5-2015

Integrating Formative Assessment into Physics Instruction: The Effect of Formative vs. Summative Assessment on Student Physics Learning and Attitudes

Chaiphat Plybour
Western Michigan University

INTEGRATING FORMATIVE ASSESSMENT INTO PHYSICS INSTRUCTION: THE EFFECT OF FORMATIVE VS. SUMMATIVE ASSESSMENT ON STUDENT PHYSICS LEARNING AND ATTITUDES

by

Chaiphat Plybour

A dissertation submitted to the Graduate College
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Mallinson Institute for Science Education
Western Michigan University
May 2015

Doctoral Committee:

David Schuster, Ph.D., Chair
Katharine Cummings, Ph.D.
Heather Petcovic, Ph.D.

INTEGRATING FORMATIVE ASSESSMENT INTO PHYSICS INSTRUCTION: THE EFFECT OF FORMATIVE VS. SUMMATIVE ASSESSMENT ON STUDENT PHYSICS LEARNING AND ATTITUDES

Chaiphat Plybour, Ph.D.

Western Michigan University, 2015

Of many instructional strategies used to improve teaching and learning in science, formative assessment is potentially one of the most effective. A central feature is timely feedback during learning, giving students the opportunity to benefit and improve while also enabling teachers to adjust instruction to learner needs. By contrast, conventional assessment tends to be mostly summative, assigning point scores, grading and ranking students, and providing extrinsic motivation. For maximum effectiveness in enhancing learning, formative assessment should be designed into instruction from the start rather than being an add-on. This project comprised development, teaching, and research aspects. Two physics topic modules, dynamics and kinematics, were structured into sets of learning units subdivided into concept aspects. Assessments were embedded appropriately in these. Each topic module was taught in two modes; in one version, formative assessment and feedback strategies were integrated into instruction, while the other version used conventional summative assessment with graded quizzes and homework. A controlled study was conducted to compare the effects of the two systems on student performance and attitude. The topic modules were implemented in two class sections of an introductory physics course, in formative and summative assessment modes. A crossover research design involved two classes, two topics, and two modes, so that all students in the two classes experienced each mode for one or the other topic, with the same instructor. Student learning was measured using pre- and posttests and calculating normalized gains. For dynamics, a conceptual unit, learning gains were substantially higher for the formative than the summative system, while for kinematics, a more formula-based unit, the difference was less marked. Student attitudes toward various aspects of two systems, and the reasons for their preferences, were ascertained using written surveys. Students much preferred the formative system, giving reasons such as feedback, chance to improve and less pressure, but they also felt the need for summative final grades. An unexpected additional result was that students who experienced the formative or summative modes last in the course gave very different formal evaluations of course and instructor. Overall, students performed significantly better in the formative system, and most students preferred a combination of formative assessment during learning and summative at the end.

ACKNOWLEDGMENTS

I would like to acknowledge the support and dedication of Dr. David Schuster, my advisor at Western Michigan University and chair of my dissertation committee. He provided advice and encouragement along the way and from him I learned a great deal about science education, teaching, assessment, writing and research.

I thank Dr. Katharine Cummings and Dr. Heather Petcovic, members of my dissertation committee, who provided valuable ideas, feedback and resources to improve my work and dissertation.

I would like to thank Dr. William Cobern, Director of the Mallinson Institute for Science Education at Western Michigan University, who helped me with support and advice during my progress through the program.

I thank my graduate colleagues in physics education, Betty Adams, Rex Taibu and David Cassidy, for advice and feedback along the way, and special thanks to Betty for help with the data analysis. All of you are good friends.

I would like to thank my cousin-in-law Dr. Tuangtip Klinbubpa-Neff, of the University of Pittsburgh at Johnstown, and her husband Dr. Bernard Glenn Neff, who have taken care of my family and me since my first day in the English language program until the end of my Ph.D. program.

I thank Dr. Seamus Cooney at Western Michigan University and his wife Dr. Chotiros Permpikul Cooney for being dear friends, who helped solve problems and cheered me on toward excellence.

Finally, I would like to thank my parents, my wife Juntima Plybour, and my sons Napat Plybour and Aruth Plybour, for their love and encouragement during my odyssey at Western Michigan University.

Chaiphat Plybour

TABLE OF CONTENTS

Table of Contents—Continued

CHAPTER

Table of Contents—Continued

CHAPTER

Table of Contents—Continued

CHAPTER

Table of Contents—Continued

CHAPTER

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Introduction

Over the years, educators and researchers have used many methods to improve students' learning and performance in science. Formative assessment, also called Assessment for Learning (H. Black, 1986), is potentially one of the most effective methods with uniquely high leverage (Black & Wiliam, 1998a; Crooks, 1988; Kluger & Denisi, 1996; Nyquist, 2003).

The main purpose of formative assessment is to support and enhance learning by assessing current understanding and providing feedback during learning. The assessment, feedback, and improvement processes are formative in nature and purpose, and the term "assessment *for* learning" is often used for formative assessment. In contrast, conventional assessment, which may be termed "assessment *of* learning," is used summatively *after* learning for earning points, grades, or ranking; to pass or fail; and as an extrinsic motivator. It should be noted that summative assessments are also often used *during* learning (e.g., graded homework) as an external behavioral inducement to study (e.g., "carrot and stick") or as an indicator, and that receiving a summative score can also have formative effects, though limited and not usually actionable. However, it may be argued that summative assessments and scores are really appropriate only *after* learning.

I first describe briefly the main characteristics of formative assessment and some of the strategies used. A more extended account will be provided in Chapter II.

Formative Assessment: Characteristics and Strategies

According to Cizek, Andrade, and Cizek (2010), the primary purpose of formative assessment during instruction is "to identify the student's strengths and weaknesses; to assist educators in the planning of subsequent instruction; to aid students in guiding their own learning, revising their work, and gaining self-evaluation skills; and to foster increased autonomy and responsibility for learning on the part of the student" (p. 4).

Some main elements of formative assessment are effective, actionable feedback; meta-cognitive involvement in learning; recognizing capability to improve; further learning from feedback; improvement toward mastery; motivation and self-esteem; and the teacher modifying pedagogy to meet students' needs.

Empirical evidence from Black and Wiliam's literature reviews shows that the use of formative assessment including several techniques benefits all types of students (Black, & Wiliam, 1998; Stiggins, 2002), but especially helps low-achieving students, reducing the gap between the weaker and stronger students' performance.

Formative assessment systems may use a number of techniques. Several were used in the present research, namely, providing detailed learning objectives, short and long concept checks, class response systems, think-pair-share, formative homework and quizzes, prompt feedback, self- and peer-assessment, discussing assessment criteria, and the opportunity to learn from mistakes and try again. These are described briefly below.

Providing Learning Objectives and Discussing Criteria

An important aspect of a formative assessment system is that students know the learning objectives for a unit; these represent the learning goals that the teacher creates for students and allow students to developmentally construct and connect learning aspects together (Donovan & Bransford, 2005). Learning objectives and criteria allow students to gauge their performance against specific goals. In this way the focus remains clear, students are less likely to get lost in the lesson, and they realize what to do to be successful in the unit.

"Concept Checks"

What I will call "concept checks" are relatively quick "real-time" formative assessments of particular "single" aspects of current learning in class. These often take the form of a sequence of quick checks of conceptual understanding along the way during topic development, usually testing the ability to use such understanding to answer short questions and problems. Quick concept checks often are often in the form of so-called "clicker questions" using multiple choice question (MCQ) format where students select a response.

Longer Assessment Items

While a short concept check usually involves a particular identified aspect of a concept, a *longer item* typically combines several conceptual aspects and may require a written constructed response, including, for example, diagrams, explanations, and problem-solving. The shortest of these might still be done using class response systems, but longer items often work best with written answers and may be combined with a peer discussion strategy.

Think-Pair-Share

The think-pair-share technique allows individual thinking as well as peer discussion in class to arrive at answers. Think-pair-share is cooperative learning designed to involve three steps: thinking individually about a question (think), then discussing with a peer (pair), and then sharing ideas with the class (share) (Lyman, 1987; McTighe & Lyman, 1988; Tyminski, Richardson, & Winarski, 2010).

Feedback

Feedback is an important feature at the heart of the formative mode. It is used to indicate the strengths and weaknesses of a student's work and to suggest how to improve. Used best it encourages students to think about the problem during feedback, and not just wait for the correct answer (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Elawar & Corno, 1985). Feedback can occur between teacher and students and between students themselves, through discussion, think-pair-share, and other techniques. Feedback during learning ideally does not involve point scores or grades; that is not its purpose.

Instructor-, Self-, and Peer-Assessment

Conventionally, assessment and feedback is provided by the instructor; such instructor assessment remained the major assessment method used in this project. However, in formative assessment systems, it is useful, and in fact desirable, for students to take a greater role in the assessment process itself. Thus, I used a combination of instructor-, peer-, and self-assessments in the study. In *self*-assessment, the teacher explains the problem aspects and solutions and discusses the criteria for assessing answers, i.e., "what to look for" in their own work, both the main essence and the details.

Students then assess their own work (Black, Harrison, Lee, & Wiliam, 2003; Klenowski, 1995). In *peer*-assessment, each *student* assesses a peer's work using the relevant criteria (Falchikov & Blythman, 2001). The teacher explains the problem and solution to the students by referring to the learning criteria and the problem at hand, and discusses important aspects of understanding to look for. Then students evaluate their peers' attempts and provide formative comments and feedback, but do not provide a "grade" (Weaver & Cotrell, 1986).

Trying Again

An important feature of assessment used formatively is that the feedback on the first assessment be *actionable*, i.e., students can learn from their mistakes and can restudy and try again. This improvement cycle is arguably one of the most important features of effective learning, in general. The aim, even if not explicit, is eventual mastery, rather than a ranking on the first assessment.

Summative Assessment

As noted, summative assessment may be seen as assessment *of* learning, rather than assessment *for* learning. Conventionally, assessment tends to be used summatively for a variety of purposes: e.g., to provide performance scores, for grading and ranking students, for signifying pass or fail, as well as to provide extrinsic motivation to study (Stiggins, 2002). Summative assessment in the form of quizzes, exams, and grade assignments is used as a basis for a "report" after completion of instruction and exams; as information for parents, teachers, and students; and, in broad form, for schools and districts, etc. (Harlen & James, 1996; Looney, 2011). Summative assessment is used for certification purposes, for example, for the SAT, GRE, GMAT, or TOEFL. Summative assessments may also used in measuring the effectiveness of instruction (Fisher & Frey, 2007).

Research and Development Project

Because of the potential effectiveness of formative assessment in enhancing learning, I was interested in using it to improve students' performance in introductory physics. Some research evidence suggests that formative assessment can also benefit students' attitudes toward the course (Andrade & Cizek, 2010; Black & Wiliam, 2010; Black, Harrison, Lee, Marshall, & Wiliam, 2004; Sargent & Curcio, 2012; Sebatane, 1998). In the current project, I aimed to investigate the comparative effects of formative or summative assessment systems on students' performance and attitudes in an introductory physics course for pre-service teachers. The project comprised development, teaching, and research aspects, with goals as follows.

Project goals

Design goal:

- To integrate formative assessment and strategies into the instructional design of two physics topic modules.

Instructional goal:

- To teach the modules in formative and summative modes as part of an introductory physics course.

Research goals:

- To conduct a study comparing student performance using formative or summative instructional systems.
- To survey student attitudes and preferences toward the two assessment modes.
- To describe how the formative system influenced teaching.

CHAPTER II

BACKGROUND AND LITERATURE

This chapter describes and discusses formative and summative assessment systems—their characteristics, implementation strategies, and relation to learning and instruction, referring this to the relevant theoretical and empirical literature.

Formative Assessment

Introduction

Formative assessment is potentially one of the most effective methods of improving learning and teaching, and thus may be said to have high leverage. The main purpose of formative assessment is to enhance learning during learning, using feedback which is "actionable" in that it can be used to re-try and improve. This is in contrast to assessment used summatively, which is given after learning, to assign points, grades, ranking, or a pass/fail; or to provide an extrinsic motivator, as in a "carrot and stick" approach.

Cizek et al. (2010) described the primary purpose of formative assessment during instruction:

> to identify the student's strengths and weaknesses; to assist educators in the planning of subsequent instruction; to aid students in guiding their own learning, revising their work, and gaining self-evaluation skills; and to foster increased autonomy and responsibility for learning on the part of the student. (p. 4)

Bell and Cowie (2001) stated that formative assessment is "the process used by teachers and students to recognize and respond to student learning in order to enhance that learning, during the learning" (p. 536). The teacher as the leader needs to provide activities and techniques appropriately, and students need to collaborate in formative assessment mode by using the formative opportunities to achieve the learning objectives.

Much empirical evidence from Black and Wiliam's intensive literature reviews shows that the use of formative assessment, including several techniques, can benefit all types of students, but it is particularly effective in helping weaker or less prepared students (Black & Wiliam, 1998a; Stiggins, 2002).

Wiliam (2006) stated that the Assessment Reform Group in the United Kingdom specifies five elements of formative assessment. These were expanded to eight in the current project, as follows:

1. Explicit learning objectives
2. Assessment of understanding during learning
3. Timely feedback to learners
4. Feedback to the teacher
5. Learners able to assess their understanding
6. No point scores or grades during learning
7. Opportunity to improve
8. Motivation and self-esteem

Note that formative assessment can be "informative" for both the learners and instructors. For learners, it provides active engagement tasks, which give an indication of understanding so far and feedback to improve. Formative assessment creates self-monitoring that allows students to check their understanding while learning. For instructors, it provides ongoing information about how things are going in the class, which enables them to make their instruction fit learner needs and progress, both at the current time and in designing future lessons. Feedback from students could be interpreted in order to modify the lesson plan. Formative assessment can also be embedded into all types of instruction, for example, both direct and inquiry-based instruction, so its use is not restricted to particular pedagogies.

Learning Objectives

Learning objectives are learning and attainment goals that the teacher creates for students. These should be shared with the students along with assessment criteria. Students tend to be more successful in learning if they have explicit learning objectives and a means to achieve them (Black et al., 2004; Sadler, 1998). Course syllabi commonly provide only content topics and broad goals, without providing detailed learning objectives with enough specificity to enable students to know what is required for each important aspect. Students should be guided to understand the learning objectives and the criteria for attaining them (Black et al., 2003). In this way, students will not become "lost" in the lesson and will know what to do to achieve the learning goals (Black et al., 2004). Sometimes teachers merely write learning objectives on the board or have students copy them into their notebooks; these are sometimes called "wallpaper objectives" (Wiliam, 2011). Certainly this is better than providing no objectives, but taking objectives seriously should involve more than just going through the motions.

Sometimes learning objectives and criteria are not formulated clearly enough or in a way easily understood or interpreted by students; both need to be made explicit and provide clear guidance for students (Leahy, Lyon, Thompson, & Wiliam, 2005; Young, 2005). Teachers can share learning objectives and criteria with students in several ways. Questions focusing on learning aspects can be used to guide learning objectives that students should be able to master at the end (Wiliam, Lee, Harrison, & Black, 2004) . Also, teachers can provide activities to let students develop the criteria, for example, by analyzing previous work (Leahy et al., 2005). However, this may not seem appropriate when time is limited, in which case teachers normally specify the criteria needed.

Feedback

Burke (2010) stated, "Feedback is the heart and soul of formative assessment" (p. 21). Feedback is used to close the gap between students' current knowledge and understanding and desired attainment standards. Two types of feedback, evaluative and descriptive, usually occur in class; however, effective formative feedback is of the latter type (Black & Wiliam, 1998a; Black et al., 2004). "Feedback to any pupil should be about the particular qualities of his or her work, with advice on what he or she can do to improve, and should avoid comparisons with other pupils" (Black & Wiliam, 1998b, p. 6). Specific descriptive feedback enables students to realize where and why they have shortcomings and how they can improve. Sadler (1998) stated that there are three processes that comprise teacher feedback: realizing how well students understand a concept, comparing it against the learning criteria, and judging and giving relevant feedback to students. To be able to provide feedback well, the teacher needs (a) to have good content knowledge to understand student responses and whether and how they are correct or not; (b) to have an attitude toward teaching that aims to show and help students improve their learning, and to be concerned with the quality of feedback offered; (c) to be able to construct or compile good assessments, using various types of questions for the same concept to promote students' meaningful understanding rather than memorizing; (d) to provide clear criteria in a form that students can easily understand, including the expectation that students will respond to teacher assessment; (e) to evaluate and compare students' current and previous work on the same concept: "This provides them with extensive, first-hand, current experience as assessors"; the teacher better understands students' ideas, guiding them to the next stage; and (f) to provide quality feedback to students: comparing the students' work against the criteria and indicating why it is or is not correct or adequate, and making suggestions as to how to improve it.

Feedback incorporates several formative techniques. Thus, self- and peer-assessment can raise students' performance (Black & Wiliam, 1998a; Black et al., 2004). Empirical evidence indicates that concept checks with feedback can improve learning (Crouch & Mazur, 2001; Mazur, 1997), with similar outcomes when feedback is employed with clickers for a real-time response (Andrade & Cizek, 2010; Freeman et al., 2014; Kowalski, Kowalski, & Gardner, 2009). Self-confidence is improved, especially when students make a mistake on their first try and then master that particular concept after receiving feedback. In this case, students have consciously used feedback to address their previous difficulty. On the other hand, "If students are unable to relate feedback to the reasons for poor performance, self-efficacy may be diminished" (Hattie & Timperley, 2007, p. 54). Feedback given to students needs to indicate the strengths and weaknesses of students' work and suggest how to improve. However, students need to think seriously about their understanding of the problem while considering feedback, and not just wait for the correct solution (Bangert-Drowns et al., 1991).

Quality of feedback (verbal and written) is important (Black & Wiliam, 1998a). Verbal feedback is frequently used in class, while written feedback can be used in the class or on students' submitted work. Verbal feedback can provide a lot of description, for example, addressing alternative conceptions or explaining concepts and phenomena that are relevant to learning the concept. For young students, oral feedback is more effective than written feedback (Gioka, 2006). The teacher can engage students to assess their own work, such as giving a challenging question that demands more cognitive skill (Gioka, 2006). Although verbal feedback or class response system feedback is quick and convenient for the teacher to use, written feedback is still needed. For instance, written feedback—for example, in the dynamics module, referring to students' force and motion diagrams, and the directions of forces, net force, velocity, and acceleration—along with verbal feedback makes it more concrete, so in this case students can perceive graphically what they must focus on. This example reflects effective written feedback, because the teacher can specify the difficulty and indicate how to improve right on the diagram. Ideally, if there is time, but not necessarily, these processes should include effective verbal feedback to complement the written.

Feedback providing actionable information and suggestions for improvement allows students to understand what is required to attain the desired level. However, feedback consisting of giving students only marks or grades interrupts substantive learning development (Butler, 1988). When students receive a good grade, it boosts students' ego, or lowers it when they receive a poor grade, but that is not the purpose (Black & Wiliam, 2009; Crooks, 1988; Gibbs & Simpson, 2004).

Feedback should be given to address alternative conceptions. Several issues may hinder students' attaining a good understanding of science, one of which is students' alternative conceptions. Students usually come to the class with different concepts and prior knowledge about nature from everyday life experiences, culture, or community. Everyday thinking and culture may either assist or hinder scientific learning, because these may or may not be compatible with scientific concepts (Cobern & Aikenhead, 1997). Concepts that appear not to be scientific are sometimes termed "misconceptions" (Pellegrino, Chudowsky, & Glaser, 2001), but usually these conceptions are misapplications of ideas and situations and may in fact form reasonable resources for further learning if applicability is recognized and the same ideas used appropriately. Formative assessment not only is designed to improve conceptual understanding in science, but also tries to identify and address such alternative conceptions.

Feedback between students is important. When students share and discuss their ideas during learning, the processes are considered as feedback between students. Feedback between students can occur within a pair, a small group, or a large group (a large group being a class discussion that the teacher is involved in). Discussion among students allows them to explore their various ideas that may be the same as or different from the scientific concept. When students speak and explain their ideas to others, it can lead them to a fuller understanding (Cooper & Cowie, 2010). Note, however, that there is a risk that alternative conceptions may be introduced or reinforced when students themselves are not yet clear about the concept, so it is important for the teacher to guide the discussion toward the desired scientific ideas and terminology. Feedback between students can also improve students' self-confidence, especially when they realize that their ideas and solutions may be similar to those of others. Students who have different ideas initially may also have a better understanding after discussion feedback. Students may also be more willing to share tentative ideas with other students than with the teacher.

One formative technique that allows discussion feedback with peers is *think-pair-share*. Several researchers have found that feedback between students facilitates concept learning. Alexopoulou and Driver (1996) found that secondary school students in a group of four progressed significantly in conceptual reasoning. Cooper and Robinson (2000) also used small group discussion in a large lecture in chemistry and found benefits. Lyman (1981) mentioned that a pair discussion using the think-pair-share technique allows students to share their knowledge and help each other in learning.

Feedback also occurs for the teacher. Feedback to the teacher, in order to adjust and improve instruction, is another important facet of a formative system. The teacher needs the information to determine if students are following and understanding the instruction. If not, instruction can be adjusted to address the problem immediately or for the next session. Students' work also provides feedback to the teacher. Such feedback can indicate what the instruction may still require before moving forward to the next topic, and targeting aspects of current difficulties to allow students to grasp a concept better.

Feedback to the teacher occurs while each group or the whole class is discussing assignments or other work, such as homework. Visiting each group to listen to students' conversations is a way for the teacher to gather information and see how students are thinking, beyond whole-class feedback. Their conversation is feedback that the teacher then uses to help students, for example, by interceding with questions providing guidance and information if students need it. When the teacher talks to a small group, feedback dialog may be more open as students feel more comfortable with group discussion as more “private” than whole-class discussion. The teacher can dig deeper into how much they understand and why. Later, in whole class discussion, bringing closure to a topic means that students can consolidate their knowledge and feel confident that they understand. Most of the time the teacher needs to adjust instruction based on such discussions. Feedback to the teacher from, for instance, homework is another good way to see if students can transfer the knowledge learned in the class to a new situation. This is a second check of whether students are meeting the learning criteria. If not, the teacher may need to adjust or extend the instruction again. The information gained from these various ways of collecting feedback for the teacher can be quite rich and informative, allowing the teacher to determine whether students have sufficient knowledge of a particular aspect and are ready to move toward another.

In addressing students’ difficulties, prompt feedback during learning is better than later feedback. Prompt feedback helps both students and teacher to address the problem right away during the task while it is uppermost in their minds.When their problems are resolved quickly, students have more confidence to test their knowledge on other questions and then feel ready to move to the next aspect. However, for difficult concepts, students may not understand immediately after getting feedback, but may need more opportunities to reinforce their knowledge before the end of each learning aspect. If feedback is delayed, on the other hand, the problem and difficulties have become more distant and need to be recapitulated, or students may have lost initial interest, or tend to be discouraged. Therefore, whenever possible, feedback should be prompt, and if necessary, repeated in another way.

Attitude and Motivation

The manner in which a teacher teaches science in class can have a greater effect on students’ feelings or attitudes than the topics themselves (Mintzes & Leonard, 2006). For instance, dynamics is a difficult topic in physics, but if the teacher teaches well and in an enthusiastic manner, the difficult topic becomes more interesting. Teachers should introduce the content to students using effective pedagogies that engage them and keep them on task. Student-centered pedagogies can benefit student attitudes toward learning, and there is a positive correlation between the course and their attitude (Osborne, Simon, & Collins, 2003; Shrigley, Koballa, & Simpson, 1988). All types of formative techniques, such as concept checks, think-pair-share, and self- and peer-assessment, constitute active-engagement formative activities that may be said to be learner-centered and learning-centered

Motivation plays a crucial role in learning. Motivation can be intrinsic, with a student being interested in the topic for its own sake and wanting to learn, or it can be extrinsic, with a student desiring to get a good grade and avoid a bad one. A formative instructional mode aims to promote intrinsic motivation so students tend to put more effort into meaningful learning with a better attitude; this takes full advantage of formative assessment (Black & Harrison, 2001b).

More intrinsic motivation. Formative assessment should, in principle, generate higher intrinsic motivation to keep learners engaged in study or activities when they have clear learning objectives to work toward and develop proficiency. When students are intrinsically motivated, they tend to increase self-regulation and become independent learners (Corno & Mandinach, 1983; Elawar & Corno, 1985). To maximize this, students should consciously understand what formative assessment is, its purpose, and its value to their learning. The teacher as the class leader needs to guide and explain the benefits. Without this, students may feel insecure about the pedagogy used, as it requires them to work on many processes and activities, some novel to them.. This is especially true for students who have been familiar for many years with conventional teaching and summative assessment and have a hard time adapting to a new type of activity or way of thinking. When students experience many new formative techniques, they may resist if they feel it is more work and not assisting their learning. The formative approach, then, may not be very successful for these students. Recognizing the value of formative assessment, especially feedback, keeps motivating students to involve themselves in the processes as long as they believe that achievement is coming in the nearby future (Ames, 1992; Jacob, 1999). In addition, when students focus on learning and improve their performance, the processes in the class flow faster, so the teacher may have the necessary time to develop assessment and use it efficiently (Stiggins, 2002).

Improved self-confidence. Most formative techniques, such as self- and peer-assessment, feedback, and concept checks, allow students to assess themselves, especially while learning, whether formally or just in their own mind. If students recognize a mistake or lack, they can take care of it successfully. They themselves have more confidence and do not give up on learning easily, even though some concepts are difficult to understand.

Teacher motivation. Not only students need to recognize the value of formative assessment, but also the teachers, so they can incorporate it into instructional design and teaching. Its perceived value motivates the teacher to devote time to helping and improving students' performance. Without this, the teacher might stop using formative assessment, as it requires good teaching skills and experience, assessment expertise, and can be very time-consuming (Andrade & Cizek, 2010). Using formative assessment and observing students' improvement, a teacher may be motivated to keep using and developing better formative lessons. Therefore, the benefits of formative assessment motivate both teacher and students to try to achieve more in their respective roles.

Assessing Knowledge and a Second Chance in Learning

The formative mode gives students the opportunity to assess their own knowledge and act on it. Effective learners need metacognition: to be aware of and manage their own learning, including setting learning goals and checking their achievement (Donovan & Bransford, 2005). To assess learning while learning, several types of assessment can be use in formative mode: formative quizzes, homework, and tests, as well as concept checks; these may comprise open-ended, multiple-choice, or true/false questions. Note that grades are preferably not included in these formative assessments. The teacher considers which formative assessment is most appropriate to use, depending on the purpose, topic, and situation (Stiggins & Chappuis, 2006). For example, if students have just started learning a particular aspect and are trying to understand the core concepts, a series of short concept-check assessments, either open-ended or multiple-choice, are appropriate, with clicker questions being quick and effective for the progressive learning purpose (Lasry, Mazur, & Watkins, 2008; Mazur & Watkins, 2009). Each answer in a multiple-choice item can be created for particular learning and discussion purposes, focusing on particular issues that commonly arise (Schuster et al., 2006), and thus questions often include alternative conceptions as answer options. The options feedback enables students to distinguish how the correct concept differs from alternative ones. If students already “display” a certain alternative conception, they need to consider their current knowledge in relation to the desired target knowledge and criteria from feedback to better develop their scientific concept understanding (Strike & Posner, 1985). The teacher needs to do this thoroughly and check with students, because such conceptual change is not easy. Once students succeed in learning this particular concept, they should be asked to transfer their knowledge to other similar questions or concepts (Donovan & Bransford, 1999, 2005). Providing a second concept check with a similar or related concept in a different problem has two purposes: to check if students have overcome their initial difficulty, and to offer a second chance for success that can motivate their learning (Bloom, Hastings, & Madaus, 1971). In addition, short concept checks take only a short time for the whole process with the whole class together, and the teacher can use them frequently. It may take longer if the concept itself is difficult to understand and students need more feedback for more cases, but the system accommodates that.

Opportunity to Practice Working in Formative Mode

Students need time and opportunity to practice during learning and problem-solving, and formative assessment techniques can make this flow smoothly and efficiently. Of course, at the beginning, students may have some confusion about the formative system, its various components, and their roles in it. For example, they may wonder: What should I do in self- and peer-assessment? How can I assess my work? How do I understand and use assessment criteria? Students need to think about these types of questions when they start participating. To address such questions, practicing the procedures helps, and of course the teacher should guide and supervise. Naturally, the teacher also needs to understand all the techniques and processes used in formative mode, for example, providing clear, effective feedback; guiding students on criteria in self- and peer-assessment; and giving feedback addressing students’ conceptions. Thus, much thought and preparation is needed, and expertise develops with experience At the beginning, both teacher and students need to understand the multiple aspects of their roles in a system of formative teaching and learning and must develop their ability with practice, to make the system most effective. Though some processes may be unfamiliar at first, since both students and teacher are used to conventional, mostly summative systems, most aspects are in fact quite sensible and natural, once one starts to adopt a formative frame of mind.

Summative Assessment

Introduction

In contrast to formative assessment, summative assessment (which may be seen as assessment *of* learning) usually serves a rather different set of purposes. When used after learning, whether that be of a small section, whole topic, or entire course, summative assessment aims to measure, record, credit, and rank students' performances. It is used in most tests and exams, such as midterm exams, final exams, quizzes, as well as external standardized tests. When summative features such as grades and points are used *during* learning, as in graded homework, grades are being used partly for behavioral purposes, as extrinsic motivation to do the work. Point scores on details of particular questions do the same, although these do provide significant (non-actionable) feedback. However, the notion of a *summative* grade while a student is still in the process of learning may, on reflection, be seen as inappropriate. Yet it is very prevalent, when what is really needed is formative comment at that stage. Summative assessment without opportunity to improve from feedback is thus not the most effective tool for improving students' learning and performance in day-to-day teaching situations. It does provides a performance record and ranking, in a summary form, of students at a certain point in time, and usually provides the results to various stakeholders, such as the students, department, school, parents, state, etc. It is assessment for the record.

The majority of people, and, of course, students and teachers, are most familiar, from their own experience, with summative assessment and its characteristics, and so it tends to be taken as "the norm," rather unquestioningly perhaps. Summative assessment is used to measure students' performance outcomes in a particular topic or course, calculating and assigning grades aggregated from scores or marks on quizzes, homework, assignments, tests, and exams. Students are assigned a grade, and that is the end of the story on that topic; the student has been assessed and there is no particular opportunity, encouragement, or motivation to learn more and possibly improve. Students are extrinsically motivated by the grades or marks in summative assessment (McKeachie & Svinicki, 2013). The system does encourage students to study hard for good grades, but not so much for intrinsic motivation to improve conceptual understanding in the subject matter. A summative exam is used after the class has finished a topic and will move on to another one, whatever the degree of understanding of the students at the time, although their understanding is documented by grade scores (Black et al., 2004).

Note, however, that in practice summative assessment does serve (limited) formative functions as well, simply by providing an indicator of current level of achievement. In practice, no system will be entirely formative or entirely summative.

Grading System

One of the characteristics of summative assessment is a system for providing point scores, marks, and grades on students' work. Stiggins (1991) stated that grading systems are a way that teachers communicate with students about their work and performance, and to motivate its importance, suggesting that it is worthwhile to put in effort. However, good and bad grades affect students' motivation, one way or another. When grades are involved, it does create anxiety for most students, some more than others, and may affect not only the extent but also the nature of their learning.

Anxiety

Hill (1984) and McKeachie (1986) found that students' anxiety is greater when they have an important test (high stakes) and expect it to be difficult. This anxiety is also associated with pressure from test time limits. Ideally, assessment aims to determine how much students know and understand after a learning unit, or topic, as a valid measure of achievement, and thus without negative influences by anxiety and pressure. Hill and Wigfield (1984) suggested that in elementary school teachers should not use grades on school work and should also provide special assistance for students who do not perform well on tests because of test anxiety. Note that anxiety affects not only students' summative test performance, but also their way of learning. Biggs and Tang (2007) noted that students are not directly motivated to learn the ideas and concepts, but may focus more on the expected answer required for a grade or score. Hill (1984) and Hill and Wigfield (1984) suggested that to reduce students' text anxiety, teachers should: provide enough time to eliminate time limit pressure; let them know enough details about the test, its format and difficulty, with practice and examples; and give an appropriate test that allows students who work hard to experience reasonable success. The teacher should also provide special help for students who are not successful on the test. Of course, these suggestions are quite in line with the philosophy and methods of formative assessment, so some of Hill and Wigfield's suggestions do constitute a natural part of a formative assessment approach.

Self-efficacy

Students who have high self-efficacy tend to work harder to achieve, while students with low self-efficacy or who fail in learning may try to avoid doing work and/or perhaps give up (Crooks, 1988). Bandura (1982) and Schunk (1984, 1995) stated that self-efficacy has a strong effect on students' behavior, especially when they experience difficulty in learning or failure on summative tests. To improve self-efficacy, methods are needed to enable students to improve their performance and achievement (Thorndike & Woodyard, 1934), and thus realize that they can do it. Here again, formative systems have that characteristic.

Formative and Summative Assessments

The characteristics and purposes of summative and formative assessment are different; a comparison of characteristics is shown in Table 1.1, which has been modified from the original table of the National Council for Curriculum and Assessment (2014).

Table 1.1

Comparison of Formative and Summative Assessment

Formative Assessment	Summative Assessment
An integral part of the learning process	Happens after the learning takes place
Information is shared with the learner	Information is gathered by the teacher
Information is available on the quality of learning	Information is usually transformed into marks or grades
Comparison with aims and objectives is important	Comparison with the performance of others
Looks forward to the next stage of learning	Looks back on past learning

From Table 1.1, the characteristics of formative and summative assessment are clearly different, which is why they are best used for different purposes. The characteristics of each one have been clarified in earlier sections. Formative assessment focuses more on students' activities that aim to improve knowledge and understanding while learning. Summative assessment focuses more on measuring students' performance at the end. These are different purposes, both legitimate. However, it is not desirable to try to use one for the purposes of the other, or for both purposes. According to their different purposes, functions, and benefits, formative and summative assessments can be blended into instruction (Biggs, 1998; Harlen, 2005a, 2005b). Thus, a particular instructional design can be mostly formative assessment, conventionally summative assessment, or a balance between both of them. Summative use of assessment is found across all levels of schooling (Harlen & James, 1997), while formative assessment often forms part of instruction to a greater or lesser degree, using certain strategies at certain times, but purpose-designed fully formative systems are much more rare. In most places, and in educational systems, summative assessments are already in place; therefore, more aspects of formative assessment should be introduced into the system as well, to achieve a balance using both of them to best benefit students (Darling-Hammond, 2010; Stiggins, 2002, 2006). Teachers should use formative assessment in class along with summative (Stiggins, 2002). To accommodate both formative and summative assessment in instruction, teachers require good knowledge and management to achieve the most benefit. However, Black and Wiliam (1998a) emphasized the use of more formative assessment in instruction, because feedback will eventually improve students' summative performance as well, since formative assessment prepares students for success on summative tests (Darling-Hammond, 2010).

In conventional, more teacher-centered teaching approaches, the teacher provides information to students, and, for the most part, explains and provides solutions to problems, with students as the recipients (Donovan & Bransford, 1999). In this way, neither the teacher nor students themselves may realize whether they are grasping core concepts while learning. The learning outcome is usually revealed only in a test, midterm or final exam, so the teacher has no chance to address the problem, and many students leave the topic or class without a proper understanding of the concepts. This way of teaching and learning can be changed to be more learning-centered, so that students are able to determine and monitor if they comprehend core concepts while learning (Bransford, Brown, & Cocking, 2000). Formative assessment strategies potentially provide better opportunities for students to learn and achieve all the objectives. A formative view of the assessment questions or problems means the purpose is not just to "get the right answer" but rather to "give your thinking," followed by discussion to lead to the understanding desired. The method and thinking is more important than the answer, during the learning process, or arguably in general anyway. By contrast, summative systems often focus more on answers and correctness, performance evaluation.

In real teaching practice, as mentioned, neither formative nor summative systems are always completely one approach or the other. Thus, in formative systems, students need some indication of attainment at the end of a unit, and in summative systems, good teachers provide comment as well as points. However, the extent of comment varies and is often brief, such as merely checks and crosses for various parts of an answer. Such feedback does constitute some formative information, but it is rather limited and not really actionable as it is in formative systems. Note that the formative use of summative assessment (tests) is also a feasible method that helps students' learning (Black et al., 2003). Students certainly learn from feedback on graded quizzes and exams. However, Butler (1988) stated that students who were given comment feedback alone, but no points or grades, displayed superior performance to when they were given both scores and comments. Therefore, if the teacher must give a grade or score, he or she should preferably give that to student only after earlier formative practices.

For formative assessment to work most effectively, it should be part of a holistic formative instructional *system*. A formative system is not just about "putting clicker questions into existing lessons," for example. Thus, in this project, within each topic my advisor and I designed "learning units," each with detailed learning objectives, near-term timeline, and proficiency goals, along with examples, so that students knew exactly the type of thing expected of them and had an attainment target date, with formative practice along the way. "Framed" this way, their learning activities could be purposeful toward that goal. For research and development purposes, an operational model was thus produced of a formative assessment system, which incorporated the desired features.

More background information and literature for specific formative techniques, such as concept checks, self- and peer-assessment, think-pair-share, assessment criteria, etc., are described in the next chapter.

CHAPTER III

THE INSTRUCTIONAL MODULES AND ASSESSMENTS, AND THEIR IMPLEMENTATION IN FORMATIVE AND SUMMATIVE MODES

Formative assessment is likely be most effective and realize its full potential if it is designed into instruction from the start as an integral, ongoing feature of teaching and learning, rather than simply being an “add on” to existing lessons. An add-on might, for example, consist of inserting “clicker questions” here and there during instruction.. Even a rather haphazard fragmented approach to including formative assessments would surely be of some benefit, but could not hope to match an approach designed from the start with embedded formative assessment in mind.

For this project, therefore, instructional modules and associated pedagogy were designed with formative features built in as an integral part of the learning units, lessons, assignments, tests, and feedback.

The Instructional Modules

Two instructional modules were developed for the topics dynamics and kinematics, and their nature is described below. For the incorporation of formative assessment into a science course and instructional design, it was useful to conceptualize and structure the content development progression and pedagogical approach in terms of *Topic Modules* and *Learning Units.*

Topic Modules

We viewed topic modules as fairly conventional subject-matter divisions, such as those found in textbooks as chapter divisions or subdivisions, which represent the logical content structure of the discipline in that area. The two physics topic modules chosen for this research study were *Dynamics* (the relation between force and motion, involving Newton’s laws), and *Kinematics* (concepts of displacement, velocity, and acceleration in motion calculations).

Learning Units. Each topic module was structured into a set of Learning Units representing the chosen progression of conceptual development in teaching and learning. Note that this *pedagogical organization* is not necessarily the same as the *content subdivisions* one finds in textbooks. The design focus was now more on a sequenced learning approach to concept development than on final content per se, i.e., on devising an effective learning path to understanding the concepts, rather than on final-product content structure.

A learning unit was conceived as a natural unit for learning a concept or some facet of it. The guiding design question was: How can a learner best to come to the concept? This put the attention on approaches to understanding some important facet of a concept or principle, rather than on how the content of the science topic was logically subdivided as established knowledge. However, of course, the two were closely related. A learning unit could have its own detailed unit-specific learning objectives, with clear starting and ending points, and attainment objectives for the students, by a certain date, in terms of unit-specific assessments. Of course, a learning unit is also connected to other units before and after.

Structural Representation

Formative assessment was embedded into learning units in both topic modules. The instructional and assessment structure is represented in Figure 3.1 as an example of how a dynamics module can be treated as comprising five learning units. This may be done in different ways depending on the objectives, approach, scope, level, and instructor, but the diagram illustrates the idea. Several formative techniques were embedded in each learning unit, in a reasonably consistent manner, though the extent depended on the nature and scope of the unit. At or near the start of a unit, learning objectives were articulated and an attainment goal and date specified.

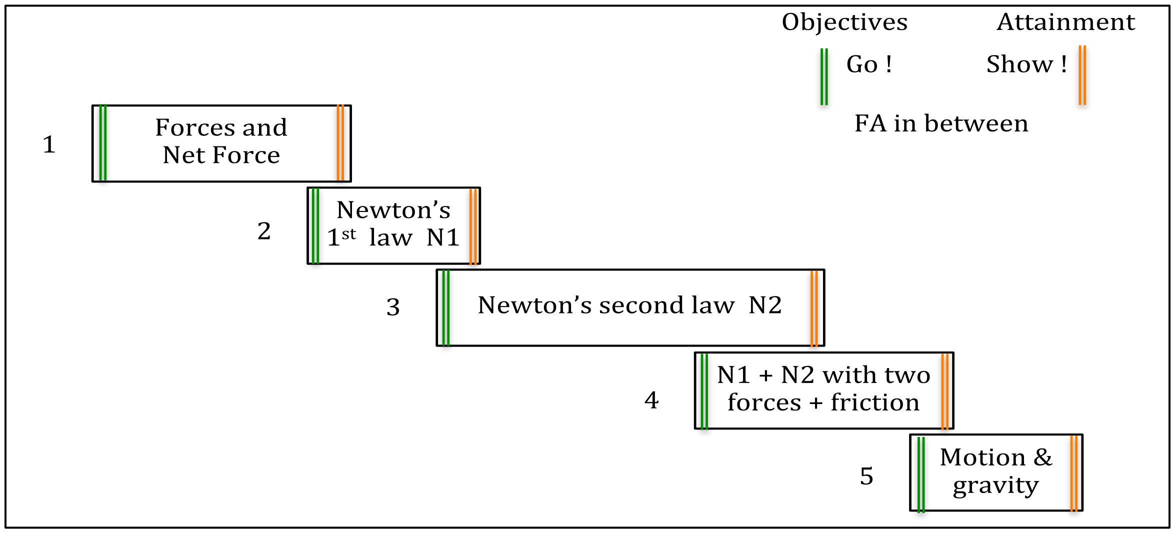


Figure 3.1. Example of structuring of the dynamics unit.

Concept Aspects Within a Learning Unit

A learning unit generally focuses on a specific concept or principle, such Newton's second law, in the above example, but within each there are several identifiable *concept aspects* involved, each to be learned. An example of concept aspects may be useful to clarify what is meant and how this was done. Consider the case of teaching and learning in dynamics, one of the topic modules in the study. Figure 3.2 below shows the example of a Newton's second law learning unit, depicted as a set of three (in this case) concept aspects, for illustrative purposes, followed by a further aspect which deals with them in combination.

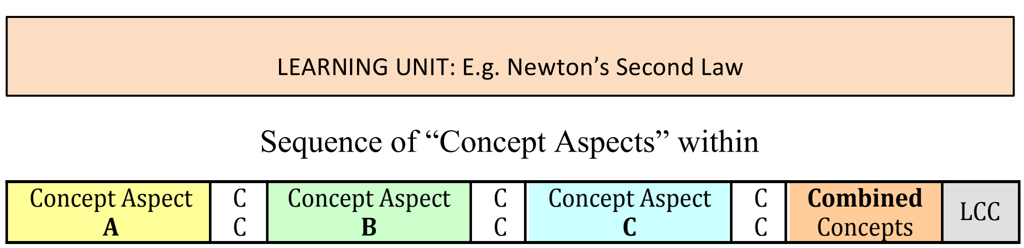


Figure 3.2. Example of concept aspects for Newton's second law learning unit.

Formative assessments, fairly brief and specific, occur along the way during instruction, as represented by concept checks (CC). At the end, a more comprehensive assessment, for example, a long concept check (LCC) and/or written assessment problems, tested to what extent the set of objectives for the unit had been attained.

Let us go into slightly more specific detail about what a concept aspect might look like within the instructional development of Newton's second law. The end-product body of knowledge was to be an understanding of Newton's second law, which is a powerful generalization valid for any type of motion, with application to many different cases. The pedagogical question was: How could one best approach this in instruction to make the central ideas accessible and conceptually meaningful to the learner? This would be different than simply presenting the law in final form as accepted scientific knowledge to be learned and used by the student in problems. The approach taken was to deal in succession with a set of specific *cases* of motion, understand each conceptually, and then generalize at the end.

The instruction for the first such case, for speeding up motion, is represented in Figure 3.3. Within this, there are two formative concept checks with feedback before students and instructor are confident about this case.

Concept Aspect A. ***Constant force —> Speeding up***

Focus Q	Exploration	Concept Formation	CC_1	Feed-back	CC_2	Feed-back	!
Constant force gives what motion?	It speeds up!	*"Constant force produced increasing speed"* (**not** constant speed!)	Case 1	Discuss	Case 2	Discuss	Got it !

Figure 3.3. Example of Concept Aspect A—in Newton's second law sequence.

After this special case, instruction in the law of motion is further developed by treating the slowing down case, in much the same way. Thereafter it is seen that the two cases can be combined by a suitable vector definition of change of acceleration. The ideas can further be extended to two dimensions and change of direction, but this was not done in the module for this project, which was limited to straight-line motions.

To summarize, the way we (note: in this dissertation the term "we" refers to the author and his advisor) chose to do it in this dynamics module, in a c*ognitive learning trajectory,* was by first considering observable *special case*s of motion (speeding up, slowing down, and changing direction). Only after observing and understanding each of these conceptually, and formulating individual rules relating force and motion for each case, did we asked the more general question of whether all might be seen as examples of a more general law of motion. This tying together led naturally to the more general form of Newton's second law, which students could now understand in terms of the particular cases now in their concept schemas.

Note that the conventional textbook topic structure and subdivisions reflect the (elegant) final structure of the science topic, presented as finished end-product content knowledge to students. But this after-the-event structure might not necessarily be ideal for learners approaching the topic for the first time (though, of course, eventually one hopes to guide learners toward the accepted end picture). Thus, our learning units were generally not identical to topic subdivisions in most textbooks. Our focus was on an identifiable and assessable natural unit of concept development.

Conceptualizing this cognitive learning trajectory in terms of learning units and concept aspects was very appropriate for this topic, with each observable case being treated as a distinct but coherent small unit of concept learning, with its own particular learning objectives, instruction, activities, and assessments. Notice that in such a conceptualization of instructional design the focus was on cognition and learning as much as on content structure. And, in this sense, the design might be seen as an aspect of teachers' pedagogical content knowledge (PCK) for teaching the particular topic. Likewise, the integrated formative assessment system might be seen as a related aspect of PCK.

An advantage of an explicit Learning Unit approach to concept development is that identified aspects of concept understanding could be taught with explicit learning objectives and using targeted formative assessment for those aspects. This instructional organization structure was then continued in the same way for the next learning unit. Custom-designed formative assessment could thus become a natural embedded part of topic teaching and concept development, within an organized structure clear to instructor and students alike.

Note that this fairly fine-grained modular learning unit approach was an appropriate framework for a formative system, which could include formative assessment during learning, targeting the concept aspect at hand. Note that it was also appropriate for an inquiry-based inductive approach to the development of these science concepts, and, in fact, a guided inquiry approach was already in the current course. Note that such a learning unit framework is by no means unique to or restricted to a formative assessment instructional system and could just as well be used with a conventional summative system; we, in fact, did just this in the study, though it was true that it tends not to be particularly prevalent in most conventional instruction.

It thus seemed appropriate for a formative system to use a learning unit structure, but at the same time it was also important for research comparison purposes to use the same structure with both the formative and summative groups, so that factors other than the assessment model should not confound the study. Thus, we used the same learning unit framework with the formative and summative instructional treatment groups. As noted above, this structure was in fact already present to a fair extent in the guided inquiry-based topic teaching design, though implicitly, so no major curriculum or materials restructuring was needed. Rather, the conscious instructional decision was to make the approach and organization explicit and align assessments with it. In fact, for the entire project there was close alignment of objectives, instruction, and assessments.

Note that although instruction was by learning units in both formative and summative systems, conventionally the “major” summative assessment was by whole-topic exams only at the end of a topic module, which in our case comprised several learning units. This whole topic summative exam setup was used for the summative system in our study also, an authentic reflection. By contrast, in the formative system, ongoing formative assessments also occurred, both within and at the end of each learning unit, in addition to the same summative topic module exam at the end as in the summative system.

Note again that learning units are not necessarily the same as topic subdivisions; rather, they were natural identifiable units of conceptual learning that were reasonably coherent within themselves, while obviously connected to other such learning units before and after. Also note that a learning unit is not the same as a “lesson.” When teachers make lesson plans, the practical unit of instruction tends to be a “class period” for understandable reasons, to make each class period “work” properly in itself, for both the teacher and students. Note, however, that such a lesson was not necessarily, or even usually, the same as a learning unit for concept development. In any event, scheduled class periods vary in length in different educational systems. In our case, for the Physics 1800 course, the scheduled class period duration was 2 hrs and 20 min. Other classes might be a 50-min. period. For learning purposes we thus prefer to think and talk in terms of learning units rather than lessons. In general, a particular learning unit may be either shorter or longer than a class period, depending on how substantial it naturally is.

Dynamics and Kinematics Topic Modules

The development phase of the project involved the re-development of two existing instructional modules to incorporate formative assessments, and the creation of sets of appropriate assessment items for both modules.

We had chosen two suitable topic modules for the study, each of about two weeks duration, with well-defined and limited scope. These units were previously written for the Physics 1800 course and had a reasonable instructional design, but without having formative assessments at the time. The topics were (1) Dynamics: Newton's 1st and 2nd Laws, and (2) Kinematics: Motion Problems. Re-developing existing modules with formative assessment in mind was much easier than developing modules from scratch. Parallel design of formative and summative instructional systems on each topic aimed to ensure equivalence in all aspects as far as possible, except for the predominant use of either formative or summative assessment.

Besides being modified to incorporate assessments, the topic development was broken into natural smaller "learning units" for formative learning purposes. The researcher also needed to determine and make sense of which formative techniques would be appropriate to use at various stages. Formative assessment is very flexible and can be used in many ways in various places; however, this did not mean that any convenient items could be thrown into instruction here and there; this would not produce the most effective instruction.

Dynamics Module. Dynamics is concerned with the relation between force and motion. When forces act on objects, how does the object behave? Students needed to learn to find the net force (resultant force), which is the effective combined force acting on an object. They then learned Newton's first law: how the object behaves when the there is no net force acting on it. After that students moved on to Newton's second law. Students observed and discovered how the object behaves when a constant net force is acting on it. Last, they investigated and learned the relationship between force and mass, for example, how the motions compare when the same force is applied on objects of different mass. They then worked with all three laws in various situations, including multiple forces and friction.

In general, dynamics may be taught at several levels, from easy to complex, and involving both conceptual understanding and formalism with calculations. However, the pre-service teacher course, Physics 1800, used in this study focused primarily on physical conceptual understanding of force and motion in a variety of one-dimensional situations.

Kinematics Module. Kinematics is the scientific description of motion, in terms of concepts of position, time, displacement, velocity, and acceleration, as well as the relations between them. Understanding and problems can be both conceptual and formula-based, involving both qualitative questions and quantitative calculations. Algebra but no calculus is involved in the Physics 1800 course.

The Assessments

Assessment embedded into a formative system enables the teacher to determine if students understand during learning. Assessments, both selected response (MCQ) and written responses, were used in the system, as appropriate to the purpose and nature of each topic and learning unit.

Assessments—Nature, Quality, and Scope

In an instructional system where formative assessment plays a significant role in learning, the nature, quality, purpose, and scope of the assessment items are important. Assessments created ranged from short concept checks to fairly comprehensive structured problems, and all were used for either formative or summative purposes. Some good items were already in use in the existing course, but not enough, and in addition, the course particularly needed short targeted items for use during concept development. Note also that items to be used formatively need to be well suited to identified purposes, but did not need to be as "perfectly" formulated as those used for summative evaluation purposes. In formative use, discussion ensued and could further clarify issues. This is an often unrecognized advantage of "real time" formative mode. This realization might help instructors who need to construct many formative items as an additional task during lesson planning.

Selected Response Items—MCQ

Selected response items were conceptual not calculational and required good understanding of fundamental concepts and principles in various physical situations. There were items at a range of levels—knowledge, comprehension, and application—depending on the purpose at hand and stage of learning. Response options often included known common alternative conceptions, so that these could be discussed in conjunction with the scientifically correct ideas. Items available elsewhere, such as in textbooks or various item banks, were often found to be of mixed or poor quality, and the weakest tended to test mainly factual recall or formula use. Thus, almost all assessment items were constructed by the advisor/course coordinator and the author/instructor. Those for dynamics were based on items previously developed by the advisor for instruction and research purposes.

Structured Problems—with Written Response

Written response questions usually were formulated as structured problems, i.e., as a sequence of sub-questions on various aspects of understanding a situation, for example, conceptual, qualitative reasoning, explanation, quantitative calculation, dependencies, etc., rather than simply requiring a numerical answer focused on one aspect, as many end-of-chapter numerical exercises unfortunately tend to do. Such structured multifaceted questions could function ideally as "teaching" problems for formative use, but they also serve well for summative assessment of the many facets of knowledge and ability required for good understanding.

There was already a good collection of multifaceted structured problems in the course, suitable for both formative and summative use, and the instructors created further items or modifications, with more fine-grained structure, for both of the modules in the project. Again, problems available elsewhere, such as textbook end-of-chapter problems, are usually not very suitable for formative use, or even summative, for that matter, especially if they are algorithmic formula-based exercises requiring students only to solve for a numerical value for a quantity.

Practice, Homework, and Quizzes

After teaching physics concepts and topics in class, the teacher usually provided opportunities for students to practice and consolidate their understanding, usually by applying their knowledge to problems, first to familiar or similar situations, and then to new or unfamiliar situations. In-class and homework assignments were used for this. The same type of questions, problems, and items were used as in the formative and summative assessments. Thus, the questions and problems assigned were excellent preparation for quizzes and tests to follow, whether these were formative or summative. In the formative system, as in the summative, students were assigned homework each session and were given a quiz at the end of a learning unit. However, in the formative mode, the teacher did not assign points or a grade to either homework or a unit quiz, but instead provided comments and discussed solutions with students instead. Thus, in this way, homework and quizzes were treated as formative assessment.

Formative System Design and Strategies

Formative System Design

To design and embed formative assessment into instruction, Ayala and Brandon (2008) stated that the six aspects should be kept in mind: (1) understand where students could improve their learning, for example, knowledge, skills, or conceptions; (2) understand techniques that could be used to measure how much students understand; (3) use formative techniques and activities appropriate to the situation; (4) understand and interpret students' responses; (5) realize when and how to address their needs; and (6) understand the benefits of formative assessment. Therefore, many aspects are involved for an effective system. The design of activities, strategies, and techniques played important roles at this point.

Formative assessments might be short, medium, or long, ranging from quick assessments of a single aspect, in class, to longer items involving several aspects of understanding and problem solving together, usually done out of class in homework, or through quizzes. The nature and quality of the assessment questions were important for the process to work best formatively. Ideally, questions should be devised for the purposes at hand and to fit a particular location in the learning sequence. Devising them during lesson preparation was a good way to do this; assessment then became a normal part of lesson planning. There were various methods and strategies to administer assessments and give feedback, which these are described below.

Short "Concept Checks"

Short-term (or "real-time") formative assessments in class during concept learning often took the form of a sequence of quick "concept checks" along the way, for concept understanding and ability to apply it. Quick concept checks were often administered as so-called "clicker questions" using selected-response items (MCQ). A short concept check should be aligned with the learning objectives and instruction and be able to motivate students to engage in class. These questions often reveal alternative conceptions. They can be used within learning or as a summary and closure for the concept. However, they should not be just a test of students' memorization of facts. Mazur (1997) uses *ConceptTests* formatively in this fashion in his Peer Instruction method at Harvard University.

Class Response Systems

There are several ways that students can respond to concept checks in class. In low-tech strategies, students raise their hands or used flash cards to give their selected response. In hi-tech methods, "classroom response systems" using clickers and computer software are now common. A clicker is a wireless handheld electronic device, functioning with software that enables "polling" of students in the class. After a question appears on the screen, students can submit answers (Beatty, Gerace, Feldman, & Leonard, 2008). A clicker has become a valuable tool for an effective pedagogy, which promotes and enables formative assessment (Beatty et al., 2008). In addition, a clicker system does not embarrass students who choose an incorrect answer, as their names are not known; this could increase students' participation and collaborative activity in a classroom. However, before choosing their best answer, students need to think carefully. Based on the responses to concept checks, the information enables the teacher to modify a lesson to address weaknesses that showed up. Formative items can be used flexibly in practice, so, for example, students could even suggest response options, not just select from them.

In this study, response systems provided a real-time indication of how thinking was distributed among students. Both low-teach and hi-tech strategies did the same thing, and which one was suitable or available was determined by the instructor.

Longer Assessment Items

Formative assessment for a complete concept or topic section occurred at the end of a section and was more comprehensive. While a short concept check usually involved a “single” aspect of comprehension, a longer item included several aspects and might require, for instance, diagrams, explanations, and problem-solving. The shortest of these longer items might still be done using class response systems and a *set* of suitable questions, which might be called a “testlet” set, but, if so, this was usefully combined with a peer discussion strategy, described later.

Hands-Down and Pens-Down Class Strategies

Having a teacher pose questions or problems for students to respond to in class is a simple form of real-time checking of student understanding during instruction. Very often the same students always raise their hands or call out answers, which means that other students do not have time to think or else know they can remain passive, and thus are not heard from. The teacher might therefore get their wrong impression of overall student understanding, be unaware of difficulties, and move on. When asked if they understand generally, students may say “yes” but don’t really understand.

Therefore, Black and Wiliam (1998a) advocated a “hands down” policy where after a question, hands must remain down, and after some wait time, the teacher would decide which students to call upon for their thinking. This produced better engagement and was more formative. Another issue was that, after students had tackled a more substantial problem and the teacher was discussing or explaining the solution at the board, students might simply take down notes rather than listening properly and thinking, so that the explanatory feedback had limited effect. Therefore, the teacher can adopt a strategy of “pens down” or “hands on heads” while explaining, with note taking later if need be. These procedures and strategies can have an effect on classroom learning behavior and how students view their role in the procedures.

Think-Pair-Share

A think-pair-share strategy allows individual thinking as well as peer discussion in class to arrive at an answer. Think-pair-share is a form of cooperative learning, designed to involve three steps: thinking about a question, discussing with peers, and sharing ideas with the class (Lyman, 1981; Tyminski et al., 2010). When the think-pair-share strategy was used in this study, after the question had been provided, students started to think and try to find their best answer. (They could either scribble or think about the question.) The teacher needed to provide a waiting time to let students think in order to find the answer (Black & Harrison, 2001a; Moriber, 1971). Then, students shared attempts with a "pair" partner. Sharing with a partner or peers meant they had chances to learn from each other. It also meant they had to explain to the other, so they had to organize their explanation. This itself was useful in learning and retaining. After pair discussions, they accepted ideas that made sense to them. Students who worked in a pair or a group felt more secure, with less anxiety and embarrassment, because normally, if they were unsure, they might not be willing to participate in public, for example, in a lecture room. However, they would be willing to first share an idea with a fellow student in the class and be more confident. In the Share phase, the teacher needed to guide the discussion if it was not on the right track or incomplete. At the end, the teacher closed with the main ideas, writing them on the board. If sharing inundated the class with many ideas, it was important to highlight the essence and the take-away message the teacher wanted students to acquire.

During in-class formative checks of understanding, a single iteration of initial and then modified responses to an item, along with instructor explanation, was usually enough to bring most students to the desired point. This meant that the next stage of learning could proceed on a solid basis, while without this, many students might be left behind and unable to build further knowledge effectively. Note that the discussion that went with a formative item was an important feature of the system.

Feedback

Whether it was for a quick check, a short quiz, or a longer test, feedback needs to be prompt. For concept checks in class, this is virtually immediate. For quizzes or tests, this ideally means feedback should be provided by the next session, while ideas are still fresh, to maximize the formative effect. The system should also involve the opportunity to try a question (or a similar question) again to improve and demonstrate proficiency. The nature of formative feedback ideally provides fairly detailed written or verbal comment and suggestions on students' attempts. In a "pure" formative system, one provides feedback but deliberately does not assign points or grades. Black and Wiliam (2004) noted from their study that students tended to pay rather little attention to feedback when points and grades seemed to be the main issue. Moreover, points allocated were often an indication of the correctness or otherwise of the answer, rather than feedback on student thinking. Note that students who were used to conventional summative scoring and grades took a while to adjust their thinking and behavior to a different system.

Out-of-Class Learning: Homework and Assessments

Formative homework and assignments can also be dealt with formatively, for example, by posing questions to reinforce or extend current learning, and and a chance to do a problem again correctly after feedback comments

Instructor-, Self-, and Peer-Assessment

In formative assessment systems it is useful, and in fact desirable, for students to take a greater role in the assessment process itself. Thus, we used a combination of instructor-, peer-, and self-assessments in the course. Self- and peer-assessment was done in class under instructor guidance, and the discussion of problem solutions and how to judge them was to enhance students' understanding and problem-solving ability. This combination system should have had benefits for both students and instructors. For students, taking some shared responsibility for assessment and how to do it, rather than viewing it purely as the instructor's role, meant they became aware of criteria for assessing knowledge and ability. Thus, they were better able to assess their own work and level of understanding, which was surely important in managing their own learning. There should also be less of a “role division” between instructors and students in this regard. Formative assessment, therefore, is not an authoritarian way of teaching; the teacher, students, and techniques play critical roles in formative assessment and must collaborate very well during instruction in appropriate situations to produce the most productive learning.

Self-Assessment

When self-assessment was used, I, as instructor, first explained the problem solution and facilitated discussion so that the class understood. I then talked about the criteria for assessing answers, that is, “what to look for” in their work, both the essence and the detail. The students, by this point, should have understood well enough after the solution discussion and were ready to use this knowledge to assess their own work. Students then assessed their own work, so-called self-assessment (Black et al., 2003; Klenowski, 1995). They were to both correct and comment. That way, they were learning from their mistakes. Self-assessment with explicit criteria should be beneficial to students because it helps them become independent learners who know how to judge their own understanding in the future (Rolheiser & Ross, 2000). For the empirical evidence, Fernandes and Fontana (1996) used self-assessment with young students for 8 months and then compared the learning outcome with a conventional group. They found that students experiencing self-assessment outperformed another group. White and Frederiksen (1998) found a similar result. Black and William (2001b) stated that student engagement in self- and peer-assessment was an important beneficial condition for formative assessment. Therefore, formative assessment strategies in the learning units in this course included both techniques.

Peer-Assessment

Falchikov and Blythman (2001) defined peer-assessment as students assessing a peer’s work by using relevant criteria. In our case, I, as instructor, explained the problem solution and discussed what to look for in assessing, as was done for self-assessment. Then students assessed their peers’ attempts, providing formative comment and feedback, but not “grading” it (Weaver & Cotrell, 1986). This was deliberate, as it is known to be better for learning, but it also avoided arguments about exact points awarded. As students assume the role of a teacher in peer-assessment, they are motivated to learn the core concepts well enough to do so. Assigning grades in peer-assessment tends to displace peer feedback to improve learning, and students resist it (Liu & Carless, 2006). Falchikov and Boud (1989) stated that the quality of peer-assessment was close to the teacher assessment when students understood the criteria well. The teacher could clarify criteria by providing a rubric and letting the students use it to assess the work. Providing a rubric and criteria for students could also help them improve their own understanding. (Murthy, 2007).

Another way to operate peer-assessment (and self-assessment) was for the teacher to discuss aspects with students along the way, and let them check and comment on the work as they continued to assess. This way seemed to take a shorter time than the first. However, the teacher needed to check closely with the students in this process. Topping (1998), after reviewing several journals, confirmed that comments and not marks in peer-assessment have a positive impact on students' attitude and performance.

For instructors, using some peer- and self-assessment in class can be an interesting aspect of teaching. In a way, this is case-based teaching; it follows the teaching of general concepts and principles and "going over" solutions by going beyond to specific cases before them. Self- and peer-assessments could also provide some welcome "marking relief" for teachers (Boud & Holmes, 1995), which could partly compensate for the extra work involved in creating formative assessments and providing feedback. Note that in a formative rather than summative system, some feedback on student work could be provided in class discussions, not just by detailed written feedback on each individual student's submission. Otherwise, the feedback provision task can become too arduous for large classes. Here, peer- and self-assessment has a workload benefit to the teacher. In addition, in many cases just the discussion was formative enough, without the formal assessment aspect.

Conventional Summative System

I have described the embedded formative assessment system in some detail, but will not do so in similar detail for the summative system, because it was the existing system, rather conventional, well known and familiar, that is commonly used in most courses, and had long been used in past years in this particular course as well. The basic characteristics will, however, be mentioned to make clear the distinctions with the formative system. In a summative system, when speaking of quizzes, homework, tests, exams, points, test scores, and grades, most people know what is conventionally meant and how the system generally operates. Of course, individual courses and instructors vary, but the overall system is recognized as a fairly well-established norm.

Our conventional pre-existing summative assessment system was used (in parallel with the formative) in the this study, in the dynamics and kinematics modules, for the performance comparison part of the project. The summative mode included homework, quizzes, exams, and grades, i.e., evaluating students' performance mostly *after* instruction and learning of a topic or subtopic, and providing a rating. In the summative assessment system, homework, quizzes, and exams were the main tools to rate students' performance. Homework was assigned to students almost every class meeting. Quizzes were given to them after a topic module. Also, a module exam was given at the end of each module. The teacher provided grades to students a few days after these summative assessments.

Homework was provided at the end of each session, usually to be turned in by the next; the idea was to have students consolidate concept understanding after everyday teaching and learning, and to keep them on task. After grading turned-in homework, the teacher regularly returned it to students with a grade and then showed the solution on the board. Students had opportunity to ask questions.

A quiz was given to students after a topic module. Students used clickers to respond to the MCQ items. The solutions were directly presented in the next class meeting without discussing incorrect choices in any detail; however, students had an opportunity to ask if they did not understand. The quizzes in kinematics used written answers. Students still did not have immediate and descriptive feedback, but they received a direct explanation about the solution with grades from the teacher, usually the next class period or the following one.

In dynamics, students took a pre- and posttest. The items were 23 conceptual multiple-choice problems.. In kinematics, students had open-ended exams for both pre- and posttest.

Implementation of the Module in Class Instruction

The Physics 1800 course took a mostly inquiry-based approach to developing physics topics, typically using a Karplus or 5-E learning cycle (Karplus, 1975; Bybee, 1997) to reflect scientific inquiry in instruction. The central phases of a learning cycle of were Exploration, Concept Formation, and Application. How might formative assessment support these phases? During the exploration and concept formation phases, formative questions were devised to help with the concept development sequence. This "assessment during concept development" use of questions was somewhat different from questions to assess understanding of concepts already developed, and it was an interesting challenge to devise items for this "along the way" inquiry process.

As an example, consider the case of developing concepts of the relation between force and motion. The inquiry issue was: "What kinds of motion might result when a (net) force acts constantly on an object?" At the start of exploring this, a formative question was posed asking or offering various possibilities to consider, such as, "it might move with constant speed," or "speed up," or "not move," or "slow down," or "depends on the size," etc. Note that this was not assessment of existing knowledge, but a formative *question* promoting scientific inquiry thinking about situations, and offering alternatives to think about while learning. Note that it was best framed in terms of generating possibilities, rather than getting the right answer before inquiry, or predicting by guessing with no basis. The question then led naturally to trying the experiment in practice and actually finding out. This example has some resemblances to "interactive lecture demonstration" strategies for physics lectures, but we deliberately asked for possibilities, rather than elicit answers or commit to predictions, to better represent real scientific inquiry.

After the concept formation stage for each aspect, short or longer conceptual checks as a formative technique can help students to check and improve conceptual understanding before moving on to the next concept. A short concept check was appropriate to use in the concept development stage because it usually did not take much time compared to other techniques. Its outcome, if different than expected, might surprise students, who thought that they completely comprehended the concept, as well as the teacher, who thought that the concept was easy and almost all students would understand it. The check outcome was thus informative to both the students and the instructor, and was thus valuable if the teacher combined it with an explanation that students still clearly needed, that is, feedback to them to reduce the gap between present and desired knowledge.

In the *application* phase of a learning cycle, that is, after concept formation, students used the science concepts to solve problems and answer questions. Application involved using the concepts that students had just learned, so students had an opportunity to apply their conceptual understanding to both similar and new situations and work on many types of practice examples. Application likewise occurred in direct instruction in the same way. In *either* method, formative assessment could be used to both assess and improve understanding.

To summarize, formative questions could be used both to generate concepts during the inquiring phases, and to assess concept understanding afterward. Note that in the mostly summative system, the same shorter and longer questions and problems were still provided to the students in instruction, but treated conventionally as homework or extra practice.

Focus Questions and Learning Objectives

For the most part for the formative mode, existing focus questions were explicitly presented also in the form of learning objectives to be attained. Focus questions were also usually a starting point in the instructional approach to both modules, which was largely inquiry-based. The focus questions, which were in statement or question form, were presented to students to introduce each aspect. The sequence of focus questions enabled students to perceive which aspect needed to be understood first before moving to the next, and to prevent them from getting lost in activities. Examples of focus questions used were the following: “If there is a net unbalanced force on an object, how will the object behave?” Students needed to observe objects, such as a bowling ball, hover puck, or person being pushed on a skateboard, under these conditions, and formulate proposed laws of motion from their observations.

In these thinking processes, and in exploring and experimenting, students were guided by the instructor. The way students first attempted this did not surprise the instructor, as they had not learned the concepts yet, so, of course, the teacher did not expect students to understand and be able answer all of the focus questions correctly at that stage. Sometimes a few of them shared their ideas and opinions by making a statement or guess and asked if it was correct. The teacher did not answer but encouraged students to discover by themselves.

Short and Long Concept Checks, Think-Pair-Share, Clickers, and Feedback

Various formative strategies were used—short and long concept checks, think-pair-share, clickers, and feedback—to improve students' conceptual understanding and influence collaborative learning.

Using short concept checks—MCQ, true or false questions, or open-ended questions—was essential to discover if students were grasping the concepts. One or two concept checks, usually multiple-choice questions, were used during learning, such as, "Is there a net force acting on a skater while moving along the ice at a steady speed?" Longer concept checks that included more aspects were given to students at the end of each learning unit. Such items required more, e.g., drawing diagrams and solving problems. Students would think and choose their best answer and share this with a student nearby or a group (Think-Pair-Share). The teacher encouraged students to go with their best answer if they felt that they had adequate evidence to support it, or if an opinion of a friend did not make sense to them. In this way, the teacher expected to see students' responses based on current conceptual understanding. Responses to questions could be either via a low-tech device, high-tech device, or a combination of both. Students regularly experienced either or all devices everyday before the project on dynamics and kinematics. Whatever type of item and response was used, the teacher needed to provide enough time for students to think.

The high-tech class response system, as mentioned before, was a clicker system with a hand unit with 5 buttons, A, B, C, D, and F, for responding. In the class, the teacher posted a concept check through PowerPoint or MS Word. To use the clicker system, a teacher connected the clicker base unit to the computer and installed clicker-system software. Each student needed to register in the clicker-system software to record the serial number of the clicker, which indicated who was participating. There were several interesting features of clickers used in this educational system. For example, the teacher could give rewards (points) to students who replied to questions correctly, or penalize those who replied incorrectly. However, this research was designed so that in the formative assessment system no points or grades were involved. In this way, students could get feedback on concept checks without pressure.

Note that students who purchased a clicker that was compatible with the university's system could also use it in other classes, so it was not a large burden financially. In class, if students did not bring their own clickers, the teacher had extra clickers to lend them.

Example of the Use of Concept Checks in a Clicker-Based Class Response System

A specific example of the formative use of clickers in the classroom with concept checks is shown in Figure 3.4. The conceptual question posed had alternative answer choices that corresponded to common alternative conceptions about force and motion.

The kind of concept check and process described above could also used as a more comprehensive long concept check later at the end of the learning unit, by requiring extra written procedures. For example, students had to draw motion diagrams, including magnitude and direction of forces, velocity, and acceleration, and explain their reasoning briefly.

1. **Question posed (on the screen and verbally)**
   An airplane is flying horizontally with constant speed.
   Which of the following is true about the forward and backward forces (if any) acting on the airplane while flying thus?
   A. There is only a forward force.
   B. There is a forward force and a backward force, but the forward force is greater.
   C. There is a forward force and a backward force, which are equal in magnitude.
   D. There are no forces acting on it while it is moving thus.

2. Teacher: "Read the question, choose your best answer and think about reasons, evidence, or example to support it. You can jot ideas on paper too. This is no grade, but do your best. (0.5–1 min)

3. Students clicked to enter their answers anonymously, and the teacher showed the response distribution but did not talk more about the question yet.

4. Teacher: "Now share your opinion with a friend nearb. Discuss. Which answer do you now prefer?? Why? Then, click again to enter your best answer." (Think-pair-share) (1–1.5 min)

5. The response system then displayed the response distribution again after this "share" stage, normally different in various respects from the original distribution. The teacher and class could see how the students overall now understood the issues conceptually from the various responses, choices, and how many thought each particular way.

6. However, the teacher (and students) still did not know the various thinking and rationales that students had behind the particular answers. In order to reveal that information, discussion and feedback between the teacher and students was now used. (2–5 min)

7. Teacher encouraged students to participate and raise questions and difficulties to help them to fully understand. They had now been told which answer choice was correct, and the class discussed the explanation for that.

8. Note that the incorrect answer choices were discussed in the same way, since students needed feedback on these too, to enable them to understand where these options and ways of thinking were not correct.

9. The teacher brought closure by briefly wrapping up the main ideas, correcting thinking and common difficulties.

Figure 3.4. Example classroom procedure for the use of concept checks with clickers.

Example of Using Concept Checks in a Low-Tech Manner

Besides the high-tech clicker class response systems with projected displays, low-tech methods were also used. While not as sophisticated and without graphic displays, these served the same function and provided formative information to both the students and the teacher. One such method used was "Close your eyes and raise your hand." Although the technique was different, the procedures were still the same. The low-tech methods do have some practical advantages, in that they are relatively simple to implement, with little or no setup, and can be done spontaneously. Neither instructor nor students need to buy equipment or have a classroom equipped with computer and projector facilities. Students don't need to buy clickers or worry about losing them or forgetting to bring them to class. However, note that students did enjoy the clicker system and its quick anonymous gathering and displaying of responses for the whole class.

"Close your eyes and raise your hand" was a simple low-tech (or no-tech) device used in the class, in conjunction with think-pair-share, feedback, and discussion. Students could respond to alternative answer options one-by-one, starting from A, B, C, and D, consecutively, by raising their hands when the option was called. They had their eyes closed so as not to be influenced by others and to feel their initial response was not known to the other students. The typical process is described in Figure 3.5.

Question: Observe the steady motion of the hover puck on the smooth floor. During this time are there any horizontal forces acting on the moving hover puck?

1. Teacher: "Consider the question and the four answer options A, B, C, and D. Choose your best answer and think about reasons, evidence, or example to support it. You can put your answer and reasons on paper also. This is no grade, but do your best. (0.5 min–1 min)
2. Teacher: "Share your opinion with your friend nearby .If your friend's idea does not convince you, please go with what you think it is correct." (Think-pair-share) (1–1.5 min)
3. Teacher: "Close your eyes. No peeking! Who think that choice A is correct? Raise your hand. B? C? and D?" (The teacher estimates how many students chose each answer option, including the correct one and those that reflected various alternative conceptions (25–45 sec)
4. Teacher: "Open your eyes."
5. Then the teacher told students how many chose A, B, C, and D.
6. Next, the teacher and students discussed each answer , including the correct one and each of the other options. (2–5 min)
7. Teacher may say: "Who knows why this answer choice is correct or incorrect. And why?" (Feedback between students and teacher)
8. Teacher brings closure by consolidation of the main ideas and difficulties.

Figure 3.5. Example classroom procedure for low-tech concept check response.

Whether using high- or low-tech methods, after two concept checks on a single issue, the teacher could usually determine if the majority of students understood well enough. If not, second concept check was either used immediately or in the next session, depending on the information from the feedback and the time available.

The system did not always work ideally or as desired, as may be expected in most teaching. For example, sometime students did not participate properly in the responses or in the feedback discussions, but instead just waited for the teacher's statement and explanation of the right answer. In such cases, the teacher could provide feedback as statements and use the "Close your eyes and raise your hand" process again to confirm if they indeed understood the statements, as in the example below (Figure 3.6).

1. Teacher: "Close your eyes. If you now think that after giving the initial push on hovercraft, there is no net force acting on it while moving because of no force from the push and the friction between the hover puck and the floor, raise your hand."
2. Students: Raise/Do not raise

Figure 3.6. Follow-up confirmation check.

Most students usually "got it" by an early stage. If a few still disagreed, the teacher asked them to come to discuss it during the break, based on the concept checks. Asking them questions and giving more feedback based on their responses showed that almost all students did eventually grasp virtually all the concepts taught this way.

Self- and Peer-Assessment, Criteria and Feedback

This section gives brief descriptions of self- and peer-assessment as used in the project.

In self-assessment, students assess their own work, usually with explicit criteria and guidance from the instructor. In peer-assessment, students assess another student's work, again with explicit criteria and guidance from the instructor. Peer-assessment may or may not be anonymous. Homework problems were given to students almost every class meeting on Tuesdays and Thursdays, and self- or peer-assessment, along with feedback, occurred in the following class session. Note that students still needed to submit their commented homework to the instructor after self- or peer-assessment, to check if they and peers had provided feedback on the work properly and correctly. For formative quizzes given at the end of a learning unit, self- or peer-assessment, with feedback, could occur after the class finished the quiz, sometimes in the same session or else in the next one.

In both self- and peer-assessment, students received input and guidance from the teacher, discussed with a neighbor, and then gave their comments and feedback on the work. Once that was done, students submitted the work with formative comments to the teacher.

Self-Assessment. An example of how homework was treated is in self-assessment given in Figure 3.7. This homework problem was given to students after they had learned the concepts and experienced concept checks. In this dynamics example, students were required for homework to make up a story or problem that involved Newton's first law, that is, to create a question concerning the case of an object moving with no force or no net force acting on it.

In revising, students could either add more information to complete the old answer or make a new story while the teacher gave feedback on each component or when they had discussion with peers. The teacher moved around the classroom to see if students needed help. Self-assessment took about 3 to 10 minutes, depending on the types of questions. Formative quizzes were treated the same manner as homework.

Task: Create a conceptual story for Newton first law, i.e., for a situation where an object is moving , with forces acting on it but no net force. Then explain it using a force diagram and including magnitude and direction of velocity and acceleration, if any.

1. Students first produced their own story for homework, for a specific moving object and situation of their choice, with explanation.
2. Teacher: "You have to think about the story and explanation step-by-step. Quantities that we need to consider—forces, net force, motion, velocity, and acceleration."
3. Students scanned through any previous work needed to focus on those components. In this way they had a chance to recall and focus on what they had done, before getting input and feedback.
4. Teacher: "First check the forces. In your story, are there any forces acting on your moving object? e.g., the pushing or pulling forces or friction, etc.?"
5. Teacher: If there are forces, is there a net or resultant force? If yes, it is not correct because this problem specifies no net force.
6. Students checked their work, discussed with peers, and corrected if mistakes were found.
7. Teacher: "Draw a force diagram for the object, showing and labeling any forces acting on it in this situation. (Waited for the answers, wrote them on the board, had feedback, and drew a force diagram.) Students checked/discussed and made comments to complete or correct their answers while the teacher walked around to help.

8. Teacher: “Next, check the velocity. What is the velocity of a moving object with no net force acting on it? (Waited for an answer, had feedback, drew on a force diagram.) Do you mention this in the story or show in a force and motion diagram that the velocity is constant? If yes, you’re correct. If not, modify it to get it right.
9. Teacher: “Now check about acceleration. What is the acceleration in this case? (Waited for an answer, feedback, and checked the force diagram.) Do you mention in the story or a diagram that acceleration is zero, i.e., that the object is not speeding up or slowing down, and why?”
10. Teacher: “Does it make sense that the object can be moving steadily in one direction even though there are forces acting on it, and even though the forces ‘cancel’?” Discussion.
11. At the end the teacher collected the commented homework and checked if they had perceived the work correctly and given appropriate feedback.. If not, written feedback by the teacher was needed, guiding them how to correct it; particular students might have to re-do it.
12. Teacher returned the work to students. Some students had to redo and submit their work to the teacher for a second check.

Figure 3.7. Self-assessment procedure example.

Peer-assessment. In peer-assessment, the basic processes were similar to self-assessment, except they performed the evaluation and feedback processes on a peer’s work rather than their own work. However, some aspects were required before the peer-assessment started. Self-assessment and peer-assessment were similar processes, but there were extra administrative procedural steps required in peer-assessment, as in the example in Figure 3.8.

1. Students wrote a 3-digit number on their work instead of their name. Only the teacher knew all the codes.
2. The teacher needed to collect all work first and then distribute it to the groups for peer assessment, and make sure that students got neither their own nor their peers’ work in the group.
3. At the end of the processes, students who assessed put their 3-digit code at the end of the problem to indicate who commented.
4. It took time since the teacher needed to check codes before the work was distributed to peers to make sure that students did not assess their own work or that of peers in the same group.
5. The teacher needed to check and record codes and names before returning commented work to students.

Figure 3.8. Peer-assessment administrative procedures.

Of course, these extra processes would not be required if peer-assessment was open rather than “double blind,” but openness creates its own problems when peers assess each others' work, for both personal interaction reasons and questions of expertise both ways.

Partly because of the extra administrative processes, the teacher preferred to use self-assessment more often than peer-assessment, both for convenience and for time, since the extra processes took almost 10 minutes to complete, affected teaching plans. Furthermore, self-assessment in itself is a valuable skill to develop and promotes a desirable degree of metacognition—self-awareness and self-management during and after learning, for specified goals.

This chapter has described the design and development of parallel formative and summative instructional units, the formative and summative systems, the assessments, and the implementation in instruction of these research units, assessments, and procedures, all as part of the project. The next chapter describes the research itself, the methods, and the collection and analysis of quantitative and qualitative data.

CHAPTER IV

RESEARCH METHODS

As stated previously, in the current project we aimed to investigate the comparative effects of formative or summative assessment systems on students' performance and attitudes in an introductory physics course. The project comprised development, teaching, and research aspects, with goals as follows.

Project goals

Instructional development and assessment goals:

- To integrate formative assessment and strategies into the instructional design of two physics topic modules.

Teaching implementation goal:

- To teach the modules in formative and summative modes as part of an introductory physics course.

Research goals:

- To conduct a study comparing student performance using formative or summative instructional systems.
- To survey student attitudes and preferences toward the two assessment modes.
- To describe how the formative system influenced teaching.

The way that the instruction and assessment goals and teaching implementation goals were accomplished has already been described in Chapter III. In the current chapter we describe the research design, the data sources, and analyses used to accomplish the research goals. This includes the quantitative comparative study of student performance, the surveys of student attitudes and preference, and the effects of the formative system on teaching.

Instructional Context for the Study

The research instructional context involved two 2-week topic modules on dynamics and kinematics in the Physics 1800 introductory physics course at Western

Michigan University. The 55 subjects were pre-service elementary education students in three separate sections (classes) of the course, with 15 to 24 students in each class. Two such classes in Fall semester 2012 provided control and treatment conditions for teaching the modules in both formative and summative modes. A follow-up class in spring semester 2013 was taught in formative mode only to test the reliability and reproducibility of the results obtained in that mode in 2012.

Implementation of the Instruction and Assessment in Formative and Summative Modes

The implementation of the topic modules in instruction in both modes, as a basis for collecting the data required in the comparative study, has already been described in Chapter III.

Experimental Comparison of Student Performance in Formative and Summative Assessment Systems

I wanted a quantitative experimental comparison of the performance of students who experienced formative and summative assessment systems for the two topic modules. I also wanted to know students' attitudes toward both systems, their preferences, learning behaviors, and the reasons they gave. Therefore, I conducted a mixed-method study to compare the effects of formative and summative systems on student performance, attitudes, preferences, and learning behaviors. Both quantitative and qualitative data were obtained.

In the "control" condition, the system was conventional as in past years, with mostly summative treatment of homework and quizzes, for which points and grades were allocated. In the treatment condition, the system was strongly formative, with concept checks during topic teaching and commented feedback on homework and quizzes, but no points or grades, and a chance to improve based on the feedback received. However, the final unit tests (module exams) for both conditions were summative.

Other than the formative or summative nature of the instruction and assessment, all other aspects of the units and instruction were kept as much the same as possible in the two modes, and all classes in both modes were taught by the same instructor. Students gained familiarity with both formative and summative methods earlier in the course, so they did not have to adapt to an unfamiliar condition during the study using the research modules. Since the formative condition spent somewhat more class time on assessment and feedback, other aspects of instruction needed to be shortened to take less time, thus were more concise or modified, such as teacher presentations and some of the demonstrations or activities. This was a natural consequence of changing modes and made for a fair research comparison with roughly equal time for both modes.

Structural Design of the Quantitative Comparative Study

The quantitative experimental comparative study was a central part of the project, to compare the effects of assessment mode on student performance, for each of the two topic modules. The research structural design aimed to control, as far as possible, for possible differences between treatment and control groups or conditions, or other factors that might possibly influence outcomes.

The structural design of the study used a crossover design involving two units, two student groups, and two assessment modes. It is most easily visualized when represented diagrammatically as shown in Figure 4.1.

Class	Earlier part of course	Pre-Test	Dynamics	Post-Test	Pre-Test	Kinematic	Post-Test
A	All get familiarity with both systems		Formative			Summative	
B			Summative			Formative	

Figure 4.1. The crossover design.

This design involved the following procedures, with classes switching modes between the two modules:

- *Dynamics Module:* Class A experienced the formative mode and Class B the summative.
- *Kinematics Module*. Class A now experienced the summative mode and group B the formative.

This design aimed to control for possible unknown or unmeasured differences between the groups; the groups were not deliberated randomly formed, but by students' own preference in registering for a particular section, or its availability, or whether it fit into their schedules. Thus, there was some element of random group formation and in fact there was no particular reason to expect the groups to be significantly different. Note that all students had experience earlier in the course with both assessment systems, so that they were all familiar with both for the comparative trials. The same instructor was involved throughout, and both classes experienced the same instructional activities, other than formative or summative mode, that the data collected depended upon. Thus, it would have been possible to obtain comparative research results even without the crossover procedure, but using it strengthened the design.

The structural design has another useful feature: each class experienced both the formative and summative systems, so we could survey student *preferences* after the experience. However, comparing the performance of one class *between the two modules* would not really be meaningful for our purposes, since one module may simply be more difficult or demanding than the other.

Assessment Instruments for the Dynamics and Kinematics Modules

- *Dynamics Module Assessment:* The pretest and posttest consisted of 23 multiple choice questions testing conceptual understanding (The instrument is in Appendix C).
- *Kinematics Module Assessment:* The pretest and posttest consisted of 7 open-ended questions, each with several sub-questions. Pre- and posttests were graded by the same instructor and were graded three times to ensure consistency. (The instrument is in Appendix D.)

Student Performance Measures and Analyses

Rather than simply using end-point performance scores as a measure of attainment in the module, we administered the assessment as pre- and posttests and determined gains in performance from pre- to post-instruction. The pre- and posttests involved both selected-response and constructed-response items, testing understanding of basic concepts and principles by the ability to apply them to problem situations. In addition, we determined normalized gains, which express gain as a proportion of the maximum gain possible given the pretest score. This allowed for possible variation in starting point knowledge between groups and between individuals. Normalized gains could be determined for the whole group and for individual students and for individual assessment items, so that data could be analyzed in a number of ways.

The formulas used for gain and normalized gain are as follows:

Gain = posttest score – pretest score

Normalized Gain = Gain / Possible gain

= (Post – Pre) / (100 – Pre)

Statistical significance of gains and of gain differences between modes was ascertained using simple *t* tests.

Surveys of Student Attitudes and Preferences for the Two Assessment Systems

We used open-ended surveys and interviews to find out about students' attitudes toward the two assessment systems and any effects on their learning behavior. Both target survey questions and free responses were used, so that we could both focus on points of interest and cue them in case students inadvertently did not think of them, while at the same time not limiting responses in advance. Useful issues and themes emerged in analyses from this qualitative information, including some aspects that we had thought of in advance.

Attitude Survey

The attitude survey, provided in Appendix E, consisted of 15 questions, both general and specific, about formative and summative assessment systems. It was given almost at the end of the course. The collected responses were reviewed and analyzed repeatedly until major themes emerged. The themes were grouped into categories, such as formative, summative, specific assessment techniques, course overall, etc. Besides the qualitative response data leading to grouping or clustering of responses, we did a semi-quantitative analysis within each cluster, giving numbers and percentages responding one way or another within each. We also extracted example quotations for each theme to illustrate typical responses and preferences.

Course and Instructor Evaluation

A formal course and instructor evaluation survey is given online for all university courses at Western Michigan University for students to provide their attitudes toward the course and instructor overall, as well as particular aspects of each. This evaluation survey uses a Likert rating scale ranging from 0 (lowest) to 5 (highest). The formal evaluation was not initially planned as part of the project, since the two research topic modules comprised only part of the entire course to be evaluated. However, the evaluation results differed unexpectedly between sections, as described in the results chapter, and are thus included in reporting.

In brief summary, the pre- and posttests provided data sources for the quantitative comparison of performance in the two modes. The student surveys and course evaluations provided qualitative and semi-quantitative data sources for student attitudes and preferences for the two systems. The qualitative survey data often suggested reasons to account for the quantitative performance differences between the two modes.

Further specifics of the procedures and analyses involved in both the quantitative and qualitative work are provided with the Results in Chapter V.

CHAPTER V

QUANTITATIVE AND QUALITATIVE RESULTS: STUDENTS' PERFORMANCE AND STUDENTS' ATTITUDES

This chapter provides the research results of students' performance and attitudes after experiencing formative or summative assessment systems in instruction in the dynamics and kinematics topic modules. Student performance data (learning gains) are obtained from quantitative scores on pretests and posttests. Attitudes and preferences toward the formative and summative systems, the course, and the instruction are obtained from qualitative survey data.

The results of the research will be presented in two parts: (1) Results Part A: Experimental Comparison of Student Performance in Formative and Summative Systems, and (2) Results Part B: Results of Attitude and Preference Surveys.

Results Part A: Experimental Comparison of Student Performance in Formative and Summative Systems

The research methods have already been described in a previous chapter, but a brief diagrammatic representation of the crossover research procedure is repeated here. The crossover study design is shown in Figure 5.1.

Class	Earlier part of course	Pre-Test	Dynamics	Post-Test	Pre-Test	Kinematic	Post-Test
A	All get familiarity with both systems		Formative			Summative	
B			Summative			Formative	

Figure 5.1. Crossover design.

Two Physics 1800 classes, Class A and Class B, both experienced formative and summative instructional modes in either the dynamics or kinematics modules; if a class experienced formative mode in one of these topic modules, it experienced summative in the other, since classes switched modes between the two topics, as illustrated. Prior to these two topic modules, students gained familiarity with both assessment systems on other topics earlier in the course.

The dynamics module data and analyes are described first and then kinematics. Results for each module include the data and results of formative and summative modes. Then the results for the two modes are compared.

Dynamics Module: Data, Analyses, Discussion and Conclusion

Formative mode. The dynamics module with embedded formative assessment was taught in Class A, involving 22 students, in Fall semester 2012. Before the module started, the class took the dynamics pretest, but 4 students were absent, so 18 took the pretest. The dynamics posttest, the same as the pretest, was administered with the final exam and none missed it. Thus, data from 18 students who took both pre- and posttest were usable and analyzed.

Table 5.1 gives students' raw scores on pre- and posttests, out of 23 points and as percentages, the calculated gain, the maximum possible gain for that student given the pre-score, and the normalized gain.

The class result for dynamics in formative mode was as follows:

Class pretest	35.5%
Class posttest	76.6%
Class actual gain	41.1%
Class normalized gain	0.64

The class gain was 41.1% between pre- and posttest. The normalized gain was 0.64, which is considered a high gain. Thus, substantial learning gains were achieved by students in formative mode in the dynamics module.

Table 5.1

Performance in Dynamics in the Formative System

Formative Assessment–Dynamics							
Participants	Pretest 23 points	Posttest 23 points	% Pretest	%Posttest	Actual gain %	Max possible Gain %	Normalized Gain
1	7	19	30	83	52	70	0.75
2	7	17	30	74	43	70	0.63
3	6	9	26	39	13	74	0.18
4	9	21	39	91	52	61	0.86
5	12	22	52	96	43	48	0.91
6	11	18	48	78	30	52	0.58
7	8	19	35	83	48	65	0.73
8	9	15	39	65	26	61	0.43
9	8	12	35	52	17	65	0.27
10	12	18	52	78	26	48	0.55
11	8	18	35	78	43	65	0.67
12	10	23	43	100	57	57	1.00
13	5	9	22	39	17	78	0.22
14	6	20	26	87	61	74	0.82
15	9	16	39	70	30	61	0.50
16	2	22	9	96	87	91	0.95
17	10	23	43	100	57	57	1.00
18	8	16	35	70	35	65	0.53
Average	**8.17**	**17.6**	**35.5%**	**76.6%**	**41.1%**	**64.5%**	**0.64**

Summative mode. The dynamics module was taught in Class B, in summative mode involving 15 students, in Fall semester 2012. Before the module started, the class took the dynamics pretest, but 2 students were absent, so 13 took the pretest. The dynamics posttest, the same as the pretest, was administered with the final exam and none missed it. Thus, data from 13 students who took both pre- and posttest were usable and analyzed.

Table 5.2 gives raw score of pre- and posttests, gain, and normalized gain, in forms of percentage of average class score.

Table 5.2

Performance in Dynamics in the Summative System

Summative Assessment-Dynamics							
Number of Participants	Pretest 23 points	Posttest 23 points	% Pretest	%Posttest	Actual gain %	Max possible Gain %	Normalized Gain
1	11	14	48	61	13	52	0.25
2	9	15	39	65	26	61	0.43
3	8	14	35	61	26	65	0.40
4	10	16	43	70	26	57	0.46
5	10	17	43	74	30	57	0.54
6	10	9	43	39	-4	57	-0.08
7	7	15	30	65	35	70	0.50
8	5	12	22	52	30	78	0.39
9	11	18	48	78	30	52	0.58
10	9	15	39	65	26	61	0.43
11	11	19	48	83	35	52	0.67
12	7	16	30	70	39	70	0.56
13	8	13	35	57	22	65	0.33
Average	**8.92**	**14.8**	**38.8%**	**64.6%**	**25.8%**	**61.2%**	**0.42**

The class result for dynamics in summative mode was as follows:

Class pretest	38.8%
Class posttest	64.6%
Class actual gain	25.8%
Class normalized gain	0.42

Comparing the class pretest and posttest, the gain for summative mode is 25.8%, and normalized gain is 0.42, which is generally considered respectable. However, it is much lower than the 0.64 value obtained for the formative mode.

Comparison of performance in the formative and summative systems. Results in the dynamics module for the formative versus summative modes are compared in Table 5.3.

Table 5.3

The Results of Formative vs. Summative Mode

Fall 2012	Dynamics	% Class pretest	*SD* pretest	% Class posttest	*SD* posttest	% Class actual gain	Class normalized gain
Class A	Formative	35.5	10.9	76.6	18.6	41.1	0.64
Class B	Summative	38.8	8.0	64.6	11.3	25.8	0.42

Note. *N* = 31 students. *SD* = standard deviation.

It is seen that the class pretest performance is similar for the two classes, but the posttest scores, actual gain, and normalized gain is much better for the class experiencing the dynamics module in formative mode.

Thus, students in the formative assessment class achieved remarkably higher learning gains than in the summative class. To check statistical significance. an independent samples *t* test was used in comparing class learning gains. The effect size was calculated as Cohen's *d* using a statistics website at the University of Colorado (http://www.uccs.edu/~lbecker/).

The statistical calculations give $t(28.061^*) = 2.834$, $p = .01$, and Cohen's $d = 0.96$. The *p* value of 0.01 indicates a statistically significant difference between students' performance in formative and summative modes. The effect size, Cohen's $d = 0.96$, means that the effect of formative mode (treatment condition) on students' performance is large compared to summative mode (control condition). Note that effect sizes greater than 0.8 are considered a large effect.

Therefore, in this research, students who experienced the instruction system with embedded formative assessments achieved substantially higher learning gains than those who experienced the existing conventional summative mode.

Kinematics Module: Data, Analyses, Discussion and Conclusion

Formative mode. The kinematics module was taught in Class B, involving 15 students, in formative mode, in Fall semester 2012. Before the topic module started, the class took the kinematics pretest, but 7 students were absent, so 8 took the pretest. The kinematics posttest was treated as the exam at the end of learning module and none of them missed it. Therefore, the data collected from 8 students, who took both pre- and posttests were usable and analyzed. It was unexpected and unclear why 7 students missed the first class after the midterm exam, making participant numbers lower than anticipated.

Table 5.4 gives raw scores in pre- and posttests, gain, and normalized gain.

The class result for kinematics in formative mode was as follows:

Class pretest	42.8%
Class posttest	83.9%
Class actual gain	41.1%
Class normalized gain	0.70

It can be seen that the class gain was 41.1% and the normalized gain 0.70. Based on both the gain and the normalized gain of 0.70, which is considered high, we conclude that the formative mode led to fairly high learning gains for the kinematics module.

Table 5.4

Raw Scores in Kinematics for the Formative System

Formative Assessment–Kinematics							
Number of Participants	Pretest 49 points	Posttest 49 points	% Pre-est	%Posttest	Actual gain %	Max possible Gain	Normalized Gain
1	23	38	46	78	32	54	0.58
2	19	49	39	99	60	61	0.98
3	9	27	19	55	36	81	0.45
4	14	43	28	87	59	72	0.82
5	31	44	62	89	27	38	0.70
6	6	45	11	91	80	89	0.90
7	32	42	65	85	20	35	0.57
8	36	43	72	88	15	28	0.56
Average	**20.9**	**41.1**	**42.8%**	**83.9%**	**41.1%**	**57.2%**	**0.70**

Summative mode. The kinematics module with summative assessment was taught in summative mode to Class A, involving 22 students, in Fall semester 2012. Before the learning module started, the class took the kinematics pretest, but 3 students were absent, so 19 students took the pretest. The kinematics posttest was treated as an exam at the end of learning module and none of them missed it. Therefore, data collected from 19 students who took both pre- and posttest were usable and analyzed.

Table 5.5 gives raw scores of pre- and posttests, gain, and normalized gain.

Table 5.5

Raw Scores in Kinematics for the Summative System

Summative Assessment–Kinematics							
Number of Participants	Pretest 49 points	Posttest 49 points	% Pretest	%Posttest	Actual gain %	Max possible Gain %	Normalized Gain
1	42	43	86	88	2	14	0.14
2	33	36	66	73	7	34	0.21
3	20	38	41	78	37	59	0.62
4	18	40	36	82	46	64	0.71
5	18	38	37	78	41	63	0.65
6	26	49	52	100	48	48	1.00
7	17	39	35	80	45	65	0.69
8	30	49	60	100	40	40	1.00
9	33	42	66	86	19	34	0.58
10	33	33	67	67	0	33	0.01
11	13	17	27	35	8	73	0.11
12	23	30	47	61	14	53	0.27
13	23	41	46	84	38	54	0.70
14	38	47	77	96	19	23	0.83
15	43	49	88	100	12	12	1.00
16	29	43	59	88	29	41	0.70
17	17	42	35	86	51	65	0.78
18	30	48	61	98	37	39	0.95
19	34	45	68	92	23	32	0.74
Average	**27.2**	**40.5**	**55.4%**	**82.6%**	**27.2%**	**44.6%**	**0.62**

The class result for kinematics in summative mode was as follows:

Class pretest	55.4%
Class posttest	82.6%
Class actual gain	27.2%
Class normalized gain	0.62

The class pretest score of 55% in kinematics for this class is higher than anticipated, indicating than students had some prior knowledge of some of the concepts involved, which is not surprising since the basic motion concepts of distance, time, and speed are taught at school. The posttest average was 82.6%, which is good, but note that since the class pretest score was 55%, the actual gain was 27.20%, making the class normalized gain 0.62. This is also good and comparable to the class normalized gain of 0.7 for the formative mode, though a little lower.

Experimental comparison of performance in formative and summative systems. Results in kinematics for the two modes are compared in Table 5.6.

Table 5.6

Formative vs. Summative Mode Comparison

Fall 2012	Kinematics	% Class pretest	*SD* pretest	% Class posttest	*SD* posttest	% Class actual gain	Class normalized gain
Class A	Summative	55.4	18.0	82.6	16.0	27.2	0.62
Class B	Formative	42.8	22.7	83.9	13.1	41.1	0.70

Note. SD = standard deviation.

The pretest performance of Class A was higher than class B, but working with learning gains and normalized gains takes that into account. The posttest score, the actual gain, and the normalized gain are better for Class B, the group experiencing the formative system. Note that this was the case for the dynamics module too, but the extent of the advantage over the summative group is not as substantial for kinematics as it was for dynamics.

Statistical calculations give t 25*) = 1.780, p = 0.344, and Cohen's d = 0.68. The p value of 0.344 indicates that the difference between students' learning gains in formative and summative modes could not be taken as statistically significant; this is likely due to the small numbers of students involved in the comparison. For this reason, student learning achievement for the two modes could best be described as comparable, or not too different, with both modes producing good gains.

Dynamics Module 2013: Data, Analyses, Discussion, and Conclusion

In Spring semester 2013 additional research data were collected. The purpose was to test with another class in another year to see if the high learning gains obtained in dynamics in formative mode in Fall 2012 were reliable and reproducible. For logistical reasons and availability of instructors, it was not possible to conduct another comparative study between formative and summative systems in 2013. However, it seemed important to repeat the formative system in instruction, since if similarly high gains were sustained, this would give further confidence in the system and the research results.

Formative mode 2013. The dynamics module was taught in Class C, involving 24 students, in Spring semester 2012. Before the learning module started, the whole class took the dynamics pretest, and the posttest was treated as a final exam, which none of them missed. Therefore, the data collected from 24 students who took both pre- and posttest were used and analyzed.

Table 5.7 gives raw scores of pre- and posttests, gain, and normalized gain.

The class result for the dynamics module 2013 in formative mode was as follows:

Class pretest	40.8%
Class posttest	77.2%
Class actual gain	36.4%
Class normalized gain	0.60

The actual class gain is 36.4%, and normalized gain is 0.60, which is considered fairly high.

Table 5.7

Performance Data for Formative Assessment Mode in Dynamics

Formative Assessment-Dynamics 2013							
Number of Participants	Pretest 23 points	Posttest 23 points	% Pretest	%Posttest	Actual gain %	Max possible Gain%	Normalized Gain
1	17	21	74	91	17	26	0.67
2	8	18	35	78	43	65	0.67
3	11	18	48	78	30	52	0.58
4	12	19	52	83	30	48	0.64
5	9	13	39	57	17	61	0.29
6	8	18	35	78	43	65	0.67
7	10	17	43	74	30	57	0.54
8	9	17	39	74	35	61	0.57
9	9	15	39	65	26	61	0.43
10	8	13	35	57	22	65	0.33
11	4	16	17	70	52	83	0.63
12	9	18	39	78	39	61	0.64
13	10	22	43	96	52	57	0.92
14	10	18	43	78	35	57	0.62
15	11	23	48	100	52	52	1.00
16	8	15	35	65	30	65	0.47
17	8	22	35	96	61	65	0.93
18	10	12	43	52	9	57	0.15
19	8	20	35	87	52	65	0.80
20	8	16	35	70	35	65	0.53
21	9	20	39	87	48	61	0.79
22	6	23	26	100	74	74	1.00
23	8	17	35	74	39	65	0.60
24	15	15	65	65	0	35	0.00
Average	**9.38**	**17.8**	**40.8%**	**77.2%**	**36.4%**	**59.2%**	**0.60**

Comparison of performance in formative mode in dynamics in 2012 and 2013. Table 5.8 compares Class A in 2012 with Class C in 2013, in regard to pre- and posttest scores, gains and normalized gains, for the dynamics module taught in formative mode.

Table 5.8

Formative Mode: Fall 2012 vs. Spring 2013

Class	Dynamics	% Class pretest	*SD* pretest	% Class posttest	*SD* posttest	% Class actual gain	Class normalized gain
Class A	Formative	35.5	10.9	76.6	18.6	41.1	0.64
Class C	Formative	40.8	11.5	77.2	13.5	36.4	0.60

Note. *SD* = standard deviation.

It can be seen that the pretest scores, posttest scores, gains, and normalized gains are similar between 2012 and 2013 for the dynamics module in formative mode. This is a satisfying result.

Statistical analysis gives $t(40^*) = 0.84$, $p = .40$, and effect size Cohen's $d = 0.26$. The p value of .40 indicates that the difference between learning gains of classes A and C, both using formative mode, is not statistically significant (mean difference 15.25, standard error difference 5.38). Therefore, the results for formative mode were essentially reproduced from one year to the next.

Discussion of Findings

This section discusses the results for both the formative and summative systems in the dynamics and kinematics modules, in relation to the characteristics of the two assessment systems, and considering reasons for the research findings. Specific aspects are treated under the headings below. Note that various of these discussion points were further elaborated or confirmed later by findings from the qualitative survey, as will be seen.

Feedback. Feedback was an important aspect of the formative system. Why should this be so effective in helping students' learning? From the study we suggest the following. Students understood what feedback was, how it functioned in formative mode, and what their own roles were. After they experienced it, they were aware of the learning benefits and many enjoyed it. Their appreciation of feedback helped create a condition that allowed students to assess their own understanding while learning, but also established a good class environment, collaborative and engaged. Students' performance in formative mode in dynamics and kinematics certainly showed higher learning gains than in summative mode.

By contrast, summative mode did not provide very much feedback to students except in the form of scores and grades. Without detailed feedback and only listening to the teacher presentations, explanations, and solutions, students became lost in the lesson more easily. The teacher provided direct explanations about the problem to help students, but this was not very successful compared to learning with a system that inherently included feedback. Summative mode is a relatively weaker system to support learning.

In the kinematics module, learning gains were good in both modes, a bit higher in formative, although the difference was not statistically significant. The reason for the similar learning achievement in the two modes might be related to the nature of the kinematics module and to the fact that considerable feedback did tend to occur in the summative mode instruction also, even if not planned for. Note that kinematics is formula-based to a fair extent, although our module also involved conceptual understanding, and kinematics assessment involves basic formulas, procedures, and calculations as well as some conceptual questions. During the teaching and learning of kinematics, students who experienced the summative mode brought up a lot of questions while the teacher was going over assigned problems and explaining the solutions. In this way, students were providing information to the teacher about what they did not understand, for example, why the direction of the car's acceleration pointed to the left and not the right. In response, the teacher naturally needed to provide feedback to the students until they asked no more questions. In this way, a fair amount of unplanned and feedback occurred, which, after all, is a natural characteristic of responsive teaching, whatever the mode. This effect might be one reason that the learning gains were comparable between modes. We note again that the formative gains were actually a little greater, but research with larger numbers of students would be required to test this with statistical significance.

Students experiencing the summative mode in kinematics, and themselves requesting more feedback, can actually be seen as confirmation that formative feedback, even unplanned, is effective. Note that all formative components—the characteristics, techniques, strategies, teacher, and students—need to work well together to create a productive learning environment for the system. In the case of summative mode in kinematics and when working problems, students played a role by providing feedback to the teacher that they needed help. While this is likely a factor in the similar gains obtained in formative and summative modes, we note also that the rather different inherent nature of the dynamics and kinematics modes, and of the aligned assessments, could also be an important factor. The possible effect of the nature of different topic modules is discussed in a separate section.

Opportunity to relearn, make corrections, and improve. Several characteristics and strategies support the formative mode in improving students' understanding and performance. Relearning is one of these. Students all come to class with different prior knowledge, so each one grasps a concept differently or to a different extent. Some students are able to succeed the first time. However, if not, and if they figure out how to improve, they become good learners and successful (Wiggins, 1998). Formative mode supports students in doing this.

A second chance to learn improves students' conceptual understanding. Learning a concept the first time may be difficult, and when students try to use a new concept to tackle a problem, they may not succeed at first. Learning another time, especially when this addresses a particular issue or mistake that occurred the first time, is a good way to improve. When students gain a stronger understanding of one concept this way, they also have better chance to master the next relevant concept as well. In formative mode, especially for conceptual dynamics, relearning or enhanced learning was simple and fast; students saw another situation involving the similar concept almost all the time, in the form of concept check MCQs. Also, students were engaged and challenged most of the time and not bored. In kinematics, a module of somewhat different nature, relearning required students to re-do a problem (open-ended written response homework or quiz) most of the time. We suggest that providing sets of concept checks on small, specific aspects of concepts and problems is productive, and that the second chance for improving understanding is more effective when the teacher gives the opportunity to students right after the first try. In this way, students get feedback immediately on the same concept twice, which may help them understand it more firmly. Therefore, we suggest that relearning should occur right away or as soon as possible after the first try.

By contrast, summative mode does not normally provide for students to improve their summative score or grade by relearning, so they have no formal chance to try again and improve. Furthermore, a poor score or failure at the start, which is unalterable in summative systems that accumulate scores, may affect students' confidence for the entire course if they think they are not able to achieve what is required. The system does not provide a second chance to relearn the concept, yet many students may not understand it soon enough, so the consequence is punitive: they get a bad grade rather than being able to improve to a better one.

Therefore, relearning in formative mode is another characteristic that helps to elevate students' performance.

No grades and less pressure. In formative mode, when no grades were involved, students experienced less pressure and were able to focus better on learning. When they understood just a partial concept, they were not afraid to try and use it. In this way, true feedback based on their current knowledge allowed the teacher to realize to what extent they grasped the concept and thus to address their needs. Although they had not mastered the concept yet, they still did not worry about getting a bad grade in formative homework and quizzes. They learned with more peace of mind. In contrast, summative mode creates a lot of pressure, with various effects on learning. Students learn with anxiety and worry that they are going to get a bad grade. No grades *during* learning is an appropriate strategy for most students, especially those who do not learn or do well under pressure. On the other hand, without grades, some students do not take learning and studying seriously. They lack motivation. This is a valid argument against having no grades at all. Note, however, that in this project we had summative final exams for a topic module even when formative assessments were used during the module.

In contrast, summative mode with a system of points and grades for homework and tests creates external pressure for students, and students may focus on grades more than on conceptual understanding in the topic. However, grades can provide strong extrinsic motivation to work hard and do well.

Motivation. Motivation in both modes will be a mixture of intrinsic and extrinsic motivations. However formative mode tends to be balanced more toward intrinsic motivation—interest in the subject for its own sake and a desire to understand it—and the mode provides opportunity to do so. Note that extrinsic motivation, the desire to get a good grade, will also play a part in a formative system, in that students realize that eventually they will likely get a better final summative grade. In formative mode, when students realize the benefits, they are keener to participate in learning, for example, in having more discussions. Especially when they do not understand something the first time, they like to be able to discuss, try again, and succeed. This also increases their confidence and self-esteem. In summative mode, when students are grade-conscious, intrinsic motivation during learning may be lower, and this may decrease further if a student gets a poor grade. Nevertheless, although a poor grade upsets students, it can, on the other hand, motivate and spur many of them to work harder. Thus there are multiple sides to consider for motivation in both modes of assessment.

Nature of topic modules. Different topics may be very different in nature. The nature of the two topic modules in this study, dynamics and kinematics, was clearly different. The dynamics module was largely conceptual rather than formula-based, and the instruction, materials, and assessment reflected that. The idea was for students to develop qualitative physical insight into force and motion situations and be able to answer conceptual questions and problems. Specially constructed selected-response assessment items were valuable for this, since the options could reflect various common alternative conceptions about physical situations as well as the desired scientific concepts and principles. Such questions were also appropriate for formative discussions because students could consider alternative ideas and discuss each case. The results in dynamics on tests of conceptual understanding were very successful for formative mode compared with summative.

The kinematics module, on the other hand, was of different character; besides conceptual understanding of the kinematics concepts, it also focused on algebraic expressions and formulas, formula-based solutions, and quantitative calculations for motion problems. The teacher noted that the majority of students in kinematics, in both modes, had no particular problems in understanding the basic concepts of distance, speed, and velocity, these being relatively simple, and they would also have learned them at some level before enrolling in this class. However, the module did extend the topic to acceleration and to motion problems for accelerated motions, which are more difficult both conceptually and mathematically. Structured written problems were used for students to apply motion concepts to new situations, and the teacher would provide explanations. The results showed that students were not easily able to use the concepts to solve such problems without a lot of practice, in both modes. Note that practice can be seen as constituting a type of formative feedback, whatever the intended mode, so this may be another reason for the comparable learning gains in formative and summative systems, when students get to practice on problems and discuss solutions.

Therefore, the nature of the topic itself, and the qualitative or quantitative approach in the module, seems to influence students' learning performance response to the formative or summative nature of instruction and assessment. In this study, the formative mode worked substantially better than summative for a conceptual module (dynamics), but both modes worked about equally for a more formalistic module (kinematics). It would be useful to explore this issue further in future research.

Results Part B: Results of Attitude and Preference Surveys

In this section, the results of the attitude and preferences survey are presented. One of purposes in the research was to learn about the attitudes of students to the formative and summative systems they had experienced, and their preferences, as well as the reasons they gave. The attitude survey in Appendix E was created for this purpose. The questions are categorized according to various aspects of the formative and summative assessment modes.

This section on the surveys includes the following aspects:

- Student understanding of the two different assessment modes.
- Overall preference of students on mode of assessment.
- Attitude toward formative assessment.
- Attitude toward summative assessment.
- Attitude toward particular assessment techniques.

Survey responses are presented for each of these aspects, in terms of percentages expressing various attitudes and with illustrative quotations from students, giving their comments and reasons.

Student Understanding of the Two Assessment Modes

Using the survey, we first wanted to find out whether students understood both modes adequately and then survey their attitude toward each mode overall, as well as toward the various aspects and strategies of each. We could also evaluate their overall understanding of both modes from analysis of their accumulated responses to the various questions on the survey. The following results present the percentages and numbers of students who understood, or did not, each type of assessment generally:

Good understanding of formative assessment	54 students	98%
Good understanding of summative assessment	54 students	98%
Weak understanding of both types	1 student	2%

Thus, almost all students understood what formative assessment and summative assessment are, presumably from class discussions of the two and from experience with each. Only one student lacked this basic understanding.

Almost all of the students understood that formative assessment enables them, as well as the teacher, to recognize if students are grasping the ideas and concepts while learning. If not, they get feedback to correct and improve. In addition, they are not grade-conscious or stressed. The clarification above is brief, while the details of each aspect of formative assessment are elaborated in a later section.

The overall picture of summative assessment from students' point of view is briefly as follows: almost all students understood that the system provides a grade for homework, quizzes, and exams, and there is grade pressure; and that there is little detailed feedback in summative assessment, and no opportunity to re-do.

Therefore, the majority of students revealed that they broadly understood each type of assessment at a satisfactory level before responding to the attitude survey. Only one student had weak understanding.

Preference of Students on Overall Mode of Assessment

Students having experienced both modes were able to express their preferences in response to survey questions about which mode suited them best for learning. The preferences are presented as follows:

Prefer formative assessment throughout	8 students	15%
Prefer summative assessment throughout	4 students	7%
Prefer a mix of formative and summative assessment	43 students	78%

Most students preferred a mix of formative and summative assessment. They responded that each one has its own nature that can benefit learning and achievement. Formative assessment creates less pressure and includes formative aspects, such as feedback, and summative assessment includes a summative grade and exam, etc. Ultimately students preferred formative assessment overall, but also wanted a final exam. Examples of students' comments are as follows:

> They are both useful in assisting my learning.
>
> Students need to have both types of assessment during the semester.
>
> Both, because students are able to get points still from the summative assessments, and with formative assessments students are able to get feedback and I feel that it benefits their learning.

> I would say both, but mostly formative. I think the only thing that should be summative are tests at the end of the unit.

About 15% of student would prefer only formative assessment throughout. They respond that formative assessment involving many strategies and being less stressful helps their learning. Examples of students' comment are as follows:

> I prefer formative assessment. The strategies are very helpful!
>
> I do better in formative assessment because that's just the way I am.
>
> Formative assessment is less stressful.

Only about 7% of student would prefer only summative assessment throughout. Examples of students' comments on summative mode are as follows:

> It requires the students to take time after class and make sure they understand the material.
>
> It gives every assignment a purpose.

Attitude Toward Formative Assessment System

The survey results of formative mode cover many aspects of the system: receive feedback, chance to relearn, correct and improve, no grades, know objectives, discussion, and teacher gets feedback to improve instruction. Each such aspect of the student responses is discussed in some detail below.

Attitude toward feedback. As feedback is used often and embedded in all types of formative strategies, it is imperative to recognize if students like it and think it benefits their learning. Results are as follows:

Like feedback generally	55 students	100%
Like feedback while leaning	53 students	96%
Like feedback later, not while learning	2 students	4%

All students liked feedback generally. Most of these (96%) students like it during learning, but 4% of students prefer it not while learning.

From the students' point of view, the nature of feedback consisted of several categories presented below.

- Receive immediate feedback
- Recognize if understand the concept.

- Learn from mistakes.
- Improve learning immediately.
- Let students know progress.
- Every learning aspect includes feedback
- Discussion with other students and the teacher.

Students like all the above aspects from their responses. Examples of students' comments are follows:

> I like the feedback because it tells me what I am excelling at but also what areas I need to improve.
>
> If we don't understand something, we know immediately and can fix misunderstanding right away.

Chance to relearn, correct, and improve. This is another aspect of formative assessment that students feel benefits learning: the chance to relearn, make corrections, and improve.

Like chances to relearn, correct and improve	53 students	96%
No response	2 students	4%

Virtually all students favored relearning, making corrections, and improving, with only 4% of students not responding. The main ideas they mentioned as reasons are:

- Revise what you did wrong to get better.
- Give enough time to improve learning.
- Master concepts before the exam.
- Formative helps students prepare for summative assessment.
- Do better next exam.
- To receive a good grade in the coming exam.

Students think that formative assessment gives another chance to revise mistakes and so are able to get a good grade in the future. In this way, they use formative assessment to prepare for the summative tests. Some students' comments are as follows:

> It gives another chance to apply knowledge to correct the mistake. It improves my learning before the test.
>
> It allows me to acknowledge my mistakes and correct them so improves my learning.

Attitudes toward no grade aspect. Formative assignments did not include a grade, which is different from a traditional course. This is how students felt about the no-grade aspect in relation to pressure:

Feel no pressure	47 students	85%
Still feel pressure	7 students	13%
No response	1 students	2%

When formative work does not involve a grade, about 85% of students feel little or no pressure. They respond that if grades are not involved, they do not feel that pressure of being penalized if they make a mistake, as they are with summative homework and quizzes. Therefore, they are able to focus more on learning. However, about 13% of students still feel pressure. They respond that although it does not include a grade, they still want to do well. Last, only one student, 2%, did not respond to this aspect.

The main ideas students mentioned in regard to the no-grade aspect are as follows:

- Not afraid to answer incorrectly.
- No external pressure.
- Only focus on learning the concepts.
- Method: formative homework and quiz do not include grade.
- Affects learning in positive way.
- Not getting penalized.
- Still feel some pressure to do well in the concept.

Examples of students' comments are as follows:

> It is not graded so even if you don't understand the concept, you can answer the best you can.
>
> I don't feel that much pressure and it allows me to learn better.
>
> We can get our understanding without being afraid to answer incorrectly.
>
> I don't like to do poorly on anything, so even though I am not getting point I want to make sure I know the material and that I am not "falling behind" in class.

Knowing/discussing the objectives and/or focus questions. Students are made aware of learning objectives for a section or activity and/or the pertinent focus questions, thus providing the purpose and desired outcome and indicating which main issues they need to focus on. This was not explicitly asked in the survey, so only a few mentioned it spontaneously, as follows:

Like discussing learning objectives	6 students	11%
Not mentioned	49 students	89%

Those who mentioned it (11%) responded that they like knowing the objectives and focus questions. Examples of students' comments are as follows:

> Discussing learning objectives are the most helpful because I learn through not taking a lecture.
>
> Discussing the objective helps me get a better grasp on the concept.

Leads to more discussion. In formative mode, discussion is a natural aspect and students have opportunities to do that within a group and within a class. The survey did not ask specifically about this aspect, and most participants did not bring it up spontaneously, except for one student who mentioned it in the open-ended survey. He stated:

> Formative leads to more class discussion

Feedback to the teacher. Feedback helps the teacher realize if students understand the concept. Students' responses about this aspect were:

Feedback to the teacher	14 students	25%
Not mentioned	41 students	75%

The survey did not ask specifically about this aspect, and thus many students did not bring it up, except for those students (25%) who mentioned it in an open-ended survey. They stated that formative mode creates feedback to instructor. This enables the teacher to determine if the lesson plan is to be adjusted to meet their needs.

An example of students' comments is as follows:

> The instructor knows where to move on to the next topic or stay and discuss the current topic more.

Attitude Toward Summative Assessment System

The student survey results relating to instruction using summative assessment involved various aspects: summative exams, points, and grades; work counting toward grades; pressure; motivation (extrinsic); being a requirement of courses; test anxiety; and summative while learning. Each of the main aspects of student responses is discussed in some detail below.

Attitude toward summative exams. At the end of a learning module, students took a summative exam. Their survey responses on this aspect were as follows:

Like an exam	14 students	25%
Do not like an exam	5 students	9%
Not mentioned	41 students	75%

Note that the survey did not ask specifically about this aspect, so those who mentioned it did so on their own in the open-ended part of the survey. Some students (25%) mentioned in the open-ended survey that they liked an exam, but others (9%) did not like it. Surprisingly, the majority of students did not mention this aspect of their own accord; only a few did. The students' point of view is elaborated in several categories, both pros and cons, as below.

Pro points mentioned were as follows:

- Measures understanding: students respond that at the end of a learning unit they feel accomplished, and having an exam at the end of learning unit is appropriate because students achieve in the concept already.
- Exam is not difficult and worth a lot, so students can earn a lot of points.

Examples of students' comments are as follows:

> I like an exam because it is the end of the unit and I am usually very confident in the material.
>
> When I get the exam back with my grade on it. I feel like I learned or accomplished something.
>
> For tests, yes it is a good way for me to see what I know at the end of a unit.
>
> It tests my overall ability on what we've learned.
>
> I like the exam, not too difficult and worth a lot of points.

Con points mentioned included:

- Worth a lot of points and stress: students respond that an exam is worth too many points compared with other summative tests, e.g., quiz or homework, that causes it too be stressful.
- No chance to restudy: A grade given to students cannot be chanced; that means there is no chance to restudy.

Examples of students' comments are as follows:

> I don't like the exam because they are worth so many points.
>
> We don't have time to go back over concepts that I don't understand because we move on right away.

Attitude toward counting toward a grade. In general, students are familiar, from almost all other courses, that assignments and tests include grades. In this research, students experienced both summative and formative modes, the latter with no grades during learning, and thus different from other courses. How these experiences of the two modes affected their attitude about counting toward grades is reported below:

Like to receive grades for every assignment	31 students	55%
Do like to receive grades, prefer formative mode	22 students	12%
Not mentioned	22 students	12%

About 56% of students liked receiving grades for every assignment. These students preferred to receive grades for various reasons that are described below. In another group, about 12% of students responded that grades did not matter, that learning is more important (preferred formative mode). In the last group, about 12% of students did not respond. The students' points of view, both pro and con, regarding grades are as follows:

Pro points mentioned were as follows:

- Grades reward the effort in learning: students feel that when they work hard, e.g., homework, they should receive a grade as the reward for their effort.
- To see if students understand the material: they would like to have a grade after the learning unit to check their performance.

Examples of students' comments are as follows:

> It gives me points to reward me for what I learned.
>
> I like that I have a chance to earn more points and raise my grade.
>
> I get a grade, and that grade gives me an understanding of how well I am doing in the class.
>
> I like that we have some summative assessments because that builds a grade.

Con points mentioned were as follows:

- Focusing on a grade instead of learning the material does not help while learning the concept: students feel that counting toward a grade while learning causes them to focus more on the grade than learning the concept.
- Not yet fully understanding a concept, but being penalized: they feel that during learning a grade should be not involved because they are trying to master things and if they are not ready yet, they will get a bad grade. They feel they are being penalized and mention that grades do not help while learning.
- "Pay a load of money to attend school, so grades are important": because he/she pays a lot of money for credits in school, getting a grade is important.

Examples of students' comments are as follows:

> I do not prefer summative assessment. They had too much focus on points than the material itself.
>
> I am too focused on the point aspect of it, taking away from true learning.
>
> I don't fully understand the questions on homework and I get marked down because of it.
>
> I never do well on summative. I know because of the pressure of getting an A.
>
> Our grades are tied to point totals, and so our grades turn into stressor.
>
> I want the best grade and I miss things because I am too focused on that.

Attitude toward pressure. Pressure is another aspect that usually accompanies the summative aspect. The students responded as follows to this aspect:

Feel pressure	53 students	96%
Do not feel pressure	2 students	4%

Thus, almost all students felt pressure while experiencing summative assessment. The students' points of view about pressure were categorized as follows:

- External pressure (grade): students respond that pressure is caused by grades, especially when they want or need a good one.
- Nervous: students are nervous to get a bad grade.
- Concern about failure.
- Takes away from the learning experience: students feel that whenever they have pressure, they are not able to focus simply on learning.
- Depend on summative situation: students say that only particular types of summative assessment, e.g., quizzes, exams, causes them pressure.

Examples of students' comments are as follows:

> It is do or die right off the bat. It affects my learning by not understanding what I did right or wrong.
>
> There is a lot of pressure on me. If I don't understand the material, I would have more anxiety about assessment.
>
> I feel pressure to do well which makes me nervous due to test anxiety. I even when I know material I get nervous.
>
> If I fail then I may fail the class and have to take it again. It makes me feel frustrated.

Summative provides extrinsic motivation. The nature of summative assessment can provide (extrinsic) motivation to study.

Summative provides motivation	22 students	40%
No response	33 students	60%

About 40% of 55 students responded that summative assessment motivates their learning. However, about 60% of students did not mention this aspect.

The students' points of view about motivation in summative assessment are categorized as follows:

- Grades in summative assessment motivate students to study harder.
- Receiving a bad grade can motivate students to study harder.
- Earning a few points from a correct answer in homework or a quiz makes students satisfied.

Examples of students' comments are as follows:

> I feel more motivated to study and actually learn the material.
>
> I receive bad grades it encourages me to study harder.
>
> I am pushed to make sure I know an answer.
>
> I do feel some pressure with the quizzes, because even though I am not being graded, it still feels like a quiz. I like to feel a little pressured, it is motivation.
>
> I know that if there is no grade, there is no reason to keep working so hard.
>
> If it’s not for a grade, I don’t have to get it right. I just have to get it done.

Summative assessment is a requirement of courses. Almost all courses in school or colleges usually involve summative assessment. Only one student mentioned without being asked an explicit question on the attitude survey. However, the majority of students did not mention it spontaneously. The student comment is the following:

> Summative assessments are what I have had all through school.

Summative assessment creates test anxiety. Although students may study hard, and perform well while learning and understanding the material, they may not do well in the summative assessment if they have test anxiety.

Exam creates test anxiety	1 student	2%
No response	54 students	98%

Only one student mentioned this spontaneously without being asked the question on the attitude survey. The student’s comment is the following:

> I lost the point, even though I knew it in the first place.

Summative assessment does not help when trying to master the concept. After experiencing summative assessment, a few students spontaneously mentioned the idea that summative assessment is not good while learning. The others did not raise the issue. An example of a student’s comment is the following:

> I think it is best to have summative for exam, but I think it is not as helpful as formative for homework and quizzes because we are still working to master information.

Attitude Toward Particular Assessment Techniques

Several strategies and techniques used were embedded in the formative and summative modes. This section clarifies how these strategies and techniques affected students' attitude and learning.

Attitude toward peer-assessment. Peer assessment was regularly used in formative mode at the beginning, especially when homework was discussed in the class. Students needed to grade and comment on a peer's work. Attitudes toward peer-assessment were as follows:

Peer assessment is helpful for learning	20 students	36%
Peer assessment is not helpful for learning	34 students	62%
No response	1 student	2%

About 36% of students responded that peer-assessment provides chances to discuss and get feedback with peers, pressures them to do well on the work, and is helpful if the instructions are clear. These are considered as pro points of peer-assessment. However, about 62% of students responded that they were concerned about the quality of peer feedback and comments. In addition, some noted they were not able to recognize and deal with their own mistakes immediately, as peer-assessment is not self-assessment. These are considered as con points of peer-assessment. Only 2% of students did not respond.

The students' points of views, pro and con, are as follows.

Pro:

- Discussion with peers (think-pair-share): students responded that they have chances to discuss problems with peers while grading, so they learned about concepts and solutions from each other.
- Peer's feedback helps to see a different view of the work: students responded that a peer is able to notice and give feedback on what is wrong on the work that they did not see themselves.
- If quality of peer's assessment is similar to teacher's it is helpful: students responded that assessing or comment by peer is useful if it is done as well as the teacher.

- There is pressure to do well on the work because someone sees it: students responded that peer-assessment pressured them to do well on the work as others would see it.

Examples of students comments are as follows:

> They are helpful in hearing others perspectives.
>
> Sometimes I can understand material better if a classmate explains it.
>
> It is nice to see how others are doing in the class as well as your own progress. You notice what you did right and wrong by your partner noticing things you did not notice.
>
> If done correctly though, it could be helpful as teachers to get experience grading/ correcting papers.
>
> I sometimes feel pressured to do good on the work.

In another group of students' responses, students who did not like peer-assessment aspects mentioned the following issues.

Con:

- Concern about quality of discussion and comment. The students responded that a peer may have not enough knowledge to assess or comment well.
- Instructions for self-and peer- grading should be clear: students needed more instructions for grading.
- Peer does not put enough effort into the work: students responded that a peer should focus more on feedback and comment.
- Does not help learning: Students responded that peer-assessment is not helpful.
- They cannot know what it is wrong with their own work right away because they are grading peer's work.

Examples of students comments are as follows:

> The peer might not know the material or how to make proper correction, so when you get it back you might not be able to lean from the assessment.
>
> The peer doesn't always know enough to grade someone else's homework and their feedback is useless.
>
> Peer assessment should not be kept unless there are more guidelines to grading.

> I think it is good to discuss work with peers, but it has to have a focus and the students need to take it seriously.
>
> I don't like peer-assessment.
>
> I don't know what I did wrong right away.

Attitude toward self-assessment. Students regularly did self-assessment when, for instance, homework was discussed. Self-assessment processes are very similar to peer-assessment, except involving self-assessing and commenting. The students' attitudes toward self-assessment were follows:

Self-assessment is helpful for learning	40 students	73%
Self-assessment is not helpful for learning	7 students	13%
No response	6 student	12%

More than a half of students responded that self-assessment is helpful; however, some of them think it is not. Twelve percent of the students did not respond.

The students' points of view, pro and con, were as follows:

Pro:

- Feedback: students responded that they were able to get feedback about their work right away.
- Improve learning: some students responded that self-assessment is useful for learning.
- Easy to grade and check my own work.
- Recognize what is wrong by themselves.

Examples of students' comments are as the follows:

> I like self-assessment.
>
> Feedback right away is good because then I don't get the wrong idea for other questions.
>
> It allows you to understand your progress or mistakes on assessments and correct them if needed.
>
> You will learn the material better, when you yourself actually put it on the paper and comment it.

Con:

- Prefer instructor feedback: students concerned about their own grading, whether it is good enough, so that they prefer to have the teacher's feedback and comment instead of self-assessment.

Examples of students comment are as follows:

> I would rather have a teacher grade it, so I know exactly what they look for on the assessment.
>
> Not sure about their own judgment if the teacher will accept it.

Attitude toward formative quizzes. Formative quizzes are used immediately after a learning unit. Students have an opportunity to improve. Their attitudes toward formative quizzes were as follows:

Like formative quizzes	18 students	33%
Do not like formative quizzes	2 students	4%
Not mentioned	35 students	64%

About 33% of students mentioned they liked formative quizzes without being asked an explicit question on the attitude survey, but about 4% of students did not like them because they still felt pressured. Last, about 64% of students did not metion this aspect spaonaneously. The students' points of view, pro and con, were as follows:

Pro:

- Students are able to apply their knowledge without being penalized and are able to learn more and able to prepare for the real quiz.

Examples of comments are as follows:

> I like formative quizzes because it's another way to apply what we know and not have to worry about being penalized for what we don't know.
>
> I like formative quizzes because it feels as if I am taking a quiz for a grade so I concentrate more and have a feel for what will be on the exam.

Con:

- Some students still felt pressured.

An example of a student comment is the following:

Formative quiz: the pressure still feels like a real quiz.

Attitude toward concept checks. Concept checks are conceptual questions that students are able to respond to by using a clicker (hi-tech) or raising a hand (low-tech). Students' attitudes are as follows:

Enjoy concept checks	48 students	87%
Useful but prefer without also drawing diagram, etc.	4 students	7%
A clicker is not necessary	3 students	5%

About 87% of students responded that they enjoyed experiencing concept check including feedback and clickers. About 7% of students responded that a concept check is good if it is not required to also draw diagrams. Last, about 5% of students responded that a clicker is not necessary in a class.

The students' points of views, pros and cons, are as follows.

Pro:

- Feedback: students responded that they are able to get the feedback from the teacher and the class immediately.
- Help to understand the concept: students responded that the concept check helped them learn the concept by giving the main concept to test if they understood.
- Clicker is an interactive way to provide feedback: Students responded that concept checks combined with the clicker make them enjoy learning in class.
- Should be used a lot more: students responded that they enjoyed learning with the concept check involving clicker, so the teacher should use them more in the class.
- Anonymity: students responded that a concept check combined with the clicker is anonymous.
- Assess current knowledge while learning: students responded that concept checks help them to assess for themselves if they understand a concept
- Feedback to teacher to recognize if the majority of the class understands the concept: students responded that concept checks create feedback to the teacher, who is able to see if they grasp the material.

- The teacher adjusts the lesson plan to students' needs: students responded that the feedback to the teacher enables adjusting the lesson plan to address students' needs.

Examples of students' comments are as follows:

> I love having clicker to do concept check because they give back instant feedback.
>
> Clickers are good because it helps us to see how the questions will be asked to us and main concept we should understand.
>
> I love doing the clicker.
>
> I really did enjoy using the clicker in this class.
>
> Allow students to be anonymous when answering. I think they should be used a lot more.
>
> It is a good way to know if your students have grasped what you taught.
>
> I also really like clicker because the instructor knows where to move on to the next topic or stay and discuss the current topic more.

Con:

- Concept check is fine without clickers: some students responded that they did not need to use clickers in class, as there are other ways to respond to the teacher.
- Drawing diagrams does not go well with concept check: the clicker should not be used with concept checks requiring drawing a diagram.

Examples of students' comments are as follows:

> As far as the clickers, we did just fine all semester, so I do not see why you need them in the first place.
>
> Clicker is a horrible thing, and most of the time they ask for a diagram.

Attitude toward summative homework. Students are given homework for every learning aspect, and their attitudes overall are as follows:

Like homework	11 students	20%
Do not like homework	4 students	7%
Not mentioned	40 students	73%

Some students mentioned they liked homework without being asked an explicit question on the attitude survey, and others did not. However, the majority of the class did not mention this aspect spontaneously. The students' points of view, pro and con, are as follows:

Pro:

- Practice the concept: students are able to practice the concept by doing homework.
- Help to prepare for summative test: homework helps them to prepare for quizzes and exams.
- Homework is not worth many points: students responded that homework is small portion of a total grade.
- Being able to use notes in homeowrk: students can use other materials, e.g., notebook to help them to solve problems.

Examples of students' comments are as follows:

I like homework because it is not worth that many points.

I like homework because it helps me practice the concept.

It helps you better prepare yourself for quizzes and exam.

This gives us a chance to use help from notes or class with topic we do not understand.

Con:

- Homework should be formative: students responded that homework in the class should be treated as formative. Did not provide reasons for that.

Examples of students' comments are as follows:

I don't like homework because I would rather have those things be formative.

I don't like homework.

Attitude toward summative quizzes. Students have a quiz at the end of a learning unit combining many aspects. A quiz is not worth many points compared to an exam. Student attitudes to the quizzes were:

Like the quizzes	14 students	25%
Do not like the quizzes	7 students	13%
Not mentioned	24 students	62%

Some students (25%) mentioned they liked the quizzes without being asked an explicit question on the attitude survey, but others did not. However, the majority of students did not mention this aspect.

The students' points of view, pro and con, are as follows.

Pro:

- Similar to an exam: students responded that quizzes are similar to the exam. In other words, it helps them to prepare for the exam.
- A quiz is not worth many points compared to the exam.
- Still have time to study following a mistake in a quiz before an exam. Although they make mistakes in a quiz, they still have time to restudy that topic before the exam.

Examples of students' comments are as follows:

> Quizzes are my favorite because they most closely resemble the exams.
>
> I like quizzes that are summative because we still have time to go over concept I don't understand that were on the quiz.
>
> They are not weighted as much as exams.

Con:

- Concern about being penalized by receiving a bad grade if they have a quiz in between learning the concept and if they yet do not understand.
- Prefer formative quizzes. Students responded that the teacher should use formative quizzes in class.

Examples of students' comments are as follows:

> I don't like quizzes because it is way too much pressure mid-way through a unit and if I don't understand something yet. I don't want to get penalized for that.
>
> I don't like quiz because I would rather have those things be formative.

Discussion of Survey Findings

This section discusses the findings of the survey of students' attitudes toward formative and summative assessment systems.

Formative and summative modes. For the survey to be meaningful, it was important that all students had experienced teaching and learning in both formative and summative modes, and understood both systems and the differences between them, before taking the attitude survey. In the study, both modes were explained and students experienced both in various topics from beginning of the semester. Their understanding was also ascertained from analysis of their written responses to the various survey questions and the open-ended parts. Therefore, we have evidence that students took the attitude survey with an understanding of the nature and functioning of the formative and summative assessment systems.

The student reflections on formative and summative systems provide support to the idea that the feedback aspect of formative assessment is particularly important for learning. In formative mode, prompt feedback is the most popular characteristic mentioned, because all students liked having feedback on their work and ideas, with reasons, and being able to discuss with the teacher and other students and thus being able to improve immediately. Only a few students disagreed that immediate feedback was the most useful: they mentioned that delayed feedback was better because students needed to understand and distill new knowledge and concepts one or two days before getting feedback. These students may not quite appreciate that immediate feedback helps to assimilate the new concepts during learning. In future, the teacher should talk about this issue to students, explaining how that immediate feedback is better during learning than delayed, as they are able to access and improve their understanding right away while fresh, rather than waiting for one and two days to understand it. Both this study and and previous research finds that feedback benefits learning substantially, and in our study all the students themselves agreed with that, so we conclude and advocate that significant feedback should be embedded in all courses and instruction.

Feedback to the teacher is another aspect of a formative system that some students think also has benefits for them. Feedback as information to the teacher allows the teacher to adjust the lesson to match student needs right away as well as for the next session. In this study, when using information from concept checks to adjust teaching, the instructor noticed that after this adjustment, the students understood the material better, even though they had not been able to "get" it the first time. In addition, feedback makes other activities in formative mode more effective, such as relearning and trying again.

Relearning to improve understanding is another characteristic that most students like. Students can try to master the material again, even though they were not successful the first time. They experienced that the relearning benefits themselves. Students also know that attaining good understanding is a priority to prepare for summative assessment, and that after doing that, their summative performance will be better; that is why relearning is effective in formative mode. However, some relearning procedures need to be improved to make it practical. Relearning by requiring students to redo and resubmit formative homework should not always be necessary, because the teacher usually sees if the majority of students have recognized their mistakes and understand well after the feedback in class. Redoing and reassigning creates an extra workload for both the students and the teacher. However, if the majority of students still experience difficulty with the problem after feedback, resubmitting homework is necessary.

The results of the attitude and preference survey give reasons from the students that help explain why the quantitative performance results were significantly better using the formative system. In essence, students get actionable feedback on their current understanding and are able to identify and remedy shortcomings.

In addition, another factor that makes relearning and other formative activities beneficial is learning without points and grades being involved initially. Without these, students learn without grade pressure, allowing them to focus more on content and learning activities. They are not afraid to apply their current knowledge to try solving problem or share their idea in the class, making feedback discussions flow freely. Sometimes, however, students are still nervous to share their opinions in class, so the teacher needs to encourage and remind them that a grade is not at stake. However, some students still feel pressure, as they want to do or show their best in every situation. Overall we conclude that initial learning without points and grades can make a positive impact on achievement.

Other positive factors mentioned by a few students are knowing the objectives and/or focus questions, and that formative assessment leads to more discussion. Students are able to learn more purposefully and grasp concepts better if they know the objectives, and discussion helps to develop better understanding.

Summative mode. Concerning the summative assessment mode, students had a variety of attitudes and opinions. There are several factors and techniques that they thought had positive or negative effects on their learning.

Some students stated that when they see a grade after an assignment, they feel the mission is accomplished. Some simply felt that a summative exam does not help them learn concepts. Furthermore, either low- or high-stakes summative tests have affective consequences. Some students like challenges and competing. On the other hand, if other students do not like taking tests or are not ready, they may feel anxious and pressured, contributing to a poor performance. It would be preferable for the teacher to prepare students by using appropriate formative methods and then letting them have a summative test at the end of a learning module. The summative assessment at the end also constitutes feedback to the teacher about how effective the instruction was overall.

Techniques of assessment.

Formative techniques. Some formative techniques used in the research were concept checks, peer- and self-assessment, clickers, and feedback, so the survey questionnaires specifically asked questions about each of these. Other techniques were not specifically asked about; however, students had chance in open-ended questions to state freely, for example, what their favorite aspects in formative or summative systems were. Naturally students may not provide opinions about aspects that have not been explicitly asked about, and this is reported in the findings.

Some students liked peer-assessment and thought that it helped because they were able to discuss and see other methods for problems that increased their knowledge broadly. However, the majority of students disagreed with that and thought there were several problems with peer-assessment. The peer may not have enough knowledge to assess another's work, or may not put much effort into written feedback—these were the main concerns. In fact, the teacher collected all the work and rechecked to see if the written feedback from peers was correct, and if not, the teacher wrote additional feedback to complete it. Therefore, students did get good feedback on their work, one way or another, but obviously they still worried about it. Another student concern was the procedure of peer-assessment. This was despite the fact that the teacher gave step-by-step guidance that was easy to follow, and in addition, students were able to discuss with peers and the class if they had difficulties or did not understand. Peer-assessment processes in principle should not be too much of a problem, in that they are similar to self-assessment, which the majority of students liked, and none complained that the directions were not clear. However, from this study alone, and without further probing by interviews, we do not know more about why some students thought negatively as they did about peer assessment.

There were also survey comments that during peer-assessment students were not able to evaluate their own work directly at the time. This is true and seems to be a weak point of peer-assessment, which does not seem to be mentioned in the literature. When students are engaged in peer-assessment, they are expected to understand how to solve the problem, and how to assess a peer's work, and can only use that knowledge later to assess and correct their own work. We suggest that the teacher needs to explain how peer-assessment benefits students in different ways from self-assessment. Lastly, peer-assessment procedures may sometimes be troublesome for the teacher, especially the confidentiality aspect. It can also be quite time-consuming. Accordingly, in this study the teacher did not use it often. In the future, peer-assessment could still be used, but the teacher would need to find an efficient way to handle the confidentiality aspect and the necessity for teacher checking of peer assessment and feedback.

Self-assessment is one of students' favorite formative techniques. The majority of students reported that it benefits their learning. Self-assessment was used often to discuss homework because it was convenient and there is no anonymous process compared to peer-assessment. Students also have more time to get feedback when they still are not sure of how to do a problem or question. Furthermore, with feedback, self-assessment allows students not only to correct and improve their work, but also to add a personal note explaining the way they were thinking, to remind themselves how to understand the concept or tackle such problems in future. Students are not able to do this for their own benefit in peer-assessment processes as they are grading a peer's work. For these various reasons, self-assessment is judged more effective for learning and improving performance.

Concept checks are another assessment aspect that play an important role along with feedback in formative mode. The use of concept checks is particularly valuable in improving students' understanding as an ongoing part of class instruction, built into lessons. This was true especially in the dynamics module, which was very conceptual in nature, involving physical insight into force and motion situations of various kinds. Such a topic, related to everyday experience, also brings out students' naïve experiential intuitions. In the attitude survey, most students said they liked the use of such concept checks. However, some students commented that requiring additional input like diagrams was more burdensome. This comment is probably only a small issue compared to the benefits of such enhanced concept checks. Requiring additional written work, such as force diagrams and explanations, and thus going beyond simply selecting a response, is more versatile and gets at more aspects of understanding than may be present in fixed response choices.

The survey did not specifically ask about quizzes and homework in the formative system, so only a few students reflected on these in their free responses; there were both positive and negative reflections on various aspects that they raised themselves.

Summative system methods. In survey responses relating to summative methods, a few people reflected on summative quizzes or summative homework and thought that quizzes and homework helped them to prepare for the exams. However, without descriptive comment feedback, these summative characteristics are generally not as effective for improving students' grasp and attainment as formative quizzes and homework. To get the most benefit, students need descriptive feedback, too. Furthermore, anxiety occurs when students worry that they are not able to do well in quizzes and homework, and this may affect the nature and extent of their learning. Therefore, summative assessment systems are limited in their benefits, unless students get more than usual feedback on their work.

After students experienced both formative and summative modes, they were able to say in the survey which mode(s) they preferred in the course, and why. I found that the majority of students preferred a mix of formative and summative, and only a few students preferred either formative or summative mode alone. This survey finding supports the contention that the formative assessment system, including various techniques, helps students to understand concepts and solve problems better than does the conventional summative system. Formative mode also helps students to prepare for summative tests and exams. However, summative assessments do have certain attractions for students, especially getting a final grade. This also motivates them to study harder toward a good grade. This is a good reason for having a grade system in a course, at least for exams occurring after learning. However, another set of students may not learn as well with grade pressure. The study findings suggest that a teacher needs to use the formative mode to help students understand the material and how to solve problems, before having summative tests and exams. In this way, the learning benefits of formative assessment will be attained, and the extrinsic motivational and attainment satisfaction aspects will also help students to do better in summative mode final exams.

Various kinds of assessments needed. When students indicated “I understand” via a concept check, which they answered correctly either right away or after discussion, the teacher tended to assumed that teaching and learning for that aspect had produced competency, and thus the class could move on to the next aspect. However, quite often, the teacher would later discover that students did not really understand fully and/or still had alternative conceptions, even though they thought they had understood at the time. This was initially a mystery to the teacher, but the reasons were often clarified later when longer concept checks and written tests were given to students at the end of a learning unit. When students were asked to answer the same or similar problem but in a longer form asking for detail, for example by drawing force diagrams and explaining, or else attempting a different problem with the same concept, some of them who had been correct on the short selected-response MCQs now revealed they had shortcomings in their understanding or could not solve the new problem variant.

The present research was not designed to investigate the above aspect more deeply. However, some possible factors could be at work. Students might answer a short concept check correctly immediately after instruction, but some of that may fade later, if not reinforced, by the time of the test. Students might not, in fact, have grasped the desired concept fully at the first MCQ concept check, and their initial alternative conceptions were likely still there in their thinking along with the correct idea, so when they attempted to apply their knowledge later in a new situation, they did not do so correctly. In addition, students might initially have just taken their best guess at an answer and "got lucky" in picking the correct one. Some students might have just agreed with what friends said after discussion in think-pair-share, without thinking independently or perhaps feeling embarrassed to reveal their limited knowledge. However, it is difficult to make firm conclusions about these possible explanations, as the project provided no direct deeper evidence on this.

To probe these kinds of issues further, interviews could be used in future research. However, it is wise to realize that students saying "I understand it" may not be a sufficient reason for the teacher to assume that they really do. Furthermore, it is best to have a repertoire available of short selected-response and longer structured-response assessments. The former are useful on-the-spot during teaching, and the latter are appropriate for fuller assessment of understanding further along and after teaching. These types complement each other, and using the combination is valuable to both assess and promote learning effectively. The observation in the study about different student performance on short concept checks versus on structured written assessments does draw attention to the limitations of using only selected-response items as a sufficient measure of full student understanding of a concept, even if these items have been expertly devised and validated. However, for a *comparative* study such as this one, such possible limitations may be less of an issue than if one were aiming to get an absolute measure of student understanding. The reason is that any possible limitations of the assessments will be the same for both the treatment and control conditions.

Course and Instructor Evaluation Results

All courses at Western Michigan University have a formal, anonymous student evaluation of course and instructor near the end of the semester, called the Instructor and Course Evaluation System (ICES). This standard university-wide evaluation was not envisaged as part of this research project, or as an additional source of data, since it involved the entire semester course and not just the two research instructional modules and modes. However, the evaluation results proved to be an unexpected outcome that did have bearing after all. Students' course and instructor evaluation ratings from Class A were substantially higher than those of Class B, across all categories. The comparative results are shown in Table 5.9.

Table 5.9

Attitude Toward Course and Instructor: Evaluation Results for Class A and Class B

Question Rate 1 (low) to (5) high	Class A	Class B
1. Rate the instructor's overall teaching effectiveness.	4.64	2.00
2. The instructor promoted an atmosphere conducive to work and learning.	5.00	2.40
3. Was the instructor able to explain difficult material to your satisfaction?	4.36	1.60
4. The instructor gave explanations/examples that were clearly to the point.	4.55	2.00
5. The instructor seemed well prepared for classes.	4.91	3.20
6. Rate the overall quality of this course.	4.36	2.00
7. How much have you learned in this course?	4.27	3.00
8. Do you feel course objectives were accomplished?	4.64	2.20
9. How well did examination questions reflect content and emphasis of the course?	4.73	3.00
10. Did this course increase your interest in the subject matter?	3.82	2.00
Response percentage	50%	33%

The percentage response rate of students completing university course evaluations is generally rather low. In line with this, only 11 of 22 students (50%) in class A and 5 of 15 students (33.3%) in class B completed the formal online course evaluations.

For Class A, the Likert scale ratings of almost all categories in the course and instructor evaluation are above 4 out of 5, which is an exceptionally high score. In class B, the ratings for most categories fell between 1 and 3, representing rather low scores. Therefore, students in Class A seemed to be dramatically more satisfied with the course and instructor than those in Class B. This was completely unexpected.

We note that dynamics was the last topic taught in the course before the evaluation, and that class A experienced the formative mode in this topic with good performance, while Class B at the same time experienced the summative mode in the topic, with reasonable but not as good performance. This suggests a reason for the difference in evaluations occurring immediately afterwards, as will be discussed later.

Spring 2013 course and instructor evaluation results. In the next semester, Spring 2013, only the formative assessment system was used, followed by the ICES course and instructor evaluation as before, and the evaluation results are shown in Table 5.10.

Table 5.10

Attitude Toward Course and Instructor: Evaluation Results for Class C in Spring 2013

Question Rate 1 (low) to (5) high	Class C
1. Rate the instructor's overall teaching effectiveness.	4.83
2. The instructor promoted an atmosphere conducive to work and learning.	4.67
3. Was the instructor able to explain difficult material to your satisfaction?	4.83
4. The instructor gave explanations/examples that were clearly to the point.	4.67
5. The instructor seemed well prepared for classes.	4.83
6. Rate the overall quality of this course.	4.67
7. How much have you learned in this course?	4.83
8. Do you feel course objectives were accomplished?	5.00
9. How well did examination questions reflect content and emphasis of the course?	4.83
10. Did this course increase your interest in the subject matter?	4.33
Response percentage	25%

All ratings are above 4, which are considered very high scores. Therefore, students in Class C were very satisfied with the course and instructor.

Note that the evaluation results for Class A in 2014 and Class C in 2015, both for formative mode, are remarkably similar.

Discussion of course and instructor evaluation results. Students in both Class A and Class B experienced both formative and summative modes from the beginning of the course. The teaching, learning, and materials used were similar, and the classes were taught by the same instructor, so I expected the course and instructor evaluations in classes A and B to be similar too. However, Class A returned higher course and instructor evaluation scores in the research semester than in previous regular semesters, while Class B returned lower scores than in previous semesters. A possible reason for this difference and shift from previous semesters is the following. Students in Class B, having experienced both formative and summative modes since the beginning of the course and liking the formative better, became dissatisfied when they were taught only in summative mode near the end of the course, and just before the evaluation. They did not have the expected opportunities to monitor their conceptual understanding with concept checks and descriptive feedback comments to address problems when they struggled, and receiving only the teacher's class explanations might not have helped as much. If students had not understood very well or had difficulties with problems, they might not be as satisfied with the course and the instructor, and did not do as well on this module, and thus may have rated the course and instructor low, based on their most recent experience. Class A, on the other hand, experienced formative mode and better performance near the end of the course, right before the evaluation, and the evaluation ratings reflected this.

The kinematics module, in which formative and summative modes produced comparable results, occurred three weeks earlier in the course and may not have been as fresh in students' minds when completing the evaluation at the end. Furthermore, as mentioned before, the summative mode for kinematics turned out to have more formative aspects than intended or planned, because students insisted on asking many questions on the more formula-based material and problems, until they felt they understood. Thus, students taking kinematics in this summative mode, followed by dynamics in formative mode, seemed to have a better recent experience than students ending with dynamics in strongly summative mode.

We suggest that students' overall impression of the course and instructor evaluation is strongly affected by their last experience in the course before the evaluation, as well as by their level of achievement during that experience.

Conclusions

For formative assessment to be most effective and meaningful, courses and lessons must be designed from the start to embed various types and techniques of formative assessment, as in this study, rather than, for example, simply adding clicker questions here and there. To use formative assessment effectively, the teacher needs to make it work as an entire coherent system, to have the maximum positive impact on teaching and learning.

In this study, students experienced both formative and summative modes during the course before collecting data on two instructional modules, dynamics and kinematics. The crossover research design allowed all students to experience each mode, in either one topic or the other. This study, treating formative assessment from an integrated system viewpoint, found that such a formative system led to significantly greater physics performance gains than the summative system, and more favorable student attitudes. Results for the formative mode were better than for the summative, for both the dynamics and kinematics modules; however, a finding not anticipated was that while formative outperformed summative substantially for dynamics, the same was not true for kinematics, where performance was roughly comparable in the two modes. We surmise that the contrasting nature of the topics contributed to this, since the dynamics module was mostly conceptual without mathematics, while kinematics was more formula-based to a fair extent. If this difference in character of the two modules is indeed a factor, then it seems that the advantage of formative mode is greater for a conceptual topic than a more mathematical one.

Survey results were that students had more favorable attitudes toward formative assessment systems than summative, for supporting their learning, but also felt the need for summative final results, so that most students preferred a combination of formative assessment during teaching and learning and summative assessment in exams at the end of the course.

Feedback was one of main aspects mentioned of formative assessment benefiting learners. The formative system also provided feedback to the instructor, allowing adjustment of instruction in real time during class or for the next session.

For formal course and instructor evaluations, students who experienced the formative mode last in the course gave much more favorable ratings of the course and instructor than those who experienced the summative mode last.

Overall, we conclude that a formative assessment system coherently integrated into topic instructional design leads to substantially higher learning gains in physics than a conventional summative system, and that the extent of the benefit is greater for a conceptual than a more mathematical topic. Surveyed students' attitudes are much more positive toward a formative system, though students still want a summative final exam. The survey suggests that reasons for better performance and attitudes in formative mode lay in various identified attributes of formative assessment.

REFERENCES

Alexopoulou, E., & Driver, R. (1996). Small-group discussion in physics: Peer interaction modes in pairs and fours. *Journal of Research in Science Teaching*, *33*(10), 1099–1114.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, *84*(3), 261.

Andrade, H., & Cizek, G. (2010). *Handbook of formative assessment*. New York, NY: Routledge.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, *37*(2), 122–147.

Bangert-Drowns, R. L., Kulik, C. L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, *61*(2), 213–238.

Beatty, I., Gerace, W., Feldman, A., & Leonard, W. (2008). Technology-enhanced formative assessment: An innovative approach to student-centered science teaching. Retrieved from http://www.srri.umass.edu/sites/srri/files/beatty-2008tia.pdf

Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education*, *85*(5), 536–553.

Biggs, J. (1998). Assessment and classroom learning: A role for summative assessment? *Assessment in Education*, *5*(1), 103–110.

Biggs, J. B., & Tang, C. S. (2007). *Teaching for quality learning at university* (3rd ed.). Society for Research into Higher Education Series. New York: Open University Press.

Black, H. (1986). Assessment for learning. In D. L. Nuttall (Ed.), *Assessing educational achievement*, 7-18. London: Falmer.

Black, P., & Harrison, C. (2001a). Feedback in questioning and marking: The science teacher's role in formative assessment. *School Science Review*, *82*(301), 55–61.

Black, P., & Harrison, C. (2001b). Self- and peer-assessment and taking responsibility: The science student's role in formative assessment. *School Science Review*, *83*(302), 43–49.

Black, P., Harrison, C., Lee, D., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, *86*(1), 8–21.

Black, P., Harrison, C., Lee, C., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. New York: McGraw-Hill International.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *80*(2), 139–148. Retrieved from Phi Delta Kappa International online article: https://webmail.wmich.edu/service/home/~/1998_Black_Inside%20the%20Black%20Box_Black_Wiliam.pdf?auth=co&loc=en_US&id=64860&part=2.2

Black, P., & Wiliam, D. (2004). The formative purpose: Assessment must first promote learning. *Yearbook of the National Society for the Study of Education*, *103*(2), 20–50.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31.

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *92*(1), 81–90.

Bloom, B., Hastings, J., & Madaus, G. (1971). *Handbook on formative and summative evaluation of student learning*. New York, NY: McGraw-Hill.

Boud, D., & Holmes, H. (1995). Self and peer marking in a large technical subject. In D. Boud (Ed.), *Enhancing learning through self assessment*. London: Kogan Page.

Bransford, J., Brown, A., & Cocking, R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, D.C.: National Academy Press, National Research Council.

Burke, K. (2010). *Balanced assessment: From formative to summative*. Bloomington, IN: Solution Tree Press.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, *58*, 1–14.

Bybee, R. W. (1997). *Achieving scientific literacy: From purposes to practices*. Westport, CT: Heinemann.

Cizek, G., & Andrade, H. (2010). An introduction to formative assessment: History, characteristics, and challenges. In *Handbook of formative assessment* (pp. 3–17). New York, NY: Routledge.

Cobern, W., & Aikenhead, G. (1997). Cultural aspects of learning science. *Scientific Literacy and Cultural Studies Project.* Retrieved from http://scholarworks.wmich.edu/science_slcsp/13

Cooper, B., & Cowie, B. (2010). Collaborative research for assessment for learning. *Teaching and Teacher Education*, *26*(4), 979–986.

Cooper, J. L., & Robinson, P. (2000). Getting started: Informal small-group strategies in large classes. *New Directions for Teaching & Learning*, *2000*(81), 17–24.

Corno, L., & Mandinach, E. B. (1983). The role of cognitive engagement in classroom learning and motivation. *Educational Psychologist*, *18*(2), 88–108.

Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, *58*(4), 438–481.

Crouch, C., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American Journal of Physics*, *69*(9), 970–977.

Darling-Hammond, L. (2010). *Performance counts: Assessment aystems that support high-quality learning.* Washington, DC: Council of Chief State School Officers. Retrieved from http://eric.ed.gov/?id=ED543057

Donovan, M. S., & Bransford, J. D. (1999). How people learn: Bridging research and practice. *National Academy of Sciences*, 88.

Donovan, M. S., & Bransford, J. D. (2005). *How students learn: History, mathematics, and science in the classroom*. *Social cciences* (Vol. 5, pp. 1–50). Washington, DC: National Academies Press.

Elawar, M. C., & Corno, L. (1985). A factorial experiment in teachers' written feedback on student homework: Changing teacher behavior a little rather than a lot. *Journal of Educational Psychology*, *77*(2), 162–173.

Falchikov, N., & Blythman, M. (2001). *Learning together: Peer tutoring in higher education* (p. 312). Psychology Press.

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59*(4), 395–430.

Fernandes, M., & Fontana, D. (1996). Changes in control beliefs in Portuguese primary school pupils as a consequence of the employment of self-assessment strategies. *The British Journal of Educational Psychology, 66*(Pt. 3), 301–313.

Fisher, D., & Frey, N. (2007). *Checking for understanding: Formative assessment techniques for your classroom*. Alexandria, VA: ASCD.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8410–8415.

Gibbs, G., & Simpson, C. (2004). Does your assessment support your students' learning? *Journal of Teaching and Learning in Higher Education*, *1*(1), 1–30.

Gioka, O. (2006). Assessment for learning in physics investigations: Assessment criteria, questions and feedback in marking. *Physics Education*, *342*(41), 341–346.

Harlen, W. (2005a). Teachers' summative practices and assessment for learning: Tensions and synergies. *Curriculum Journal*, *16*(2), 207–223.

Harlen, W. (2005b). Trusting teachers' judgment: Research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research Papers in Education*, *20*(3), 245–270.

Harlen, W., & James, M. (1996). Creating a positive impact of assessment on learning. *American Educational Research Association*, 8–12.

Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Library*, *4*(3), 365–379.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112.

Hill, K. T. (1984). Debilitating motivation and testing: A major educational problem, possible solutions, and policy applications. *Research on Motivation in Education: Student Motivation*, *1*, 245–274.

Hill, K. T., & Wigfield, A. (1984). Test anxiety: A major educational problem and what can be done about it. *The Elementary School Journal*, 105–126.

Jacob, E. (1999). *Cooperative learning in context: An educational innovation in everyday classrooms*. Albany, NY: SUNY Press.

Karplus, R. (1975). The learning cycle. In F. Collea et al. (Eds.), *Workshop on physics teaching and the development of reasoning.* Stonybrook, NY: American Association of Physics.

Klenowski, V. (1995). Student self-evaluation processes in student-centred teaching and learning contexts of Australia and England. *Assessment in Education: Principles, Policy & Practice*, *2*(2), 145–163.

Kluger, A. N., & Denisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284.

Kowalski, S. E., Kowalski, F. V., & Gardner, T. Q. (2009, October). Lessons learned when gathering real-time formative assessment in the university classroom using tablet PCs. In *Frontiers in Education Conference, 2009* (pp. 1–5). IEEE.

Lasry, N., Mazur, E., & Watkins, J. (2008). Peer instruction: From Harvard to the two-year college. *American Journal of Physics, 76*(11), 1066–1069.

Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment minute by minute day by day. *Educational Leadership*, *63*(3), 18–24.

Liu, N.-F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education*, *11*(3), 279–290.

Looney, J. W. (2011). Integrating formative and summative assessment: Progress toward a seamless system? *OECD Education Working Papers, No. 58.* OECD Publishing.

Lyman, F. (1987). Think-pair-share: An expanding teaching technique. *MAA-CIE Cooperative News*, *1*(1), 1–2.

Lyman, F. T. (1981). The responsive classroom discussion: The inclusion of all students. *Mainstreaming Digest*, 109–113.

Mazur, E. (1997). *Peer instruction: A user's manual*. Upper Saddle River, NJ: Prentice Hall.

Mazur, E., & Watkins, J. (2009). Just-in-time teaching and peer instruction. In S. Simkins & M. Maier (Eds.), *Just in time teaching: Across the disciplines, and across the academy* (pp. 39–62). Sterling, VA: Stylus.

McKeachie, W., & Svinicki, M. (2013). *McKeachie's teaching tips*. Belmont, CA: Wadsworth, Cengage Learning.

McKeachie, W. J. (1986). *Teaching and learning in the college classroom: A review of the research literature* (Vol. 86). Ann Arbor: University of Michigan Press.

McTighe, J., & Lyman, F. (1988). Cueing thinking in the classroom: The promise of theory-embedded tools. *Educational Leadership*, *45*(7), 18–24.

Mintzes, J. J., & Leonard, W. H. (2006). *Handbook of college science teaching* (p. 416). NSTA Press.

Moriber, G. (1971). Wait-time in college science classes. *Science Education*, *55*(3), 321–328.

Murthy, S. (2007). Peer-assessment of homework using rubrics, *American Institute of Physics Conference Proceedings, 951*(1), 156–159.

National Council for Curriculum and Assessment. (2014). AfL overview. Retrieved from http://www.ncca.ie/en/Curriculum_and_Assessment/Post-Primary_Education/Junior_Cycle/Assessment_for_Learning_AfL_/Overview/AfL_-_Overview.html

Nyquist, J. B. (2003). *The benefits of reconstruing feedback as a larger system of formative assessment: A meta-analysis* (Doctoral dissertation). Vanderbilt University, Nashville, TN.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, *25*(9), 1049–1079.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Rolheiser, C., & Ross, J. A. (2000). Student self-evaluation: What research says and what practice shows. *Orbit*, *30*, 33–36.

Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education*, *5*(1), 77–84.

Sargent, C. S., & Curcio, A. A. (2012). Empirical evidence that formative assessments improve final exams. *Journal of Legal Education*, *61*(3), 379–405.

Schunk, D. H. (1984). Self-efficacy perspective on achievement behavior. *Educational Psychologist*, *19*(1), 48–58.

Schunk, D. H. (1995). Learning goals and self-evaluation: Effects on children's cognitive skill acquisition. In *Annual Meeting and Exhibit of the American Educational Research* (p. 14). San Francisco, CA. Retrieved from http://eric.ed.gov/?id=ED389385

Schuster, D., Cobern, W., Applegate, B., Schwartz, R., Vellom, P., & Undreiu, A. (2006). Assessing pedagogical content knowledge of inquiry science teaching: Developing an assessment instrument to support the undergraduate preparation of elementary teachers to teach science as inquiry. *STEM Assessment Conference*, 247.

Sebatane, E. M. (1998). Assessment and classroom learning: A response to Black and Wiliam. *Assessment in Education*, *5*(1), 123.

Shrigley, R. L., Koballa, T. R., & Simpson, R. D. (1988). Defining attitude for science educators. *Journal of Research in Science Teaching*, *25*(8), 659–678.

Stiggins, R., & Chappuis, J. (2006). What a difference a word makes. *Journal of Staff Development*, *27*(1).

Stiggins, R. J. (1991). *A practical guide for developing sound grading practices.* Available from the Northwest Regional Educational Laboratory, 101.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, *83*(10), 758–765.

Strike, K., & Posner, G. (1985). A conceptual change view of learning and understanding. *Cognitive Structure and Conceptual Change*, *211*, 231.

Thorndike, E., & Woodyard, E. (1934). The influence of the relative frequency of successes and frustrations upon intellectual achievement. *Journal of Educational Psychology*, *25*(4).

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *68*(3), 249–276.

Tyminski, A. M., Richardson, S. E., & Winarski, E. (2010). Enhancing think-pair-share. *Teaching Children Mathematics*, *16*(8), 451.

Weaver, R. L., & Cotrell, H. W. (1986). Peer evaluation: A case study. *Innovative Higher Education*, *11*(1), 25–39.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, *16*(1), 3–118.

Wiggins, G. P. (1998). *Educative assessment: Designing assessments to inform and improve student performance* (Vol. 1). San Francisco, CA: Jossey-Bass.

Wiliam, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment*, *11*(3), 283–289.

Wiliam, D. (2011). *Embedded formative assessment*. Solution Tree Press.

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy & Practice*, *11*(1), 49–65.

Young, E. (2005). *Assessment for learning: Embedding and extending.* Retrieved from http://www.educationscotland.gov.uk/resources/a/publication_tcm4509456.asp?strReferringChannel=educationscotland&strReferringPageID=tcm:4-615801-64

Appendix A

Human Subjects Institutional Review Board
Letter of Approval

WESTERN MICHIGAN UNIVERSITY

Human Subjects Institutional Review Board

Date: November 29, 2012

To: David Schuster, Principal Investigator
Chaiphat Plybour, Student Investigator for dissertation
Rex Taibu, Student Investigator

From: Amy Naugle, Ph.D., Chair

Re: HSIRB Project Number 12-11-37

This letter will serve as confirmation that your research project titled "Integrating Formative Assessment into Physics Instructional Design, and the Effect of Formative vs. Summative Assessment on Student Learning in Two Physics Units." has been **approved** under the **exempt** category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note: This research may **only** be conducted exactly in the form it was approved. You must seek specific board approval for any changes in this project (e.g., ***you must request a post approval change to enroll subjects beyond the number stated in your application under "Number of subjects you want to complete the study***)." Failure to obtain approval for changes will result in a protocol deviation. In addition, if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

Reapproval of the project is required if it extends beyond the termination date stated below.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: November 29, 2013

Walwood Hall, Kalamazoo, MI 49008-5456
PHONE: (269) 387-8293 FAX: (269) 387-8276

Appendix B

Consent Form

Western Michigan University
Mallinson Institute for Science Education

CONSENT FORM FOR USE OF STUDENTS' ACHIEVEMENT SCORES AND GRADES AND ATTITUDE SURVEY

Principal Investigator: Dr. David Schuster
Student Investigator: Chaiphat Plybour

Integrating Formative Assessment into Physics Instructional Design, and the Effect of Formative vs. Summative Assessment on Student Learning in Two Physics Units

You have been invited to participate in a research project entitled "Integrating Formative Assessment into Physics Instructional Design, and the Effect of Formative vs. Summative Assessment on Student Learning in Two Physics Units". The information gained from this project will be used in Chaiphat Plybour's PhD dissertation on science education. The consent document will detail the purpose of the research project, the time commitment, the procedures used in the study, and the risks and benefits of participating in the research. Please read this consent form carefully and completely and please ask any questions if you need more clarification.

What are we trying to find out?
The researcher attempts to seek 1) which method between formative and summative assessment is better to improve students' achievement 2) whether or not students who have no prior physics education have the same learning gain as students who do 3) which method can improve students' attitude toward learning achievement.

Who can participate in this study?
To participate in this study you need to (1) be enrolled in PHYS 1800 courses and (2) sign this consent form. The consent form has two parts: scores & grades and attitude survey. To give us consent to use your scores and grades information, you may sign the scores and grades part. Signing both parts means you allow us to use both your scores and grades, and attitude survey. Signing the attitude survey will only mean that we may select you for attitude survey, but will not use your scores and grades result for the research.

Where will this study take place?
This study will take place in room 1413, Wood Hall, Western Michigan University.

What is the time commitment for participating in this study?
Paticipants in each section will take approximately 2 hours and 30 minutes for pre- and post-test and survey that is part of the standard classroom activity.

What will you be asked to do if you choose to participate in this study?
You will be asked to permit the researchers to use pre-test and post- test scores and grades on kinematics and dynamics that are part of standard classroom activity. Those who participate in the attitude survey will answer questions about the physics course.

What information is being measured during the study?
The pre-and post-test scores will be used to calculate the normalized gain between the sections taught using formative assessment and summative assessment. The score and grade may help us determine which instruction is more effective in teaching kinematics and dynamics. The attitude survey will be used to determine how each instruction influences participants' learning.

What are the risks of participating in this study and how will these risks be minimized?
There are no known risks for your participation in the study.

What are the benefits of participating in this study?
There are no benefits for participating in the study. The kinematics and dynamics data as a part of activities in the class may help the investigators to design effective instructions in both topic areas in the future.

Are there any costs associated with participating in this study?
There are no costs to be incurred as a result of participating in the study other than the time it takes to complete the tests and participate in the survey.

Is there any compensation for participating in this study?
There are no compensations for participating.

Who will have access to the information collected during this study?
Only the Principal Investigator and the Student Investigator will have access to the study data and information. There will be no identifying information when analyzing and reporting data from this study. The name of participants participating on attitude survey will be replaced by pseudonyms. The results of the research may be published in the form of a research paper, a doctoral dissertation of the students' investigator or presented at professional meetings. The Principal Investigator and the Student Investigator will not have access until final grades for the semester are submitted.

What if you want to stop participating in this study?
You can choose to stop participating in the study at anytime for any reason. You will not suffer any prejudice or penalty by your decision to stop your participation. You will experience NO consequences either academically or personally if you choose to withdraw from this study.

The investigator can also decide to stop your participation in the study without your consent. Should you have any questions prior to or during the study, you can contact the Principal Investigator Dr. Schuster at **(269) 387-5844** or david.schuster@wmich.edu. You may also contact the Chair, Human Subjects Institutional Review Board at (269) 387-8293 or the Vice President for Research at (269) 387-8298 if questions arise during the course of the study.

This consent document has been approved for use for one year by the Human Subjects Institutional Review Board (HSIRB) as indicated by the stamped date and signature of the board chair in the upper right corner. Do not participate in this study if the stamped date is older than one year.

I have read this informed consent document. The risks and benefits have been explained to me. I agree to take part in this study.

____

Please Print Your Name

________________________ ____________________

Participant's signature Date

Appendix C

Pre- and Posttest Dynamics

Name:____ Group______

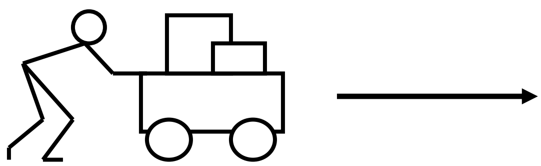
Dynamics Fall 2012 Date: Dec 10, 2012

Number of physics courses taken

In High school__________ In College___________

Please read each item carefully. Then circle the correct answer to the question and Transfer your answers A, B, C, or D to the Scantron sheet. Turn in Both the question paper and the scantron sheet.

1. Amy pushes a wagon containing some packages along a level floor. The wagon has such good wheels that friction is very small.

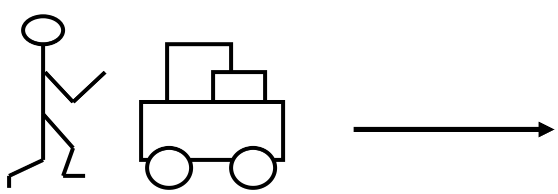


If Amy pushes continuously all the time with a **constant pushing force**, and friction can be ignored, then the wagon will . . .

A. not move.
B. move with a constant speed.
C. speed up steadily.
D. slow down.

1. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

2. Amy suddenly **stops pushing** when the wagon is moving quite fast**.**

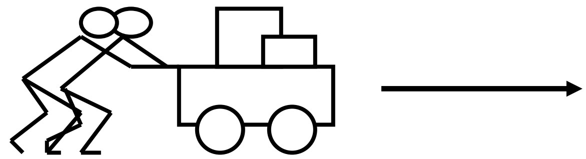


What will happen next? (Remember that friction can be ignored). The wagon will . . .

A. stop immediately.
B. continue moving at the speed it was going.
C. speed up.
D. slow down quickly, then stop.

2. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

3. Amy's sister joins her and **both push together** on the loaded wagon, **each** pushing just at hard as Amy did before.

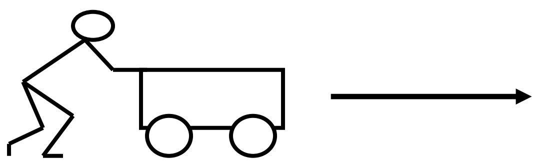


Compared to when Amy was pushing alone, the wagon will now . . .

A. travel at the same steady speed.
B. travel at a faster steady speed.
C. speed up just as before.
D. speed up quicker than before.

3. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

4. Amy now **unloads** the packages from the wagon and then pushes it with the **same constant force** as she used before.

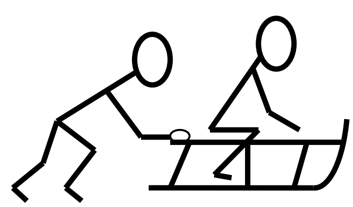


The empty wagon will now . . .

1. move just the same as when it was loaded.
2. move at a steady speed, but faster than before.
3. move at a steady speed, but slower than before.
4. speed up, more rapidly than before.

4. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

5. You are pushing a sled through the snow by applying a horizontal force with a magnitude of 70 newtons. There is a force of friction between the sled and the snow with a magnitude of 20 newtons.



What is the resulting (net) force on the sled?

A. 20 newtons
B. 50 newtons
C. 70 newtons
D. 90 newtons

5. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

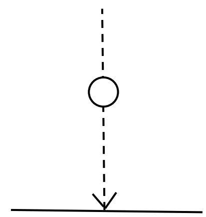
6. We see a car moving straight past our window at a steady speed. What can we say about the possible forces acting on the object?

A. There are no forces at all acting on the object.
B. There may be forces on the object, but if so they are equal and opposite (balanced).
C. There is a net (unbalanced) force on the object in the direction it is moving.
D. There is only *one* force on the object, in the direction it is moving.

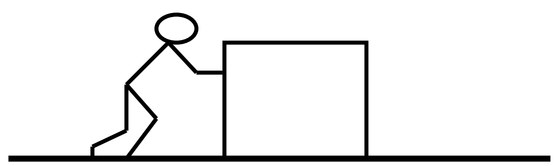
6. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

7. A stone falls to the ground. Any air resistance can be ignored. As the stone falls, what is true of its **speed** and the **force**, if any, acting on it?

A. Its speed is constant <u>and</u> there is a constant force on it.
B. Its speed is constant <u>and</u> the force on it is zero.
C. Its speed increases <u>and</u> the force on it is constant.
D. Its speed increases <u>and</u> the force on it is increasing.



7. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess



8. Tom is pushing on a crate with a constant force of **90 newtons,** across a level floor. The floor creates a **friction force** on the box of **40 newtons**.
What is the size of the net (unbalanced) force on the box causing it to speed up?

A. 50 newtons
B. 65 newtons
C. 90 newtons
D. 130 newtons

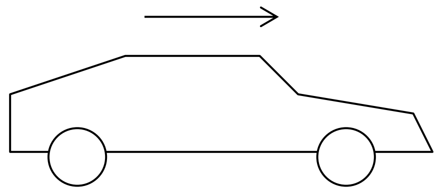
8. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

9. After the box in question 8a has gained some speed, Tom reduces his applied force to 40 newtons (which is the same as the friction force). What will be the motion of the box when he is pushing with 40 newtons?

A. It will stop immediately.
B. It will still speed up, but not as quickly as before.
C. It will move at a steady speed.
D. It will slow down.

9. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

10. A car is traveling at a **steady** 70 mph along a straight level highway, not speeding up or slowing down.



What is true of the forward and backward force(s) acting on the car in this situation?

A. There are no forward or backward forces acting on the car.
B. There is a forward force and a backward force, which are exactly equal and opposite (balanced).
C. There is a forward force and a backward force, but the forward force is bigger.
D. There is only a forward force on the car.

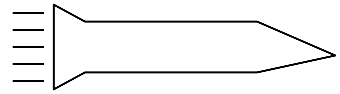
10. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

11. A passenger is in a car traveling at a steady speed along a straight level road. What is true of the forward and backward force(s), if any, on the seated passenger? (Note: this asks about forces on the passenger, not on the car).

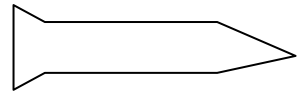
A. There are no forward or backward forces on the passenger.
B. There is a forward force on the passenger, but no backward force.
C. There are forward and backward forces on the passenger, but the forward force is bigger.
D. There are forward and backward forces on the passenger, exactly equal and opposite.

11. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

12. A rocket is out in space where there is no friction at all to affect its motion. You can also ignore any effect of gravity. At first the rocket engine is operating continuously as shown in the top diagram, and the rocket is speeding up.



Then the engine suddenly shuts off.

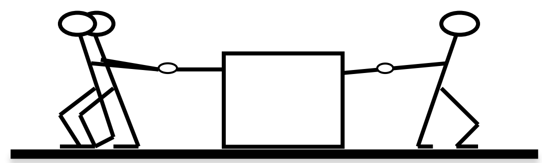


After the engine shuts off, what is true of the forces acting on the rocket and how it moves?

A. The force remains as before, and the rocket keeps speeding up.
B. The force remains as before, and the rocket travels at a constant speed.
C. The force becomes zero, and the rocket travels at constant speed.
E. The force becomes zero, and the rocket stops.

12. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

13. A heavy box full of toys weighing 100 newtons is on rough ground. Andy starts pulling on the box to the *right* with a force of 60 newtons, but then Betty and Cathy start pulling to the *left* with forces of 50 newtons and 30 newtons. To everyone's surprise, when all three are pulling as shown, the box does not move one way or the other.



We can conclude that the size of the **friction** force on the box is . . .

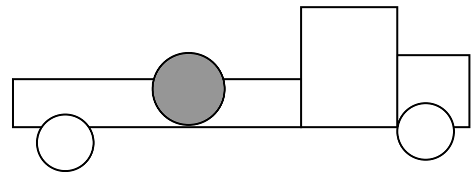
A. 20 newtons.
B. 40 newtons.
C. 100 newtons.
F. 140 newtons.

13. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

14. The *direction* of the friction force in question 13a above is . . .

A. down.
B. up.
C. to the left.
D. to the right.

14. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess



15. A round boulder is placed loose in the back of a pickup truck. The driver starts the truck and begins speeding up. As the truck starts moving forward, the driver hears a noise, and looking back notices that the boulder has hit the back (tailgate) of the pickup. The driver wonders about the reason for this.

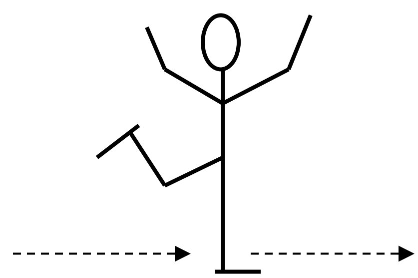
Sally and Sam give different explanations, as follows.

1. Sally says that the boulder tended to 'stay put' as the truck started moving, so it did not move forward with the truck. It only appeared to move backward in the truck because the truck moved forward.
2. Sam says that since the boulder hit the back of the truck, there must have been a backward force acting on it, causing it to do so.

What is the best comment about these two statements?

A. Only Sally's statement (#1) is correct.
B. Only Sam's statement (#2) is correct.
C. Both Sally and Sam have made correct statements.
D. Neither Sally's nor Sam's statements are correct.

15. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

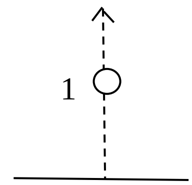


16. An ice skater is skating on an indoor ice rink. She first gets up to a fast speed, then stands on one skate and keeps going steadily across the ice without any apparent effort.
The reason for her keeping moving like this without effort is that . . .

A. the ice must be sloping slightly downwards.
B. the air is pushing her forward.
C. this is the natural behavior of objects with no net force on them.
D. the force that got her moving is still acting on her.

16. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

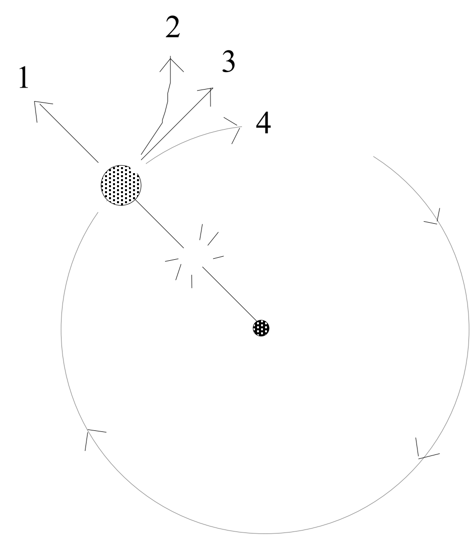
17. A ball is thrown upward as shown, and as it moves upwards its speed gets less. Air resistance is small and can be ignored. When the ball is on the way up as shown (at position 1), what forces, if any, are acting on it?



A. No forces.
B. Only an upward force.
C. Only a downward force.
D. An upward and a downward force, the upward force being larger.

17. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

18. A stone is being swung around in a circle at the end of a string. The diagram shows a top view of the path of the stone. The string suddenly breaks when the moving stone is in the position shown.

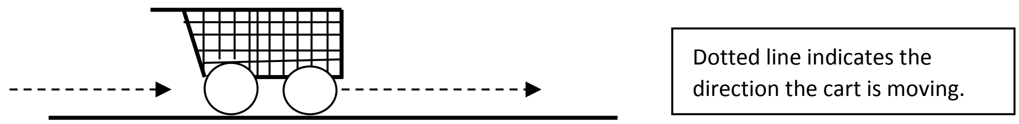


Which numbered path best shows the motion of the stone *after* the string breaks?

A. Path 1
B. Path 2
C. Path 3
D. Path 4

18. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

19. A shopping cart is given an initial push to get it moving on a carpeted floor. The cart then travels along the carpet on its own for a while, gradually slowing down.

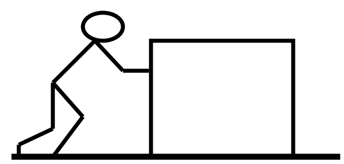


What horizontal force(s), if any, are acting on the shopping cart while it is still moving forward but slowing down?

A. Only a forward force, which diminishes with time
B. Only a backward force
C. A forward force and a backward force, but the forward force is bigger
D. A forward force and a backward force, but the backward force is bigger

19. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

20. A box is at rest on the floor. You apply a horizontal force to it by pushing from the side, but it does not move. What can be said about the friction force, if any, acting on the box while you are pushing?



A. There is no friction force.
B. The friction force is less than the force you apply.
C. The friction force is equal to the force you apply.
D. The friction force is greater than the force you apply.

20. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

21. There are plates, cups and other dishes placed on a tablecloth lying on a dining table. You hold the edge of the tablecloth with both hands and pull it toward you with a quick jerk. The tablecloth comes out and the dishes stay pretty much where they were! The reason the dishes don't get pulled along with the tablecloth and fall off the table is that while the cloth is being pulled . . .

A. There is no friction between tablecloth and dishes, so the dishes will experience no horizontal force and remain where they are.
B. A force due to friction with the tablecloth <u>does</u> act on the dishes, but it is insignificant and cannot affect the dishes.
C. A force due to friction with the tablecloth does act on the dishes, but acts for a very short time if the cloth is pulled out quickly, so the dishes have little time to move.
D. No force is being applied directly by your hands to the dishes, so there is no reason for the dishes to move.

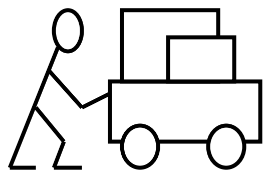
21. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

22. Our hero notices that a wheeled cart is rolling by itself toward a baby on the level sidewalk. She rushes in and pushes backward on the cart as hard as she can. What will happen?

A. She will have no effect if the cart is a lot bigger than she is.
B. The cart will stop immediately, since there is nothing pushing it forward.
C. The cart keeps going forward even though she is pushing backward, but it will slow down.
D. The cart will instantly go backward instead of forward.

22. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

23. Students are investigating the motion of a loaded cart and find that it speeds up when they push it.



They suspect that how quickly the cart speeds up might depend on **two** things, namely: the **size of the force** they apply to the cart, and the **mass** of the cart. They decide to investigate this, but two groups use different methods, as follows.

First Method (Group 1)	**Second Method (Group 2)**
Group 1 first keeps the mass of the cart constant, and applies three different forces, to find the effect of force size. Next they keep the force constant and apply it to the cart loaded with three different masses, to find the effect of mass.	Group 2 varies both the force and the mass, from one trial to the next, to get six different combinations of force and mass. Then they look at their data for all six trials, seeking possible patterns that show the effect of both force size and of mass.

Which of these two methods is better for finding out how speeding up depends on force size and on mass?

A. The first method is better.
B. The second method is better.
C. Both methods are equally good.
D. Neither method is good.

23. How sure are you of your answer? A. Very sure B. Somewhat sure C. My best guess

Appendix D

Pre- and Posttest Kinematics

(Pretest 1) Physics 1800 Fall 2012. Name: . . . . Section time: . . . .

First basic problem on the speed concept – parallel tabular form (variant)

Problem. *An athlete runs 18 meters in 3 seconds at a steady speed. What is her speed?*

Rough sketch of situation, labeled with the given quantities

Solution by word-logic and mental arithmetic

Question re-casting

The question is “What is her speed?” Reformulate this without using the words ‘speed’ or ‘velocity’, to bring out the underlying conceptual meaning of this question.

Word-logic statements and mental arithmetic

Well, she runs _18_ meters in _3_ seconds (given),

so she will run ___ meters in 1 second.

So her speed is . . . .

Answer:

Thus her speed is . . . . meters per second.

Solution by symbolic method (algebra then substitution)

Distance traveled □s = 3 m
Time taken □t = 3 s

Answer:

Thus her speed is . . . . meters per second.

Any note or comment?

(Pretest 1) Physics 1800 Fall 2012. Name: . . . . Section time: . . . .

Third basic problem involving the speed concept – parallel tabular form (variant)

<table>
<tr><td colspan="2">Problem. A person is running at a steady speed of 4 meters per second. (4 m/s). How long does it take her to travel 20 meters?</td></tr>
<tr><td colspan="2">Rough sketch of situation, labeled with the given quantities</td></tr>
<tr><td>Solution by word-logic and mental arithmetic

Re-cast the information given (without using the word speed).

Start with the given information “at a steady speed of 4 meters per second”. Reword this, without using the word ‘speed’, to bring out the basic conceptual meaning of the speed concept.

Word-logic statements and mental arithmetic</td><td>Solution by symbolic method (algebra then substitution)

Given: v = 4 m/s</td></tr>
<tr><td>Answer:</td><td>Answer:</td></tr>
<tr><td colspan="2">Any note or comment for us?</td></tr>
</table>

(Pretest 1) Physics 1800 Fall 2012. Name: . . . . Section time: . . . .

First basic problem on the acceleration concept – parallel tabular form (variant)

Problem. A car speeds up from a speed of 5 meters per second to 13 meters per second, taking 4 seconds to do so. Find the acceleration.

Rough sketch of situation, labeled with the given quantities

<table>
<tr>
<td>

Solution by word-logic and mental arithmetic

Re-cast the question

The question is "find the acceleration". Re-word this without using the word "acceleration" to bring out the underlying conceptual meaning of the question.

Word-logic statements and mental arithmetic

The speed changes by . . . . m/s in . . . . seconds,

Therefore it changes by . . . . .m/s in 1 second.

Therefore

</td>
<td>

Solution by symbolic method (algebra then substitution)

Initial speed v_1 = 5 m/s. Final speed v_2 = 13 m/s

</td>
</tr>
<tr>
<td>

Answer
The car's acceleration is. . . . . . meters per second every second.

(or; the speed changes by . . . meters per second, every second).

</td>
<td>

Answer
The car's acceleration is . . . . . .m/s/s (meters per second per second)

</td>
</tr>
</table>

Any note or comment for us?

(Pretest 1) Physics 1800 Fall 2012. Name: . . . . Section time: . . . .

Third basic problem on the acceleration concept – parallel tabular form (variant)

<table>
<tr><td colspan="2">Problem. A car accelerates uniformly at 2 m/s/s in order to overtake another car. If the initial speed was 20 m/s and the car accelerated until its speed was 26 m/s, for how long did it accelerate?</td></tr>
<tr><td colspan="2">Rough sketch of situation, labeled with the given quantities</td></tr>
<tr><td>Solution by word-logic and mental arithmetic

Re-cast the question
The given information is “a car accelerates at 2 m/s/s”. Re-word this without using the terminology “accelerates” to bring out the underlying conceptual meaning of the information, in terms of speed changes.

Word-logic statements and mental arithmetic</td><td>Solution by symbolic method (algebra then substitution)</td></tr>
<tr><td>Answer
The time taken accelerating was. . . . . . seconds.</td><td>Answer
The time taken accelerating was. . . . . . seconds.</td></tr>
<tr><td colspan="2">Any note or comment for us?</td></tr>
</table>

(Posttest 1) Physics 1800 Fall 2012. Name: . . . . Section time: . . . .

Problem on the speed concept

<table>
<tr><td colspan="2">Problem. A cyclist travels a distance of 20 meters in 4 seconds at a steady speed. What is her speed? (7)</td></tr>
<tr><td colspan="2">Rough sketch of situation, labeled with the given quantities (1)</td></tr>
<tr><td>Solution by word-logic and mental arithmetic

Question re-casting
The question is “What is her speed?” Reformulate this without using the words ‘speed’ or ‘velocity’, to bring out the underlying conceptual meaning of this question.

Word-logic statements and mental arithmetic (2)
Well, she travels <u>20</u> meters in <u>4</u> seconds (given),
so she will travel ___ meters in 1 second.
So her speed is . . . .</td><td>Solution by symbolic method or algebra (show the relevant formula then substitution) (2)

Distance traveled $\Delta s = 20$ m
Time taken $\Delta t = 4$ s</td></tr>
<tr><td>Answer: (1)
Thus her speed is . . . . meters per second.</td><td>Answer: (1)
Thus her speed is . . . . meters per second.</td></tr>
<tr><td colspan="2">Any note or comment?</td></tr>
</table>

(Postest 1) Physics 1800 Fall 2012. Name: ....Group:.................................

A problem involving the speed concept

<table>
<tr><td colspan="2">Problem. A person is running at a steady speed of 5 meters per second. (5 m/s). How long does it take her to travel 20 meters? (7)</td></tr>
<tr><td colspan="2">Rough sketch of situation, labeled with the given quantities (1)</td></tr>
<tr><td>Solution by word-logic and mental arithmetic

Re-cast the information given (without using the word speed).
Start with the given information “at a steady speed of 5 meters per second”. Reword this, without using the word ‘speed’, to bring out the basic conceptual meaning of the speed concept.

Word-logic statements and mental arithmetic (2)</td><td>Solution by symbolic method or algebra (show the relevant formula then substitution)

Given: v = 5 m/s (2)</td></tr>
<tr><td>Answer: (1)</td><td>Answer: (1)</td></tr>
<tr><td colspan="2">Any note or comment for us?</td></tr>
</table>

(Postest 1) Physics 1800 Fall 2012. Name: ....Group:..........................

A problem on the acceleration concept

<table>
<tr><td colspan="2">Problem. A train speeds up from a speed of 5 meters per second to 15 meters per second, taking 5 seconds to do so. Find the acceleration of the train. (7)</td></tr>
<tr><td colspan="2">Rough sketch of situation, labeled with the given quantities (1.5)</td></tr>
<tr><td>Solution by word-logic and mental arithmetic

Re-cast the question
The question is “find the acceleration”. Re-word this without using the word “acceleration” to bring out the underlying conceptual meaning of the question.

Word-logic statements and mental arithmetic (1.5)
The speed changes by . . . . m/s in . . . . seconds,
Therefore it changes by . . . . .m/s in 1 second.
Therefore</td><td>Solution by symbolic method or algebra (show the relevant formula then substitution) (2)

Initial speed $v_1 = 5$ m/s. Final speed $v_2 = 15$ m/s</td></tr>
<tr><td>Answer
The train’s acceleration is. . . . . . meters per second every second. (1)
(or; the speed changes by . . . meters per second, every second).</td><td>Answer
The train’s acceleration is . . . . . .m/s/s (meters per second per second) (1)</td></tr>
<tr><td colspan="2">Any note or comment for us?</td></tr>
</table>

(Postest 1) Physics 1800 Fall 2012. Name: .... Group:..........................

A problem on the acceleration concept

Problem. *A car accelerates uniformly at 3 m/s/s in order to overtake another car. If the initial speed was 30 m/s and the car accelerated until its speed was 36 m/s, for how long did it accelerate? (7)*

Rough sketch of situation, labeled with the given quantities (1)

Solution by word-logic and mental arithmetic (2)

Re-cast the question

The given information is "a car accelerates at 3 m/s/s". Re-word this without using the terminology "accelerates" to bring out the underlying conceptual meaning of the information, in terms of speed changes.
Word-logic statements and mental arithmetic

Answer
The time taken accelerating was. . . . . . seconds. (1)

Solution by symbolic method and algebra (show the relevant solution then substitution) (2)

Answer
The time taken accelerating was. . . . . . seconds. (1)

Any note or comment for us?

(Pretest 2)

Name: . . . . Group. . . . .

Average speed and mean speed basic problems

Average speed basic problems

In each case below, work out the average speed during the motion.

a) A car travels 150 miles in 3 hours,.

b) A sprinter takes 10 seconds to run the 100 meter dash.

c) A car takes 4 seconds to brake to a stop in a distance of 20 meters

Does your calculation of average speed depend on whether the speed varies or not during the motion? Yes [ ] No [ ]

Mean speed value basic problems: for uniformly accelerated motions

In the cases below, a car is changing speed uniformly. Work out the *mean speed value* in each case.

a) A car speeds up uniformly from a speed of 10 mph to 20 mph.

b) A car starts from rest and reaches a speed of 40 mph.

c) A car is traveling at 10 m/s and slows to a stop.

d) A car slows down from 50 mph to 30 mph.

(Pretest 2)

Name: . . . . Group. . . . .

Car braking ahead of a speed trap – will the driver get a ticket?

A car is traveling in a residential neighborhood at 16 m/s (nearly 60 mph, which is above the speed limit of 10 m/s or about 35 mph). The driver suddenly sees a speed trap 30 meters ahead and brakes hard for 2 seconds until the car has slowed down to the speed limit, then takes her foot off the brakes. The speed trap consists of two wires close together laid across the road, and when a car passes over them its speed is determined.

a) What **distance** does the car travel while braking?
   Do this both by conceptual word-logic and by algebraic symbols.

By conceptual word-logic: and mental arithmetic:

Algebraically:

b) Given that the car was 30 meters from the wires when the driver started braking, will the car still be exceeding the speed limit when it reaches the wires, or will the driver be OK (no ticket)? Say why.

c) Determine the acceleration of the car while braking. Does this seem a reasonable value? Why? In what direction is this acceleration?

(Pretest 2)

Name: . . . . Group. . . . .

How deep is the well?

Villagers want to know the depth of a deep water well. One of them drops a large stone down the well and counts off 3 seconds before hearing the sound of the splash below. She says “Our well is 3 seconds deep! No, just kidding, we can figure it out”. How deep is the well?

(Sound travels much faster than the stone so the time for the sound to travel up can be ignored).

(Posttest 2)

Name: . . . . Group. . . . .

Given information for the exam: the acceleration due to gravity is 10 m/s/s

Average speed and mean speed problems

Average speed basic problems (8 points)

In each case below, work out the average speed during the motion.

a) A train travels 120 miles in 2 hours. (1)
b) A sprinter takes 9 seconds to travel a distance of 90 meters. (1)
c) A car takes 5 seconds to brake to a stop in a distance of 25 meters(1)

Does your calculation of average speed depend on whether the speed varies or not during the motion? Yes [ ] No [ ] (1)

Mean speed value problems: for uniformly accelerated motions

In the cases below, a car is changing speed uniformly, and initial and final speeds are given. Work out the *mean speed value* in each case. That is, the speed value midway between the initial and final values.

a) A car speeds up uniformly from a speed of 10 mph to 20 mph. (1)
b) A car starts from rest and reaches a speed of 40 mph. (1)
c) A car is traveling at 10 m/s and slows to a stop. (1)
d) A car slows down from 50 mph to 30 mph. (1)

(Posttest 2)

Name: . . . . Group. . . . .

Given information for the exam: the acceleration due to gravity is 10 m/s/s

Train slowing down

A heavy freight train is traveling at speed 25 m/s. (nearly 55 mph). It must slow down to 5 m/s as it nears the city. The driver applies the brake for 100 s to do so.

What **distance** does the freight train travel while braking? (2)
Do this both by conceptual word-logic and by algebraic symbols.

By conceptual word-logic: and mental arithmetic:

Algebraically (show and use the relevant formula):

a) Given that the fright train was 2000 meters from the city outskirts when the driver started braking, will the train have slowed down enough when it reaches the outskirts? Say why. (1)

b) Determine the train's acceleration while braking. Does this seem a reasonable value for a train? Why? In what direction is this acceleration? (2)

(Posttest 2)

Name: . . . . Group. . . . .

Climber and cliff (6)

A climber wants to use a rope to rappel down a high cliff and wonder whether her rope is long enough to reach the bottom. So, she drops a rock over the edge and sees it hit the ground 4 seconds later. How high is the cliff?

Appendix E

Attitude Survey

Name____Section:____________

Attitude Survey

Opinions and preferences about formative and summative assessment

*Formative assessme*nt. The primary purpose of formative assessment is to get information to improve learning and teaching. Formative assessment -- continued timely feedbacks on understanding during learning -- gives students the opportunity to see what they do or don't understand so far, to learn and improve their understanding of the lesson topic. Formative assessment involves feedback but usually not points or grades. Students may get a chance to try again and improve. The system also provides the teacher with information so adjust the lesson accordingly.

Summative assessment. Conventionally, assessment has mostly been used s*ummatively*, i.e., quizzes and exams for scoring student performance, awarding points toward grades and ranking. Summative exams are also a way of providing extrinsic motivation for studying and doing well.

Each form of assessment has a number strategies that can be useful. Student work can be instructor-assessed for both types of assessment, but peer- and self-assessment can also be particularly useful for formative assessment. In formative assessment, feedback should be prompt, either in class by concept checks or by the next session for homework and tests.

You are invited to give your comments and views on formative and summative assessment as used in Physics 1800.

Formative Assessment

1. Do you like formative assessment in Physics 1800? Why or why not? Are there aspects you like and others not? Explain.

2. Does formative assessment including the strategies mentioned above help your **learning** of the topic? Why or why not?

3. **Which formative strategies are your favorite** or help you understand the lesson or discussion better? **Which one is not your favorite?** (e.g. self-and peer-assessment, concept check with feedback while learning, formative quiz, discussing objectives, etc.) You can choose more than one strategy. Explain your thinking about each strategy.

4. Do you feel any pressure with formative assessment— just getting feedback without points while learning? Why or why not? How does it affect your learning?

5. Do you learn more when you can re-do you work? **Why or why not?**

6. Are you willing to revise or improve formative work, e.g., homework or quiz question, when you will not get a grade? **Why or why not?**

Summative Assessment

7. Do you like *summative* assessment in Physics 1800? Why or why not? Are there aspects you like and others not? Explain.

8. Does summative assessment including strategies mentioned above help your **learning**? **Why or why not?**

9. Which summative strategies are your favorite or help you to understand the concept? Which one is not your favorite? Why? You can choose more than one strategy.

10. Do you feel pressure with summative assessment -- working for points or grades while learning? Why or why not? How does it affect your learning?

General

11. Do you put as much effort into formative homework or quizzes as you do on summative homework (for grades), when you know that there are no grades involved? Why or why not?

12. If you can choose the kind of assessment to be used in PHYS 1800 course, will you prefer only formative assessment, only summative assessment, or both? Why or why not?

13. Do you think that peer assessment, if well done, is useful or not for learning? Should we keep it? Explain.

14. Do you think that self-assessment, if well done, is useful or not for learning? Should we keep it? Explain.

15. Give your thoughts on the value of having concept checks and feedback immediately after learning a concept? Also say whether clickers are a good way to do this, or would it be fine without?

Appendix F

ICES Course Evaluation Results

Western Michigan University
ICES Online

INSTRUCTOR & COURSE EVALUATION SYSTEM ices Online

Course Evaluation Results

PHYS 1800 - Physics: Inquiry and Insights **Fall, 2012**
Section 105, Lab/Discussion (Chaiphat Plybour)
TR, 1pm, WOOD 01413

Evaluations were completed by **11** out of **22** students (50.0%).

Click a plus or minus symbol to expand or collapse an open-ended item.

Demographic Items

Class Status:

Freshman	Sophomore	Junior	Senior	Graduate	Other	Omitted
18% (2)	18% (2)	9% (1)	55% (6)	-	-	-

This course was:

Elective	Required, But a Choice	Specifically Required	Omitted
-	55% (6)	45% (5)	-

This course was in my:

Major	Minor	Graduate Program	Other	Omitted
82% (9)	18% (2)	-	-	-

What was your pre-course opinion of the instructor?

Negative	No Opinion	Positive	Omitted
-	91% (10)	9% (1)	-

What was your pre-course opinion of the course?

Negative	No Opinion	Positive	Omitted
9% (1)	73% (8)	18% (2)	-

Expected grade in the course:

A	B	C	D	E	Omitted
82% (9)	18% (2)	-	-	-	-

University Items

Rate the instructor's overall teaching effectiveness. [Exceptionally Low ... Exceptionally High]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	9% (1)	18% (2)	73% (8)	-	**4.64**	0.67

Rate the overall quality of this course. [Exceptionally Low ... Exceptionally High]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	-	64% (7)	36% (4)	-	**4.36**	0.50

How much have you learned in this course? [Very Little ... A Great Deal]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	9% (1)	55% (6)	36% (4)	-	**4.27**	0.65

Was the grading system for the course explained? [No, Not At All ... Yes, Very Well]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	9% (1)	9% (1)	82% (9)	-	**4.73**	0.65

The instructor seemed well prepared for classes. [No, Seldom ... Yes, Always]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	-	9% (1)	91% (10)	-	**4.91**	0.30

The instructor promoted an atmosphere conducive to work and learning. [Strongly Disagree ... Strongly Agree]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	-	-	100% (11)	-	**5.00**	0.00

College Core Items

Do you feel course objectives were accomplished? [No, Not At All ... Yes, To A Great Extent]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	-	36% (4)	64% (7)	-	**4.64**	0.50

The instructor was a good classroom leader. [Strongly Disagree ... Strongly Agree]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	-	18% (2)	82% (9)	-	**4.82**	0.40

Was the instructor able to explain difficult material to your satisfaction? [Almost Never ... Almost Always]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	9% (1)	45% (5)	45% (5)	-	**4.36**	0.67

How accessible was the instructor for student conferences about the course? [Never Available ... Available Regularly]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	-	18% (2)	82% (9)	-	**4.82**	0.40

Department Core Items

PHYS

How well did examination questions reflect content and emphasis of the course? [Poorly Related ... Well Related]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	-	27% (3)	73% (8)	-	**4.73**	0.47

Did this course increase your interest in the subject matter? [No, Not Much ... Yes, Greatly]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	45% (5)	27% (3)	27% (3)	-	**3.82**	0.87

The instructor's knowledge of subject was: [Poor ... Excellent]

1	2	3	4	5	Omitted	Mean	St. Dev

-	-	-	-	100% (11)	-	**5.00**	0.00

Was the instructor enthusiastic about teaching? [Very Unenthusiastic ... Very Enthusiastic]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	-	-	100% (11)	-	**5.00**	0.00

The instructor gave explanations/examples that were clearly to the point. [Strongly Disagree ... Strongly Agree]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	9% (1)	27% (3)	64% (7)	-	**4.55**	0.69

The instructor stimulated my intellectual curiosity. [Almost Never ... Almost Always]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	18% (2)	36% (4)	45% (5)	-	**4.27**	0.79

Rating Scale Item Means

	1 2 3 4 5	
Rate the instructor's overall teaching effectiveness.		**4.64**
Rate the overall quality of this course.		**4.36**
How much have you learned in this course?		**4.27**
Was the grading system for the course explained?		**4.73**
The instructor seemed well prepared for classes.		**4.91**
The instructor promoted an atmosphere conducive to work and learning.		**5.00**
Do you feel course objectives were accomplished?		**4.64**
The instructor was a good classroom leader.		**4.82**
Was the instructor able to explain difficult material to your satisfaction?		**4.36**
How accessible was the instructor for student conferences about the course?		**4.82**
How well did examination questions reflect content and emphasis of the course?		**4.73**
Did this course increase your interest in the subject matter?		**3.82**
The instructor's knowledge of subject was:		**5.00**
Was the instructor enthusiastic about teaching?		**5.00**
The instructor gave explanations/examples that were clearly to the point.		**4.55**
The instructor stimulated my intellectual curiosity.		**4.27**

= below 3.0 / = 3.0 - 4.0 / = above 4.0

ICES Open-Ended Items

What are the major strengths of the instructor/course?

Western Michigan University ICES Online

ices INSTRUCTOR & COURSE EVALUATION SYSTEM Online

Course Evaluation Results

PHYS 1800 - Physics: Inquiry and Insights **Fall, 2012**
Section 125, Lab/Discussion (Chaiphat Plybour)
TR, 10:30am, WOOD 01413

Evaluations were completed by **5** out of **15** students (33.3%).

Click a plus or minus symbol to expand or collapse an open-ended item.

Demographic Items

Class Status:

Freshman	Sophomore	Junior	Senior	Graduate	Other	Omitted
60% (3)	-	-	40% (2)	-	-	-

This course was:

Elective	Required, But a Choice	Specifically Required	Omitted
-	40% (2)	60% (3)	-

This course was in my:

Major	Minor	Graduate Program	Other	Omitted
60% (3)	20% (1)	20% (1)	-	-

What was your pre-course opinion of the instructor?

Negative	No Opinion	Positive	Omitted
-	100% (5)	-	-

What was your pre-course opinion of the course?

Negative	No Opinion	Positive	Omitted
60% (3)	40% (2)	-	-

Expected grade in the course:

A	B	C	D	E	Omitted
40% (2)	60% (3)	-	-	-	-

University Items

Rate the instructor's overall teaching effectiveness. [Exceptionally Low ... Exceptionally High]

1	2	3	4	5	Omitted	Mean	St. Dev
40% (2)	20% (1)	40% (2)	-	-	-	**2.00**	1.00

Rate the overall quality of this course. [Exceptionally Low ... Exceptionally High]

1	2	3	4	5	Omitted	Mean	St. Dev
40% (2)	20% (1)	40% (2)	-	-	-	**2.00**	1.00

How much have you learned in this course? [Very Little ... A Great Deal]

1	2	3	4	5	Omitted	Mean	St. Dev
20% (1)	20% (1)	20% (1)	20% (1)	20% (1)	-	**3.00**	1.58

Was the grading system for the course explained? [No, Not At All ... Yes, Very Well]

1	2	3	4	5	Omitted	Mean	St. Dev
40% (2)	20% (1)	40% (2)	-	-	-	**2.00**	1.00

The instructor seemed well prepared for classes. [No, Seldom ... Yes, Always]

1	2	3	4	5	Omitted	Mean	St. Dev
-	40% (2)	20% (1)	20% (1)	20% (1)	-	**3.20**	1.30

The instructor promoted an atmosphere conducive to work and learning. [Strongly Disagree ... Strongly Agree]

1	2	3	4	5	Omitted	Mean	St. Dev
20% (1)	20% (1)	60% (3)	-	-	-	**2.40**	0.89

College Core Items

Do you feel course objectives were accomplished? [No, Not At All ... Yes, To A Great Extent]

1	2	3	4	5	Omitted	Mean	St. Dev
40% (2)	20% (1)	20% (1)	20% (1)	-	-	**2.20**	1.30

The instructor was a good classroom leader. [Strongly Disagree ... Strongly Agree]

1	2	3	4	5	Omitted	Mean	St. Dev
40% (2)	20% (1)	40% (2)	-	-	-	**2.00**	1.00

Was the instructor able to explain difficult material to your satisfaction? [Almost Never ... Almost Always]

1	2	3	4	5	Omitted	Mean	St. Dev
40% (2)	60% (3)	-	-	-	-	**1.60**	0.55

How accessible was the instructor for student conferences about the course? [Never Available ... Available Regularly]

1	2	3	4	5	Omitted	Mean	St. Dev
-	40% (2)	40% (2)	-	20% (1)	-	**3.00**	1.22

Department Core Items

PHYS

How well did examination questions reflect content and emphasis of the course? [Poorly Related ... Well Related]

1	2	3	4	5	Omitted	Mean	St. Dev
-	40% (2)	40% (2)	-	20% (1)	-	**3.00**	1.22

Did this course increase your interest in the subject matter? [No, Not Much ... Yes, Greatly]

1	2	3	4	5	Omitted	Mean	St. Dev
60% (3)	-	20% (1)	20% (1)	-	-	**2.00**	1.41

The instructor's knowledge of subject was: [Poor ... Excellent]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	20% (1)	40% (2)	40% (2)	-	**4.20**	0.84

Was the instructor enthusiastic about teaching? [Very Unenthusiastic ... Very Enthusiastic]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	40% (2)	20% (1)	40% (2)	-	**4.00**	1.00

The instructor gave explanations/examples that were clearly to the point. [Strongly Disagree ... Strongly Agree]

1	2	3	4	5	Omitted	Mean	St. Dev
60% (3)	-	20% (1)	20% (1)	-	-	**2.00**	1.41

The instructor stimulated my intellectual curiosity. [Almost Never ... Almost Always]

1	2	3	4	5	Omitted	Mean	St. Dev
60% (3)	-	40% (2)	-	-	-	**1.80**	1.10

Rating Scale Item Means

Item	1 2 3 4 5	Mean
Rate the instructor's overall teaching effectiveness.		**2.00**
Rate the overall quality of this course.		**2.00**
How much have you learned in this course?		**3.00**
Was the grading system for the course explained?		**2.00**
The instructor seemed well prepared for classes.		**3.20**
The instructor promoted an atmosphere conducive to work and learning.		**2.40**
Do you feel course objectives were accomplished?		**2.20**
The instructor was a good classroom leader.		**2.00**
Was the instructor able to explain difficult material to your satisfaction?		**1.60**
How accessible was the instructor for student conferences about the course?		**3.00**
How well did examination questions reflect content and emphasis of the course?		**3.00**
Did this course increase your interest in the subject matter?		**2.00**
The instructor's knowledge of subject was:		**4.20**
Was the instructor enthusiastic about teaching?		**4.00**
The instructor gave explanations/examples that were clearly to the point.		**2.00**
The instructor stimulated my intellectual curiosity.		**1.80**

= below 3.0 / = 3.0 - 4.0 / = above 4.0

ICES Open-Ended Items

What are the major strengths of the instructor/course?

Western Michigan University ICES Online

INSTRUCTOR & COURSE EVALUATION SYSTEM ices Online

Course Evaluation Results

PHYS 1800 - Physics: Inquiry and Insights **Spring, 2013**
Section 135, Lab/Discussion (Chaiphat Plybour)
TR, 10:30am, WOOD 01413

Evaluations were completed by **6** out of **24** students (25.0%).

Click a plus or minus symbol to expand or collapse an open-ended item.

Demographic Items

Class Status:

Freshman	Sophomore	Junior	Senior	Graduate	Other	Omitted
17% (1)	17% (1)	17% (1)	50% (3)	-	-	-

This course was:

Elective	Required, But a Choice	Specifically Required	Omitted
-	33% (2)	67% (4)	-

This course was in my:

Major	Minor	Graduate Program	Other	Omitted
33% (2)	50% (3)	17% (1)	-	-

What was your pre-course opinion of the instructor?

Negative	No Opinion	Positive	Omitted
-	83% (5)	17% (1)	-

What was your pre-course opinion of the course?

Negative	No Opinion	Positive	Omitted
33% (2)	50% (3)	17% (1)	-

Expected grade in the course:

A	B	C	D	E	Omitted
50% (3)	50% (3)	-	-	-	-

University Items

Rate the instructor's overall teaching effectiveness. [Exceptionally Low ... Exceptionally High]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	-	17% (1)	83% (5)	-	**4.83**	0.41

Rate the overall quality of this course. [Exceptionally Low ... Exceptionally High]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	-	33% (2)	67% (4)	-	**4.67**	0.52

How much have you learned in this course? [Very Little ... A Great Deal]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	-	17% (1)	83% (5)	-	**4.83**	0.41

Was the grading system for the course explained? [No, Not At All ... Yes, Very Well]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	-	33% (2)	67% (4)	-	**4.67**	0.52

The instructor seemed well prepared for classes. [No, Seldom ... Yes, Always]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	-	17% (1)	83% (5)	-	**4.83**	0.41

The instructor promoted an atmosphere conducive to work and learning. [Strongly Disagree ... Strongly Agree]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	-	33% (2)	67% (4)	-	**4.67**	0.52

College Core Items

Do you feel course objectives were accomplished? [No, Not At All ... Yes, To A Great Extent]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	-	-	100% (6)	-	**5.00**	0.00

The instructor was a good classroom leader. [Strongly Disagree ... Strongly Agree]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	17% (1)	-	83% (5)	-	**4.67**	0.82

Was the instructor able to explain difficult material to your satisfaction? [Almost Never ... Almost Always]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	-	17% (1)	83% (5)	-	**4.83**	0.41

How accessible was the instructor for student conferences about the course? [Never Available ... Available Regularly]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	-	-	83% (5)	17% (1)	**5.00**	0.00

Department Core Items

PHYS

How well did examination questions reflect content and emphasis of the course? [Poorly Related ... Well Related]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	-	17% (1)	83% (5)	-	**4.83**	0.41

Did this course increase your interest in the subject matter? [No, Not Much ... Yes, Greatly]

1	2	3	4	5	Omitted	Mean	St. Dev
17% (1)	-	-	-	83% (5)	-	**4.33**	1.63

The instructor's knowledge of subject was: [Poor ... Excellent]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	-	17% (1)	83% (5)	-	**4.83**	0.41

Was the instructor enthusiastic about teaching? [Very Unenthusiastic ... Very Enthusiastic]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	-	17% (1)	67% (4)	17% (1)	**4.80**	0.45

The instructor gave explanations/examples that were clearly to the point. [Strongly Disagree ... Strongly Agree]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	-	33% (2)	67% (4)	-	**4.67**	0.52

The instructor stimulated my intellectual curiosity. [Almost Never ... Almost Always]

1	2	3	4	5	Omitted	Mean	St. Dev
-	-	17% (1)	17% (1)	67% (4)	-	**4.50**	0.84

Rating Scale Item Means

	1 2 3 4 5	
Rate the instructor's overall teaching effectiveness.		**4.83**
Rate the overall quality of this course.		**4.67**
How much have you learned in this course?		**4.83**
Was the grading system for the course explained?		**4.67**
The instructor seemed well prepared for classes.		**4.83**
The instructor promoted an atmosphere conducive to work and learning.		**4.67**
Do you feel course objectives were accomplished?		**5.00**
The instructor was a good classroom leader.		**4.67**
Was the instructor able to explain difficult material to your satisfaction?		**4.83**
How accessible was the instructor for student conferences about the course?		**5.00**
How well did examination questions reflect content and emphasis of the course?		**4.83**
Did this course increase your interest in the subject matter?		**4.33**
The instructor's knowledge of subject was:		**4.83**
Was the instructor enthusiastic about teaching?		**4.80**
The instructor gave explanations/examples that were clearly to the point.		**4.67**
The instructor stimulated my intellectual curiosity.		**4.50**

= below 3.0 / = 3.0 - 4.0 / = above 4.0

ICES Open-Ended Items

What are the major strengths of the instructor/course?