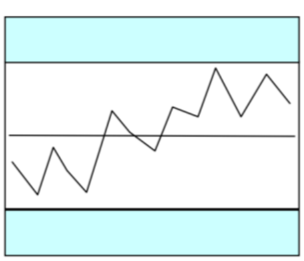
a. A machine press produces 1,000 stamped metal parts over a one hour period (*the parts produced by the machine*)
b. A quality inspector selects 100 parts at random from a batch and measures the width of each part (*the measurements taken by the quality inspector*)
c. A bottle-filling machine dispenses 20oz of soda into 100 bottles (*the bottles full of soda*)
d. A single piece of equipment used on the International Space Station is installed into the Space Shuttle payload bay (*the hardware secured inside of the payload bay*)

Question 7 (5 points)

An engineering specification calls for a machine shaft to be produced with a width of 3.0 inches. The specification does not allow for deviations from this 3.0-inch width. Is it possible for *any* machining process to produce this part to specification?

Question 8 (10 points)

The picture below shows a series of measurements taken from a machining process (where a single lathe is used to cut away metal from base stock to produce the finished part). This data is the result of a quality engineer using a digital caliper to measure the width of every 25th part out of the machine. As the data shows, there is variation from one measurement to the next . Please indicate which of the following factors (if any) affected each measurement such that it could explain the small differences in values between measurements.



a. Environmental conditions (temperature, humidity, pressure) changing over time

b. Differences in the tools installed in the machine (they wear down over time, so must be replaced every so often)
c. Variation in the quality engineer's technique when measuring each part
d. Vibrations in the floor caused by a construction crew just outside of the building
e. A faulty coolant pump in the machine, which stops for a few minutes at a time, then starts back up
f. The discovery that the digital caliper was out of calibration – every measurement taken was .003 inches less than the true value.

Question 9 (3 points)

A part is specified to be 5 inches wide with .001-inch bi-lateral tolerances (5.00 ± .001 inches). To verify the finished part, an experienced operator uses a standard ruler, which has 1/32-inch markings on it. Is this measurement system acceptable?

Question 10 and 11 (3 points each)

Choose the option that best describes the target next to the letter [10]"C" / [11]"B" in the figure below.

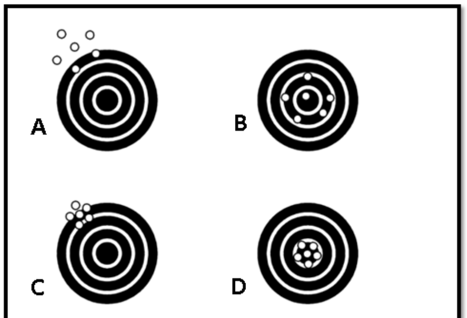


Figure: Measurement/Target Shooting Analogy

a. Neither accurate nor precise
b. Precise but not accurate
c. Accurate but not precise
d. Both accurate and precise

Question 12 (2 points)

Suppose we have a set of measurements of a system. What statistical value would give us the best indication of a system's precision?

a. Mean
b. Standard Deviation
c. p-Value
d. t-Ratio

Question 13 (2 points)

True/False In terms of accuracy/precision in a target analogy, *precision* means that we consistently hit the center of the target.

Probabilistic and Statistical Reasoning Questions

Question 14 (2 points)

A coin has been tested and proven to be fair within a reasonable tolerance. This coin is flipped 8 times. The resulting sequence of heads/tails is H-T-T-T-H-T-T-T. For the 9th flip of the coin, choose the option below that best describes the probability the coin will land heads-up.

a. The probability is greater than 60%
b. The probability is between 50% and 60%
c. The probability is about equal to 50%
d. The probability is between 40% and 50%
e. The probability is less than 40%

Question 15 (2 points)

A different coin is flipped 25 times. The resulting sequence contains 21 tails and 4 heads. For the 26th flip of the coin, choose the option below that best describes the probability the coin will land heads-up.

a. The probability is greater than 70%
b. The probability is between 50% and 70%
c. The probability is about equal to 50%
d. The probability is between 30% and 50%
e. The probability is less than 30%

Question 16 (2 points)

Poker Chips, Part 1. A black bag contains 50 red chips and 50 blue chips all mixed up. For each trial, you shake the bag, pull a chip out, write down the color, then put it back. Which of the following sequences is **most likely** to result from doing this five times?

a. R R R b b
b. b R R b R
c. b R b b b
d. R b R b R
e. All four sequences are equally likely.

Question 17 (2 points)

Poker Chips, Part 2. Listed below are the same sequences of R's and b's as listed in Part 1. Which of the sequences is **least likely** to result from the same procedure followed in part 1?

Question 18 (5 points)

Dr. Jackman computed the grade of every student in the IE 448 class at the end of the fall semester. The average (out of 100) was 74.1, with a standard deviation of 15.7. Assume that the grades were symmetrically (normally) distributed, and that the assessment method used in the course accurately measured conceptual knowledge. What is the implied "population" (the subject studied) in these statistics?

a. Every student who completed IE 448 in the fall semester
b. Any single student who completed IE 448 in the fall semester
c. The conceptual knowledge conveyed to students in IE 448 that semester
d. Dr. Jackman's ability to teach IE 448 that semester

Question 19 (5 points)

Two parts, A and B, had their diameters measured several times. Given the degree of uncertainty involved in the measurement process, we are 95% confident that Part A is between 2.997 and 3.001

inches in diameter. Likewise, Part B is between 3.000 and 3.010 inches in diameter. Assume calibrated measurement devices and symmetrical distributions. Select the statement that is best supported by this data.

a. Part B is larger than Part A by 0.006 inches, on average
b. Parts A and B are the same size
c. Parts A and B are not significantly different in size
d. Part B was produced with less precision than Part A

Specific Subject-Matter Questions

Question 20 (3 points)

In general, *measurement error* describes

a. Mistakes made by operators while using measuring tools
b. Inconsistencies in the readings of measurements from a specific device
c. Unavoidable variability within the measurement process
d. Measurements that are not accurate or precise

Question 21/22 (3 points each)

Part 1 of 2: A process can be characterized as having a stable component and at least one variable component. When computing statistics from the output of a process,
(a) which component is best described by the mean?

Part 2 of 2: which component is best described by the standard deviation?

Post-Lab Exam, Part 1: Multiple Choice

General Problem-Solving Questions

Question 1 (2 points)

Mrs. Carson took a taxi to meet her three friends for lunch. They were waiting for her outside the restaurant when she pulled up in the car. She was so excited to see her friends that she left her tote bag in the taxi. As the taxi pulled away, she and her friends took notice of the license plate number so they would be able to identify the car when they called the taxi company.
#1: The four women seem to agree that the plate starts out with the letter J.
#2: Three of them agree that the plate ends with 12L.
#3: Three of them think that the second letter is X, and a different three think that the third letter is K.
The four license plate numbers below represent what each of the four women thinks she saw. Which one is most likely the license plate number of the taxi?

a. JXK 12L
b. JYK 12L
c. JXK 12H
d. JXX 12L

Question 2 (2 points)

Determine who among the below provided experts would be the most reliable source to settle the following issue: “Does nudity on television contribute to chauvinistic attitudes of young people toward women?”

a. The president of the National Association of Broadcasters
b. A sociology professor
c. The president of the American Civil Liberties Union
d. A television talk show host

Question 3 (2 points)

Determine whether the two people participating in the following conversation are addressing the same issue.
Michelle: "I think that we should ban advertisements for striptease dancers, because I see striptease as being demeaning and exploitative."
Natasha: Relax. "These advertisements depict both male and female strippers, so it is not as bad as you think."

True or False: The two people participating in the conversation are addressing the same issue.

Question 4 (2 points)

Consider the following statement and decide which part is necessary for this argument to be complete: “Everything that exists has a cause of its existence. The universe must have a cause of its existence, for it exists. If the universe has a cause of its existence then that cause must be the first cause. Therefore, God exists.” The described passage is missing:

a. Conclusion: "God exists because everything that exists has a cause."
b. Premise: "The universe has a beginning of its existence."
c. Premise: "God is the first cause."
d. Premise: "Not all existence requires a cause."
e. Conclusion: "The existence of God explains the existence of causality."

Question 5 (2 points)

Rita, an accomplished pastry chef who is well known for her artistic and exquisite wedding cakes, opened a bakery one year ago and is surprised that business has been so slow. A consultant she hired to conduct market research has reported that the local population doesn't think of her shop as one they would visit on a daily basis but rather a place they'd visit if they were celebrating a special occasion. Which of the following strategies should Rita employ to increase her daily business?Health insurance premiums are very expensive.

a. Make coupons available for a 25% discount on wedding, anniversary, or birthday cakes.
b. Exhibit at the next Bridal Expo, having pieces her wedding cake available for tasting.
c. Place a series of ads in the local newspaper that advertise the wide array of breads.
d. Move the bakery to the other side of town.

Probabilistic and Statistical Reasoning Questions

Question 6 (5 points)

Determine the correct relationship between the following two situations:
(1) A fair coin is tossed 5 times. It lands heads-up exactly 3 times.
(2) A fair coin is tossed 500 times. It lands heads-up exactly 300 times.The probability is greater than 60%

a. Situation (1) is much more likely than (2)
b. Situation (1) is slightly more likely than (2)
c. Situations (1) and (2) have about equal probabilities of occurring.
d. Situation (2) is slightly more likely than (1)
e. Situation (2) is much more likely than (1)

Questions 7/8 (5 points each)

Poker Chips. There are two black bags full of chips. Bag (1) contains 70 red chips and 30 blue chips. Bag (2) contains 30 red chips and 70 blue chips. You do not know which is which.
Question 7: For each trial, you shake the bag, pull a chip out, write down the color, then put it back. You perform this procedure five times on one of the bags, and the following sequence results: **R R R R b.**
Question 8: You perform this procedure 30 times on one of the bags, and you end up with 22 blues and 8 reds.
Which of the following statements is best?

a. The chips probably came from Bag 1
b. The chips probably came from Bag 2
c. The chips could have come from either bag (the probabilities are about equal)

Question 9 (10 points)

Indoshell Precision Machine Company machines hydrostatic transmission barrels. To inspect these barrels, they place every 25^{th} unit on an automated CMM (a highly-precise inspection instrument). The operator is instructed to run the CMM, and look at the readout. If any of the parts' features are out of tolerance, the operator is instructed to run the part through the CMM again before rejecting the part. **Using this procedure, assuming all other factors remain consistent, which type of error is more likely to result?**

a. Type I (failure to reject bad parts)
b. Type II (failure to accept good parts)
c. Both errors are equally likely
d. Errors will not occur with this system

Questions 10/11 (10 points each)

In a set of data, Variance is defined conceptually as the average squared distance of each individual data point from the mean. We square this distance, because if it was not squared then the total variance in a symmetric data set would always equate to zero. Standard Deviation is defined as the square root of variance.

(Question 10) If a data set contains values measured in inches, what units do **mean** and **standard deviation** for this data have, respectively?

a. Inch, inch2
b. No units, no units
c. Inch, inch
d. Inch, no units

(Question 11) Choose the correct match for each value if you were to use them to evaluate a measurement system

a. Mean: Accuracy
   Standard Deviation: Precision
b. Mean: Precision
   Standard Deviation: Accuracy

Measurement Lab

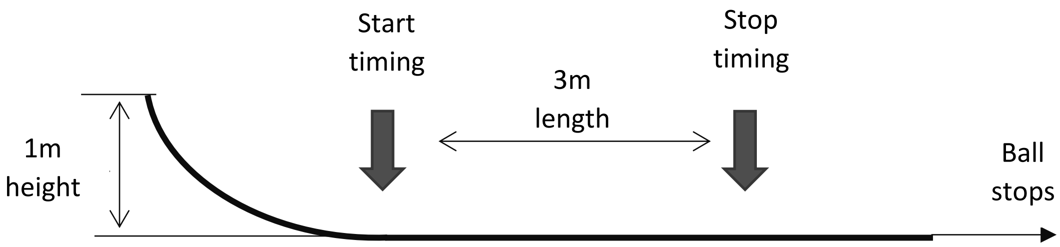
Part 2 - Final Examination Short Answer Questions

Instructions:

1. Please write your name on the back of this exam and nowhere else.
2. This exam is closed book.
3. There are five questions for a total of 50 points. There are five pages (including cover sheet).
4. Answer each item to the best of your ability in the space provided**. Any writing outside of the box will be ignored**. You may use words or pictures as necessary.
5. You have 30 minutes to complete the exam.

Question 1

Consider a simple physics experiment where a ball is dropped from a pre-determined position on a curved track (see figure below). The ball is released from height *x* and allowed to roll down the track, as its gravitational potential energy is converted into kinetic energy. To compute velocity, a stopwatch is used to time how long it takes the ball to travel 3 meters along the smooth concrete floor, marked at the points in the figure.



This experiment is repeated 10 times by the same group of four students (one drops the ball, one uses the stopwatch, one observes, and one records data). The students switch roles throughout the trials so they aren't always doing the same job. The total kinetic energy of the ball is computed for each trial by ($KE = \frac{1}{2}mv^2$), where m is the mass of the ball (a constant). The students compare the measured kinetic energy with the original gravitational potential energy ($PE = mgh$), and compute the difference. The following is a sample of their data:

Trial	PE	KE	Difference
1	4.91 J	4.64 J	**0.27 J**
2	4.91 J	3.78 J	**1.13 J**
3	4.91 J	4.98 J	**0.07 J**
4	4.91 J	4.27 J	**0.64 J**

(a) (5 points) There is variability in present this data; however, there is one row of data that is particularly problematic. Which row is this, and what is the issue?

(b) (5 points) Using the same physics experiment, label each of the following factors based on how it might have affected the outcome of each trial. Your choices are (**c**) consistent effect, (**v**) variable effect, or (**n**) no effect (write in the blank before the line).

___ Air resistance

___ Friction – ball on the ramp

___ Friction – ball across the floor

___ Measurement error – stopwatch

___ Differences in release height

~~___ Room temperature~~

~~___ Atmospheric pressure~~

___ Computational errors

___ Change in gravity due to height above sea level

___ Procedure – moving tasks among operators

Question 2

A digital micrometer is used to measure the thickness of an aluminum block produced by a machining process. List the types of variability present under each of the following scenarios. Your choices are (**A**) Repeatability Variation, (**B**) Reproducibility Variation, and (**C**) Process Variation (write only the letters in the box).

(a) (5 points) A single inspector takes five measurements of the first block produced.

(b) (5 points) A single inspector measures the first ten blocks to come out of the process. Each is measured once.

(c) (5 points) Three different inspectors measure 30 of the next 100 blocks to come out of the process.

Question 3

Figures 1 and 2 below show the distribution of measurements taken from the same feature on a set of 70 aluminum blocks. The measurements in Figure 1 were obtained by several inspectors each using a digital caliper to measure the feature, while those in Figure 2 were obtained using a Coordinate Measuring Machine (CMM). (A CMM is a computer-controlled device used to obtain high-quality measurements, but it takes much longer to measure each part.) Each block was measured exactly once under each measurement system.

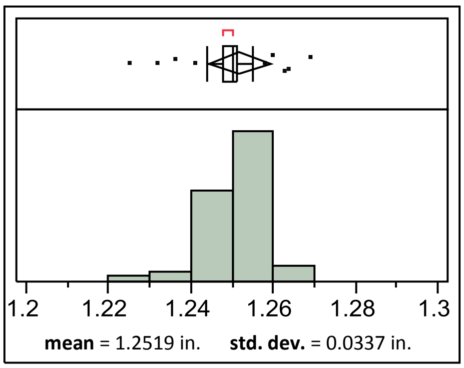


FIGURE 1: MEASUREMENTS FROM CALIPERS/INSPECTORS

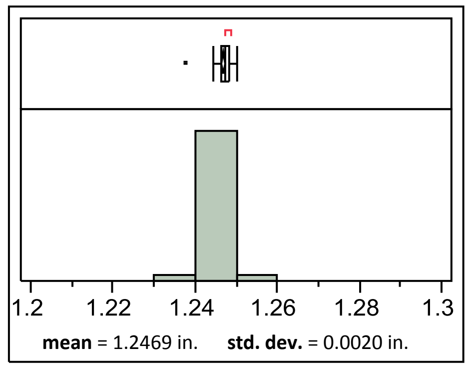


FIGURE 2: MEASUREMENTS FROM CMM

(a) (5 points) How do the two measurement systems compare in terms of accuracy and precision?

(b) (3 points) The customer specified that 100% of the parts produced by this process must be measured. If the tolerance is +/- 0.05 in., which measurement system should be used?

(c) (3 points) Suppose the customer changes the tolerance to +/- 0.5 in. Would you use a different measurement system than in (b)? Explain.

Question 4

(4 points) A manufacturer produces a component with a tolerance specification of +/- .025". A new operator using a digital caliper has a repeatability standard deviation of .03". Should measurements from this operator be used to make accept/reject decisions on the parts? Explain.

Question 5

The NCAA College Basketball Rules, Section 16, Article 1, specifies the shape of the basketball: "The ball shall be spherical. *Spherical shall be defined as a round body whose surface at all points is equidistant from the center except at the approved black rubber ribs (channels and/or seams).*" (emphasis added). Article 8 further defines the size: "The circumference of the ball shall be within a maximum of 30 inches and a minimum of 29½ inches." You are the engineer at Spalding where the basketballs will be produced.

(a) (5 points) Can the NCAA's specifications be met as-stated? If not, what needs to be changed?

(b) (5 points) Assume a basketball from your factory was produced to the NCAA specification. How would you verify that this basketball meets those requirements?