Chapter 5. Measurements of Expectations:
The Maryland Physics Expectations (MPEX) Survey

I. OVERVIEW

A. What are Expectations?

In chapter 2, several examples were used from the Physics Education Research literature to demonstrate that students do not come into the introductory physics class as blank slates. In the previous chapter, we showed that concept tests like the FCI[1] and FMCE[2] have determined that many students have common sense beliefs about physics concepts that can hinder their learning of the physics concepts taught in the introductory class. However, the discussion in chapter 2 showed that it is not only physics concepts that a student brings into the physics classroom. Each student, based on his or her own experiences, brings to the physics class a set of expectations about what sorts of things they will learn, what skills will be required, and what they will be expected to do. In addition, their view of the nature of scientific information affects how they interpret what they hear. As we saw in chapter 2, students' expectations can have a strong effect on what they get out of a physics class.

B. Why Study Expectations?

While it is important for students to learn a functional understanding of the physics concepts, it is also important for the students to develop expectations favorable for developing a deep, functional understanding of physics. As we saw in the examples of Mazur,[3] Hammer,[4] and Tobias[5] in chapter 2, the students' expectations have a significant effect on what the students do to learn physics, their idea of what physics is, the type of understanding they build, and ultimately what they get out of the class. In many traditional introductory physics classes, some students' expectations and their desire for a good grade with a minimum of work may lead them to a false sense of what learning physics is about. Like Hammer's student Liza in chapter 2, these students may believe that by memorizing formulas and problems solutions and using what they've memorized to solve typical end of chapter problems on exams, that they have successfully learned physics in addition to doing well on the exams.

C. Why a Survey?

The FCI and other concept tests like it, as described in the previous chapter, have played a major role in convincing physics instructors of the validity of studies of the nature and persistence of students' common sense beliefs in traditional instruction as well as evaluations of curricula designed to improve students' conceptual understanding. There are several reasons why tests like the FCI have become an almost indispensable assessment instrument to physics instructors. Unlike many research-based assessment methods, they can easily be used and interpreted by instructors who are not physics education researchers. They can be used to roughly determine the distribution of Newtonian and common sense beliefs over a whole class. In addition, they can be used as a pre/post evaluation tool to see if and how student responses change. Several curricula that take into account the students' common sense beliefs have shown significant improvements in students' conceptual understanding as measured by multiple-choice concepts tests compared to traditional instruction (see chapter 4).

However, while a great deal is known about student common sense beliefs about concepts in introductory physics courses, very little is known about the nature and effects of student expectations in physics classes. The few studies that exist have mainly used individual student interviews.[6] What they have taught us about student learning is alarming, but we need to learn how serious and how widespread the problems demonstrated by these studies are. If inappropriate expectations play a significant role in the difficulties our students commonly have with introductory calculus-based physics, physics education researcher and physics instructors need way to track and document them. In particular, we need an instrument equivalent to study this issue. It needs to be convenient to give and easy to analyze for a physics instructor who is not a physics education specialist to use with their own classes as a pre/post evaluation. This would give us a sense of the distribution and evolution of these expectation "misconceptions" in the introductory physics classroom.

In this chapter, I describe our development of the MPEX survey, a 34-item Likert-scale (agree-disagree) survey that probes students' cognitive expectations.[7] The MPEX survey is one of several instruments being developed to meet this need.[8] Note that because of their brevity, these instruments are surveys and not diagnostics. (See the discussion on this issue in chapter 4, page 129.) While they are useful as an instrument for learning how broad and prevalent expectation misconceptions are in a class, they are often less reliable measures than diagnostics for evaluating individual students.[9]

D. Chapter Layout

In this chapter, I describe the development, the structure, and testing for reliability and validity of the MPEX Survey. The MPEX survey results from introductory physics classes are presented in chapter 10. In section II, I describe the development of the MPEX survey over the last five years by Redish, Steinberg, and the author. In section III, I describe both the 34 survey items and the construction of seven dimensions of expectations. Section IV contains results on two tests of the validity of the MPEX survey deals, the results of the survey with five calibration groups and the results from interviews with students taking the survey and going over their responses. Tests of the MPEX survey reliability including factor analysis, Cronbach alpha, and reproduction of results are presented in section V together with a brief discussion of the survey data's statistical uncertainty.

II. DEVELOPMENT OF THE SURVEY

This study began as an attempt to extend Hammer's work. Detailed interviews are too slow and expensive to allow the determination of distribution functions – how many students in various classes hold particular views. We want to know such things as what is the distribution of students' expectations coming into the introductory physics class, whether students in universities or junior colleges have different attitudes or distributions of attitudes, and whether the distribution of attitudes found in a study at a particular college or university are representative of what is found throughout the country. To determine these distribution functions requires an efficient and easily delivered instrument — a questionnaire or short answer test. A major goal of the MPEX project has been to develop and evaluate such an instrument.

We also want to know the role of dynamics. While Hammer did not observe students changing their views, the studies by Perry[10] and Belenky *et al.*[11] (see the discussion on expectations in chapter 2) found that many of their young adult subjects were able to evolve more sophisticated expectations about general knowledge. Although neither Perry's nor Belenky *et al.*'s studies focused specifically on scientific situations, these two studies give us hope that many students can be moved from Hammer's category B to the more sophisticated and scientific category A. A second goal of this project is to understand whether such transitions are possible and to begin to specify what activities in a university physics class can help such transitions take place.

Redish and Steinberg began to develop the MPEX survey in the 1992 fall quarter at the University of Washington. The first version of the expectations survey was delivered in the spring quarter of 1993 to students in the three-quarter calculus-based introductory physics sequence at the University of Washington. The students in first-quarter course were surveyed at the beginning of the quarter and the students in the third-quarter course were interviewed at the end of the quarter. Responses from more than 100 students were obtained from each course. These students in the introductory calculus-based physics class were given a survey with 51 statements about the nature of physics, the study of physics, and their relation to it. They rated these statements on a five point Likert scale[12] from strongly disagree (1) to strongly agree (5). The survey was also given to eight faculty members who had taught the class before. The survey items were chosen as a result of a detailed literature review, discussions with physics faculty, and the designers' combined 35 years of teaching experience.

Upon his return to University of Maryland, Redish gave a new version of the survey to students in an experimental class of the engineering physics sequence. Based on the analysis of the University of Washington results, this second version of the survey was pared to 35 statements. My involvement with this project began with the second semester of this sequence. Up to this point, the surveys were given to the students as paper and pencil instruments that had space for comments. For the second version of the survey, I validated the survey items by interviewing students and by examining students' written comments from the class surveys. Beginning with the third version of the survey and widespread distribution with other schools, Redish and I developed a scantron version of the survey to allow for easy processing of large numbers of completed surveys from the 17 participating colleges and universities.

We[13] validated the survey items from the current version of the survey (version 4.0) in a number of ways: by discussion with other faculty and physics education experts, through student interviews, by giving the survey to a variety of "experts", and through repeated delivery of the survey to groups of students. I conducted over 125 hours of interviews with over 100 students at eight of the participating colleges and universities.

The MPEX survey was iteratively refined through testing in more than 17 universities and colleges during the last four years. The final version of the survey presented here has 34 items and typically takes twenty to thirty minutes to complete.[14] The survey items are designed so that those students with sophisticated expectations will agree with some items and disagree with others. The results of the MPEX survey given in the introductory courses at ten colleges and universities using traditional and research-based curricula are given in chapter 10 of this dissertation. (The curricula and their implementations at the ten schools are described in chapter 8.)

III. CHOOSING THE ITEMS OF THE MPEX SURVEY

A. Cluster Descriptions

The cognitive beliefs that we have referred to as "student expectations" clearly are complex and contain many facets. We decided to focus on six issues or dimensions along which we might categorize student attitudes about learning and physics. Three of these are taken from Hammer's study and we have added three of our own. Building on the work of Perry and Songer and Linn,[15] Hammer proposed three dimensions along which to classify student beliefs about the nature of learning physics:

1. *Independence* — beliefs about learning physics — whether it means receiving information or involves an active process of reconstructing one's own understanding;
2. *Coherence* — beliefs about the structure of physics knowledge — as a collection of isolated pieces or as a single coherent system;
3. *Concepts* — beliefs about the content of physics knowledge — as formulas or as concepts that underlie the formulas.

In the MPEX survey, we also seek to probe the three additional dimensions described below:

4. *Reality Link* — beliefs about the connection between physics and reality — whether physics is unrelated to experiences outside the classroom or whether it is useful to think about them together;
5. *Math Link* — beliefs about the role of mathematics in learning physics — whether the mathematical formalism is just used to calculate numbers or is used as a way of representing information about physical phenomena;
6. *Effort* — beliefs about the kind of activities and type of work necessary to make sense out of physics — whether they expect to think carefully and evaluate what they are doing based on available materials and feedback or not.

The extreme views associated with each of these variables are given in Table 5-1. We refer to the extreme view that agrees with that of most mature scientists as the “expert” or “favorable” view, and the view that agrees with that of most beginning students as the “novice” or “unfavorable” view. The survey items that have been selected to probe these six expectation dimensions are given in the right hand column of the table. We

Table 5-1. MPEX dimensions of student expectations

	Favorable	**Unfavorable**	**MPEX Items**
independence	takes responsibility for constructing own understanding	takes what is given by authorities (teacher, text) without evaluation	1, 8, 13, 14, 17, 27
coherence	believes physics needs to be considered as a connected, consistent framework	believes physics can be treated as unrelated facts or "pieces"	12, 15, 16, 21, 29
concepts	stresses understanding of the underlying ideas and concepts	focuses on memorizing and using formulas	4, 19, 26, 27, 32
reality link	believes ideas learned in physics are relevant and useful in a wide variety of real contexts	believes ideas learned in physics has little relation to experiences outside the classroom	10, 18, 22, 25
math link	considers mathematics as a convenient way of representing physical phenomena	views the physics and the math as independent with little relationship between them	2, 6, 8, 16, 20
effort	makes the effort to use information available and tries to make sense of it	does not attempt to use available information effectively	3, 6, 7, 24, 31

refer to the collection of survey items designed to probe a particular dimension as a "cluster." Note that there is some overlap, as these dimensions are not independent variables.

B. Survey Items & Responses

The survey has undergone several iterations of development and testing. A copy of the current pre-course versions (version 4.0) of both the scantron and paper & pencil survey forms are included in Appendix C. Note that the pre- and post-course versions of the forms have the same survey items but with different tenses to differentiate between what students expect to do at the beginning of the sequence and what they have done at the end of the course. In the current section, each of the thirty-four survey items are presented and interpreted. The items are listed under their corresponding clusters.

1. Student beliefs about learning physics: The independence cluster

One characteristic of the binary thinker, as reported by Perry and Belenky *et al.*,[16] is the view that knowledge comes from an authoritative source, such as an instructor or a textbook, and it is the responsibility of that authority to convey this knowledge to the student. This is a key element of the transmissionist view of learning, that knowledge is presented to the student who learns it and repeats it back on assignments and exams. The more mature students understand that developing knowledge is a participatory process. They understand that they must actively think about what they are learning to build an understanding of the course material. Hammer classifies these two extreme views as "by authority" and. "independent."[17] The survey items 1, 8, 13, 14, 17, and 27 are designed to probe students' views along this dimension.

Item 1. *All I need to do to understand most of the basic ideas in this course is just read the text, work most of the problems, and/or pay close attention in class.*

At first glance, item 1 seems like a fairly innocuous question that all students should agree with. After all this is what instructors expect most students to do to succeed in the course. It is only when one realizes that with the word "just" there is nothing in this statement about understanding the material that the point of this item becomes clear. As students become more independent in building their own understanding, they start to disagree with this item. Disagreeing with this item is a strong indication that the student is moving beyond "binary" thinking.

Item 8. *In this course, I do not expect to understand equations in an intuitive sense; they just have to be taken as givens.*

All instructors would expect their students to disagree with this item. However, many students attempt to learn physics by memorizing equations and mathematical problem solutions without understanding. Some students do this because they feel this is the most efficient way to learn the material. Others like Hammer's student Ellen[18] do it because they haven't been able to understand the material and this is their only way to succeed. However, because many students recognize that they should disagree with this statement, agreement with this statement is a strong indication of learning by authority. Since in this view, concepts are divorced from the equations, this item is also part of the Math Link cluster.

Item 13. *My grade in this course is primarily determined by how familiar I am with the material. Insight or creativity has little to do with it.*

As Tobias' student observers commented,[19] many students believe that if you know physics, there is one and only one right way to do things like solving problems, particularly at the introductory course level. When students believe this, they do not see the value of insight or creativity because they are focusing on learning the "right" answers. They see science as a set of facts laid down by authority rather than as a process to understand physical phenomenon. Students who agree with this item are emphasizing learning the course material rather than trying to understand and think about the course material.

Item 14. *Learning physics is a matter of acquiring knowledge that is specifically located in the laws, principles, and equations given in class and/or in the textbook.*

While learning physics involves learning the laws, principles, and equations, it is more than that. Learning physics also involves building a functional understanding of the laws, principles, and equations including understanding their implications and interconnections. "Independent" students who are actively building their own understanding of the material will disagree with this statement.

Item 17. *Only very few specially qualified people are capable of really understanding physics.*

Many students come into the introductory physics class believing that they cannot do physics but maybe they can learn something about it. One of our goals of instruction is that students learn how to do physics themselves, that is, to make observations, to generalize from what they observe and construct models, and to make and test predictions rather than to just receive physics knowledge. We expect that students who believe that they can do physics to disagree with this statement

Item 27. *Understanding physics basically means being able to recall something you've read or been shown.*

Item 27 is an example of the extreme “by authority” view. We would hope that most students would disagree with this item. Since such a view precludes conceptual understanding, item 27 is also part of the concepts cluster.

2. Student beliefs about the structure of physics knowledge:

The coherence cluster

Most physics faculty feel strongly that students should see physics as a coherent, consistent structure. A major strength of the scientific worldview is its ability to describe many complex phenomena with a few simple laws and principles. Students, like Liza from Hammer’s study in chapter 2, who emphasize science as a collection of facts, fail to see the integrity of the structure, an integrity that is both epistemologically convincing and useful. The lack of a coherent view can cause students many problems, including a failure to notice errors in their reasoning and an inability to evaluate a recalled item through crosschecks. Survey items 12, 15, 16, 21, and 29 have been included in order to probe student views along this dimension.

Item 12. *Knowledge in physics consists of many pieces of information each of which applies primarily to a specific situation.*

Item 12 reflects the view of students who focus on memorizing information. These students have a great deal to memorize because there is so much material and so many different situations. Their focus is on the equations used in all these different situations, not on the more general equations and principle from which the situation specific equations can be derived. These students tend not to see physics knowledge as a consistent framework and they don’t see the connections and underlying themes in the course material. A student who sees either the framework, the connections, or the underlying themes should disagree with this statement.

> Item 15. *In doing a physics problem, if my calculation gives a result that differs significantly from what I expect, I'd have to trust the calculation.*

As we saw in the discussion of Hammer’s doctoral study in chapter 2, the type B students were very casual about making and breaking associations between different aspects of their knowledge. Since they don’t expect physics to be coherent or even to really make sense, these students tend to trust their calculation more and change their intuition to suit a particular problem. The type A students saw physics as more coherent and were much more cautious about modifying their understanding. They tend to trust their intuition more than their calculations.

> Item 16. *The derivations or proofs of equations in class or in the text have little to do with solving problems or with the skills I need to succeed in this course.*

The derivations of equations show how the equations are related to the coherent framework. They show where they come from and how they relate to the main principles. A student who agrees with item 16 does not see either the relationship or the coherent structure as useful. While this is usually an indication of the students’ expectations, it can also be an indication of the types of physics knowledge valued in the class. Since derivations are an important part of the relationship between concepts and equations, this item is also part of the math link cluster.

> Item 21. *If I came up with two different approaches to a problem and they gave different answers, I would not worry about it; I would just choose the answer that seemed most reasonable. (Assume the answer is not in the back of the book.)*

Coming up with two different answers using two different approaches indicates that something is seriously wrong with at least one of the solutions and perhaps with the students' understanding of physics and how to apply it to problems. A student who sees physics as a set of many equations that apply to many different situations and believes that there is only one right solution would assume that the less reasonable answer was produced from using an incorrect equation. A student who sees physics as a coherent and consistent whole will disagree with item 21.

Item 29. *A significant problem in this course is being able to memorize all the information I need to know.*

A sophisticated student will realize that the large number of different equations and results discussed in a physics text can be structured and organized so that only a small amount of information needs to be memorized and the rest can be easily rebuilt as needed. Item 29 is part of a probe into whether or not students see this structure or are relying on memorizing instead of rebuilding. However, if students are permitted to use a formula sheet or if exams are open book, they may not perceive memorization as a problem. This does not mean that they see the coherence of the material.[20] If extensive information is made available to students during exams, item #29 needs to be interpreted carefully.

3. Student beliefs about the content of physics knowledge: The concepts cluster

The group of items selected for the concepts cluster (items 4, 19, 26, 27, and 32) are intended to probe whether students are viewing physics problems as simply a mathematical calculation or the application of an equation, or whether they are aware of the more fundamental role played by physics concepts and principles in complex problem solving. The intent of these items and the student views that would lead students to agree or disagree with these items is clear.

> Item 4. *Problem solving in physics basically means matching problems with facts or equations and then substituting values to get a number.*
>
> Item 19. *The most crucial thing in solving a physics problem is finding the right equation to use.*
>
> Item 26. *When I solve most exam or homework problems, I explicitly think about the concepts that underlie the problem.*
>
> Item 27. *Understanding physics basically means being able to recall something you've read or been shown.*
>
> Item 32. *To be able to use an equation in a problem (particularly in a problem that I haven't seen before), I need to know more than what each term in the equation represents.*

The intent of these items ranges from statements of extreme novice views like item 27 to the more sophisticated views expressed in items 19 and 26. Students who disagree with item 27 can at least distinguish between memorizing and understanding. Students who rely heavily on equation manipulation to get through the introductory physics course will agree with item 4. However, it is interesting to compare the results of items 4 and 19. A more experienced student could disagree with item 4 and yet still agree with item 19 either because of or despite the use of the words "most crucial." As students become more experienced with complex problem solving, the importance of finding the correct equation decreases and more emphasis is placed on a better understanding of the problems and the conceptual issues involved. Another sign of sophistication in problem solving is the explicit use of physics concepts and principles as queried by item 26. Many novice students will write out equations in problem solutions with little or no idea of the relation between the equations and the concepts or principles. These students often apply whatever equation they can find or remember that seems to have the right quantities to solve the problem. These students tend to disagree with item 32.

4. Student beliefs about the connections between physics and the real world: The reality-link cluster

Although physicists believe that they are learning about the real world when they study physics, the context dependence of cognitive responses (see ref. 5) opens a possible gap between faculty and students. Students may believe that physics is related to the real world in principle, but they may also believe that it has little or no relevance to their personal experience. This can cause problems that are both serious and surprising to faculty. The student who does a calculation of the speed with which a high jumper leaves the ground and comes up with 8000 m/s (as a result of recalling numbers with incorrect units and forgetting to take a square root) may not bother to evaluate that answer and see it as nonsense on the basis of personal experience. When an instructor produces a demonstration that has been "cleaned" of distracting elements such as friction and air resistance, the instructor may see it as displaying a general physical law that is present in the everyday world but that lies "hidden" beneath distracting factors. The student, on the other hand, may believe that the complex apparatus is *required* to produce the phenomenon, and that it does not occur naturally in the everyday world, or is irrelevant to it. A failure to make a link to experience can lead to problems, not just because physics instructors want students to make strong connections between their real-life experiences and what they learn in the classroom, but because learning tends to be more effective and robust when linked to real and personal experiences.

The four items included as the reality-link cluster (items 10, 18, 22, and 25) do not just probe whether the students believe the laws of physics govern the real world. Rather, these items probe whether the students feel that their personal real-world experience is relevant for their physics course and vice versa. In our interviews, we observed that many students show what we would call, following Hammer, an "apparent reality link." That is, they believe that the laws of physics govern the behavior of the real world in principle, but that they do not need to consider that fact relevant or necessary to their physics course.

Item 10. *Physical laws have little relation to what I experience in the real world.*

Item 18. *To understand physics, I sometimes think about my personal experiences and relate them to the topic being analyzed.*

Item 22. *Physics is related to the real world and it sometimes helps to think about the connection, but it is rarely essential for what I have to do in this course.*

Item 25. *Learning physics helps me understand situations in my everyday life.*

Item 10 is fairly obvious and most, but not all, students usually disagree with it at the beginning of the course. Note that the question does not ask if physical laws are related to the real world but if they are related to the students' experiences in the real world. Even many students with novice views will disagree with this item. As students start thinking about what they are learning, they usually start to see examples of physics in everyday situations and agree with Item 25. More sophisticated students will agree with item 18 and disagree with item 22. While many students see examples of physics in everyday life, only a few use their own experiences to help them understand the physics they are learning or in solving problems. While many students find that thinking about physics in connection with the real world is useful, only a fraction of them feel that this helps them in most introductory physics courses.

We expect at least some students to respond as if they believe physics and physical laws are closely related to their experiences, but this connection is not something they use to help them learn physics in or out of class. These students are aware of the connection and some even know that in ideal circumstances they should be using the link, but they don't actively integrate their experiences and the physics they are learning.

5. Student beliefs about the role of mathematics in learning physics: The math-link cluster

An important component of the calculus-based physics course is the development of the students' ability to use abstract and mathematical reasoning in describing and making predictions about the behavior of real physical systems. Expert scientists use mathematical equations as concise summaries of complex relationships among concepts and/or measurements. They can often use equations as a framework on which to construct qualitative arguments. Many introductory students, however, fail to see the deeper physical relationships present in an equation and instead use the math in a purely arithmetic sense, i.e. as a way to calculate a numeric answer. When students have this expectation about equations, there can be a serious gap between what the instructor intends and what the students infer. For example, a professor may go through extensive mathematical derivations in class, expecting the students to use the elements of the derivation to see the structure and sources of the relationships in the equation. The students, on the other hand, may not grasp what the professor is trying to do and reject it as irrelevant "theory." Students who fail to understand the derivation and structure of an equation may be forced to rely on memorization — an especially fallible procedure if they are weak in coherence and have no way to check what they recall. The survey items probing students' apparent expectations[21] of the role of mathematics are 2, 6, 8, 16, and 20.

> Item 2. *All I learn from a derivation or proof of a formula is that the formula obtained is valid and that it is OK to use it in problems.*
>
> Item 6. *I spend a lot of time figuring out and understanding at least some of the derivations or proofs given either in class or in the text.*
>
> Item 8. *In this course, I do not expect to understand equations in an intuitive sense; they just have to be taken as givens.*
>
> Item 16. *The derivations or proofs of equations in class or in the text have little to do with solving problems or with the skills I need to succeed in this course.*
>
> Item 20. *If I don't remember a particular equation needed for a problem in an exam there's nothing much I can do (legally!) to come up with it.*

Items 2, 6, and 16 deal with the different roles of derivations in student learning. Many students find looking at derivations done by the professor in class or in the textbook to be useful, but don't actually work derivations out themselves. And while students might find derivations useful for learning physics, some do not see them as necessary for doing well in the course.

Some students don't try to understand the equations, they just use them to solve problems where the variables match the conditions of the problem. Students who use this approach should agree with item 8. One disadvantage of this approach on exams is that if you forget the correct equation, you are either stuck or forced to use a different approach. However, if a student understands the equation as a relationship and remembers where it comes from and how it connects to the concepts, they could rebuild the forgotten equation. Students who have this ability should disagree with item 20.

6. Student beliefs about studying physics: The effort cluster

Many physics lecturers do not expect most of their students to follow what they are doing in lecture during the lecture itself. They expect students will take good notes and figure them out carefully later. Unfortunately, many students do not take good notes and even those who do may rarely look at them. When physics begins to get difficult for students, most instructors expect them to try to figure things out using a variety of techniques — working through the examples in the book, trying additional problems, talking to friends and colleagues, and in general trying to use whatever resources they have available to make sense of the material. Some students, on the other hand, when things get difficult, may be at a loss for what to do. Some students do not have the idea that if they do not see something right away, there are steps they can take that will eventually help them make sense of the topic.[22] An important component of the tools that help build understanding is making the effort to go over the book and the class activities (lecture in the traditional course format). Another important component is the appreciation that one's current understanding might be wrong, and that mistakes made on homework and exams can give guidance in helping to correct one's errors. This dimension is probed by items 3, 6, 7, 24, and 31 on the survey.

Item 3. *I go over my class notes carefully to prepare for tests in this course.*

Item 6. *I spend a lot of time figuring out and understanding at least some of the derivations or proofs given either in class or in the text.*

Item 7. *I read the text in detail and work through many of the examples given there.*

Item 24. *The results of an exam don't give me any useful guidance to improve my understanding of the course material. All the learning associated with an exam is in the studying I do before it takes place.*

Item 31. *I use the mistakes I make on homework and on exam problems as clues to what I need to do to understand the material better.*

Items 3, 6, and 7 describe activities beyond doing the homework that help students develop a good understanding of the material. We expect that good students will agree with these items. Items 24 and 31 ask if students make the effort to use the graded exams and homework as feedback to debug their understanding of physics or problem solving methods.

7. Other expectation issues

Not all the items in the survey are part of the clusters. Since this is a survey instrument to study student expectations, we have included the following items to probe addition expectation issues:

Item 5. *Learning physics made me change some of my ideas about how the physical world works.*

As we saw in chapters 2 and 4, many students come into introductory physics courses with common-sense beliefs about how the world works that are incompatible with what they learn in the course. Students who develop a good conceptual understanding of physics will need to reconcile what they learn with what they thought they knew about how things work in the physical world. Even students who come into an introductory physics class with a more-physics like view find new applications and subtleties that help them see the world in new ways. These students should agree with this item.

Item 9. *The best way for me to learn physics is by solving many problems rather than by carefully analyzing a few in detail.*

Item 9 is subtler. As mentioned above, as students become more sophisticated learners, they shift their emphasis from the equations to the concepts and principles. The studies on expert and novice problem solvers by Chi *et al*.[23] (discussed in chapter 2) suggest that these students also begin to classify problems not by their surface features but by the concepts needed to understand the problem and reach a solution. Thus, these students often find solving a few problems in great detail teaches them how to use concepts effectively for problem solving. On the other hand, students who focus more on the equations like to solve many problems so they are better prepared to apply them in many different situations. We would expect that as students become more sophisticated learners, they will disagree with this item.

Item 11. *A good understanding of physics is necessary for me to achieve my career goals. A good grade in this course is not enough.*

Since most of the students surveyed are not physics majors and few non-majors take physics out of interest, most of the students surveyed are taking the introductory physics because it is required by their program. While many of these students just look at physics as another general education requirement, the more sophisticated students see how understanding physics will be useful in their careers. These students will agree with this item. In the pre-course survey, this item can be used to gauge the distribution of majors in a class. The more engineering and physical science majors in given class, the higher the percentage of student responses agreeing with this item.

Item 23. *The main skill I get out of this course is learning how to solve physics problems.*

Item 30. *The main skill I get out of this course is to learn how to reason logically about the physical world.*

Items 23 and 30 are unusual in that both responses are necessary to evaluate student expectations on the course objective. Novice students who emphasize mathematical problem solving will agree more with item 23 than 30. As students become more sophisticated learners, they agree less with item 23 and more with item 30. A comparison of the pre and post course distributions on these two items is an indication of the course goals communicated to the students as well as their expectations.

Item 28. *Spending a lot of time (half an hour or more) working on a problem is a waste of time. If I don't make progress quickly, I'd be better off asking someone who knows more than I do.*

Most physics instructors recognize that having their students struggle with a problem and working through it on their own helps the students piece their knowledge together and build their confidence. Students who recognize the value of this struggle in building their own understanding should disagree with this item.

Item 33. *It is possible to pass this course (get a "C" or better) without understanding physics very well.*

This item tell us more about students' perception of the class rather than the physics expectations of the students. Ideally, understanding physics should be required to pass an introductory physics course. Students who feel this way should disagree with this item. Some false positives ('disagrees") can be expected from students who confuse familiarity with understanding.

Item 34. *Learning physics requires that I substantially rethink, restructure, and reorganize the information that I am given in class and/or in the text.*

In science we make observations, test predictions, and evaluate what we find. Often as we learn more, we need to rethink, restructure, or reorganize what we thought we knew when it seems to contradict what we learn. Reflection is an important tool for building a consistent, coherent knowledge framework. Often students in introductory physics will reach a similar point when they learn something new that either contradicts what they already knew or helps them to think about what they know in different ways. Students who reflect and rethink to add what they learn to what they know should agree with this item. A favorable response on this item suggests a sophisticated learner.

8. Additional survey questions

In addition to the 34 survey items, we also collected data with the survey on the students' background, study habits, and self-appraisal of skills. These background questions ask about the students' major, whether or not they had physics before, and the students' math courses in high school and college. Questions on study habits inquire into how much time students spend studying each week, preparing for exams, and working with others. The students are then asked to rate themselves with regard to understanding course materials, useful skills, and mathematical ability. This background data has not been used for the analysis in this dissertation but will be used in later work.

IV. VALIDITY

Any measurement instrument, particularly an attitude measurement like the MPEX survey, needs to be checked for both validity and reliability. A measurement instrument is validated when one can demonstrate that it measures what was intended. Reliability is an indication of to what extent the instrument measurements are free of unpredictable kinds of error. Reliability is discussed in section VI. In the current section, I discuss the results of two kinds of construct validity procedures. Construct validity determines how well the instrument measures the construct being measured, in this case the students' cognitive expectations. For our purposes, construct validity is important for verifying that the MPEX survey is measuring cognitive attitudes and beliefs about the course and learning physics. However, since we believe that student expectations play a major role in what students learn in an introductory course, it would be useful to see if the MPEX survey results correlate with other measures of student learning such as grades and the FCI. This correlation analysis will be performed in later work.

A. Surface Validity - Measurements of Calibration Groups

In order to test whether the survey correctly represents elements of the hidden curriculum, it was given to a variety of students and physics instructors. The "expert" response was defined as the response that was given by a majority of experienced physics instructors who have a high concern for educational issues and a high sensitivity to students. Redish, Steinberg, and I conjectured that experts, when asked what answers they would want their students to give, would respond consistently.[24]

1. The calibration groups

We tested the response of a wide range of respondents by comparing five groups:

Group 1: engineering students entering the calculus-based physics sequence at the University of Maryland,

Group 2: members of the US International Physics Olympics Team

Group 3: high school teachers attending the two-week Dickinson College Summer Seminar on new approaches in physics education

Group 4: university and college teachers attending the two-week Dickinson College Summer Seminar on new approaches in physics education

Group 5: college faculty who are part of a multi-university FIPSE-sponsored project to implement Workshop Physics at their home institutions.

The University of Maryland students (UMD) are a fairly typical diverse group of primarily engineering students at a large research university. The entering class average on the FCI is 51.1% ± 2.4% (standard error). The number of students in the sample is N = 445.

The US International Physics Olympics Team (POT) is a group of high school students selected from applicants throughout the USA. After a two week training session, five are chosen to represent the US in the International Physics Olympics. In 1995 and 1996, this group trained at the University of Maryland in College Park and we took the opportunity to have them complete survey forms. The total number of respondents in this group is N = 56. Although they are not teachers, they have been selected by experts as some of the best high school physics students in the nation. Our hypothesis was that they would prove to be more expert than the average university physics student, but not as expert as our groups of experienced instructors.

The physics instructors who served as our test groups were all visiting Dickinson College. Attendees came from a wide variety of institutions. Many have had considerable experience in teaching, and all of them were sufficiently interested in educational development to attend a workshop. We separated them into three groups. Group 3 — high school teachers (HS) attending a two-week summer seminar (N = 26), Group 4 — college and university teachers (College) attending the two-week summer seminar (N=56), and Group 5 — college and university teachers implementing Workshop Physics (Expert) in their classroom (N = 19). The teachers in Group 5 were committed to implementing an interactive engagement model of teaching in their classroom. We asked the three groups of instructors to respond with the answer that they would prefer their students to give after instruction. We expected these five groups to show an increasing level of agreement with answers that we preferred.

2. The responses of the calibration groups

The group we expected to be the most sophisticated, the group 5 instructors, agreed strongly as to what were the responses they would like to hear from their students. On all but three items, ~80% or more of this group agreed with a particular position. These three items, numbers 7, 9, and 34, had a strong plurality of agreement, but between $\frac{1}{4}$ and $\frac{1}{3}$ of the respondents chose neutral. We define the preferred response of Group 5 as the expert response. We define a response in agreement with the expert response as "favorable" and a response in disagreement with the expert response as "unfavorable". A list of the favorable responses to the survey items is presented in Table 5-2.

Although the survey itself uses a five point Likert scale (strongly disagree = 1 to strongly agree = 5),[25] we have chosen to group the survey responses into three categories: agree, disagree, and neutral. Someone who responds either "agree" or "strongly agree" for a survey item is considered to agree with that item. Someone who responds either "disagree" or "strongly disagree" on a survey item is considered to disagree with that item. Someone who responds "neutral" for or does not answer a survey item is considered to be neutral with regard to that item. This was done for two reasons. One, it is not clear that the intervals of the Likert scale are the same for every respondent with regard to any particular survey item. For example, one person who agrees with a particular survey item might have the same expectations as another person who strongly agrees with the same item. Because of this, Redish and I feel that the three-point scale does not unduly reduce the resolution of the survey. The second reason is to sharpen the interpretation of the data. Because we are looking for shifts in student expectations, changes from agree to disagree or even agree to neutral are more significant than changes from strongly agree to agree.

Table 5-2. Prevalent responses of our expert group for the MPEX Survey items. Where the respondents did not agree at the >80% level, the item is shown in parentheses and the majority response is shown. The response "A" indicates agree or strongly agree. The response "D" indicates disagree or strongly disagree.

1	D	8	D	15	D	22	D	29	D
2	D	9	(D)	16	D	23	D	30	A
3	A	10	D	17	D	24	D	31	A
4	D	11	A	18	A	25	A	32	A
5	A	12	D	19	D	26	A	33	D
6	A	13	D	20	D	27	D	34	(A)
7	(A)	14	D	21	D	28	D		

Figure 5-1. A-D plot for the calibration groups, average of all survey items. The percentage of respondents agreeing with the majority of expert views (favorable) is plotted against the percentage disagreeing the those views (unfavorable).

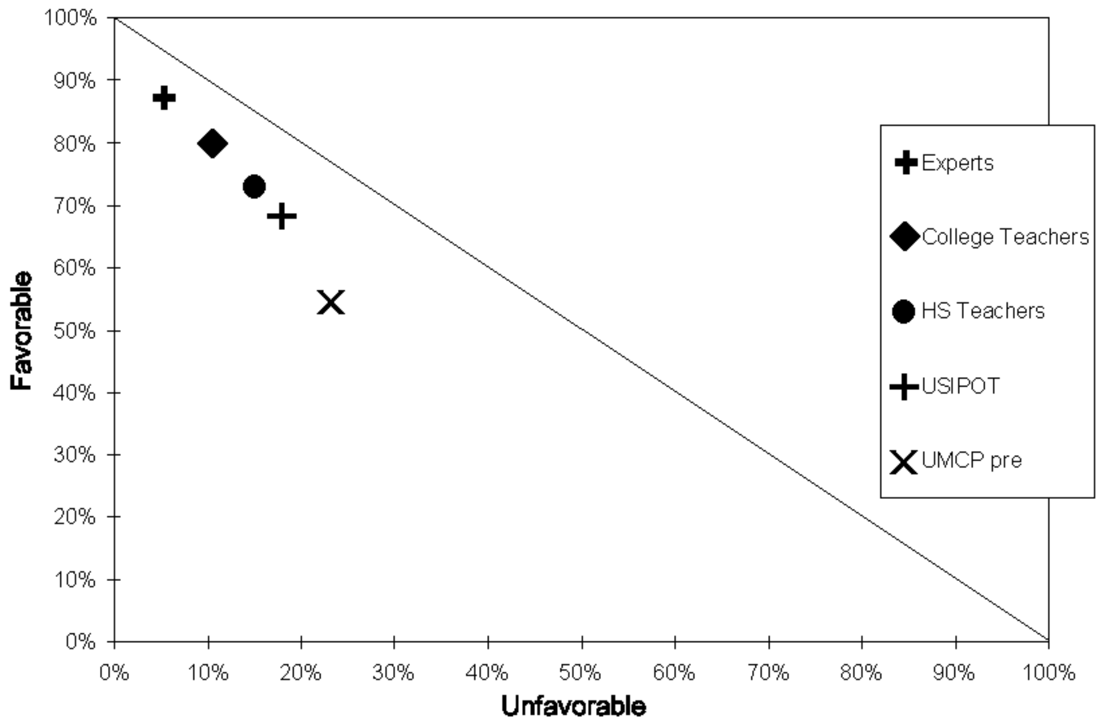


Table 5-3. Percentages of the calibration groups giving favorable / unfavorable responses on Overall and Cluster MPEX survey.

MPEX clusters	Experts	College	HS	POT	UMD Pre
Overall	87 / 6	81 / 10	73 / 15	68 / 18	54 / 23
Independence	93 / 3	80 / 8	75 / 16	81 / 12	47 / 34
Coherence	85 / 12	80 / 12	62 / 26	79 / 8	53 / 24
Concepts	89 / 6	80 / 8	71 / 18	73 / 13	54 / 27
Reality Link	93 / 3	94 / 4	95 / 2	64 / 20	61 / 14
Math Link	92 / 4	84 / 9	67 / 21	85 / 8	67 / 22
Effort	85 / 4	82 / 6	68 / 13	50 / 34	67 / 13

To display our results in a concise and easily interpretable manner, we introduce the *agree-disagree* (A-D) or *Redish plot*. In this plot, the percentage of respondents in each group answering favorably is plotted against the percentage of respondents in each group answering unfavorably. Since the fraction of students agreeing and disagreeing must add up to less than or equal to 100%, all points must lie in the triangle bounded by the corners (0,0), (0,100), (100,0). The distance from the diagonal line is a measure of the number of respondents who answered neutral or chose not to answer. The closer a point is to the upper left corner of the allowed region, the better the group's agreement with the expert response.[26]

The results on the overall survey are shown in Fig. 5-1. In this plot, the percentages are averaged over all of the items of the survey, using the preferred responses of calibration group 5 as favorable. The groups' responses are distributed from less to more favorable in the predicted fashion.[27]

Although the overall results support the contention that our survey correlates well with an overall sophistication of attitudes towards doing physics, the cluster results show some interesting deviations from the monotonic ordering. These deviations are quite sensible and support the use of clusters as well as overall results. In order to save space and simplify the interpretation of results, the data is presented in Table 5-3. Displayed in this table are the percentages of each group's favorable and unfavorable responses (in the form favorable/unfavorable). The percentage of neutrals and those not answering can be obtained by subtracting the sum of the favorable and unfavorable responses from 100.

From the table we see that most of the fraction of respondents agreeing with the favorable response tends to decrease monotonically from group 1-5 with a few interesting exceptions. The high school teachers (group 3) are farther than their average from the favorable corner in the coherence and math clusters, while the Physics Olympics team is closer to the favorable corner in those categories than their average. These results are plausible if we assume that high school teachers are less concerned with their students forming a coherent and a mathematically sophisticated view of physics than are university teachers. The results also agree with Redish and Steinberg's personal observations[28] that the members of the POT are unusually coherent in their views of physics and exceptionally strong in their mathematical skills

Note also that the Olympics team results are very far from the favorable corner in the effort cluster. The main discrepancies are in items 3 and 7. These items represent highly traditional measures of effort (reading the textbook, going over one's lecture notes) which we conjecture are not yet part of the normal repertoire of the best and brightest high school physics students before they enter college. Redish, Steinberg, and I also conjecture that most of them will have to learn to make these kinds of efforts as they progress to increasingly sophisticated materials and the level of challenge rises.

This analysis of both the overall responses of the calibration groups and the variations in the ordering confirms that the MPEX survey provides a quantitative measure of characteristics which experts hope and expect their students to have.

B. Validation with Student Interviews

I conducted more than 120 hours of videotaped student interviews in order to validate that our interpretation of the survey items matched the way they were read and interpreted by students. I asked students (either individually or in groups of two or three) to describe their interpretations of the statements and to indicate why they responded in the way that they did. In addition, students were asked to give specific examples from class to justify their responses. The protocols for these interviews (the MPEX Survey Protocol and the Open MPEX Protocol) are discussed in chapter 7.

From these interviews, we found that students are not always consistent with their responses to what appear to us to be similar questions and situations. We feel that this does not represent a failure of the survey, but properly matches these students' ill-defined understanding of the nature of physics. One reason for this was described by Hammer (see the section of expectations in chapter 2). He observed that some students in his study believed that professional physicists operated under the favorable conditions, but that it sufficed for these students to behave in the unfavorable fashion for the purposes of the introductory course.[29] We refer to this type of characteristic as an "apparent" expectation. This is only one aspect of the complex nature of human cognition. We must also be careful not to assume that a student exists in one extreme state or another. A student's attitude may be modified by an additional attitude, as in Hammer's observations, or even exist simultaneously in both extremes, depending on the situation that triggers the response.[30] This is one reason why considerable care must be taken in applying the results of a limited probe such as our survey to a single student.

We are also aware that students' self-reported perceptions may not match the way they actually behave. However, the interviews suggest that if a student's self-perception of the learning characteristics described in Table 5-1 differs from the way that student actually functions, the self-perception has a strong tendency to be closer to the side chosen by experts. We therefore feel that while survey results for an individual student may be misleading, survey results of an entire classroom might understate unfavorable student characteristics.

Because of the length of this dissertation, I have included in Appendix F selected interview responses for four survey items from nine of the nearly 100 students I have interviewed. The responses represent all the interviews conducted during a site visit at Nebraska Wesleyan University (NWU) two weeks before the end of the second semester. I have selected this set for two reasons. First, these were the only MPEX interviews conducted with a good sample from a sizable calculus-based class that used the most current version (version 4.0) of the survey. Second, two additional controls were added to address concerns raised in the earlier interviews.

- To prevent the interview itself from unknowingly influencing the student responses, the students completed the survey before starting the interview (within 24 hours of the interview).
- This student sample was representative of the entire class. At least three students were interviewed from the top third, middle third, and bottom third of the class. The students were rated by the instructor based on their overall grade at the time of the interviews.

The class was taught with the Workshop Physics curriculum. A description of the curriculum and details of the implementation at NWU can be found in chapter 8.

Below, I discuss student responses from the four MPEX survey items with regard to how students are interpreting the survey items and whether their answers make sense. (A detailed analysis of the interviews in terms of the expectations of these students for evaluation of student learning and the curriculum is presented in chapter 10.) I selected items 2, 6, 14, and 22 because they are a good cross section of both the expectation issues and the difficulties in interpreting the survey data. (Note that all interviewed student names used in this dissertation are code names.)

Items 2 and 6 are two of the items that look at students' perception of derivations. Item 2 asks students if derivations are good for anything beyond demonstrating the validity of an equation. Item 6 asks the students if they do or go over derivations themselves.

All nine students' responses indicated they interpreted item 2 correctly. For example:

> John: *I put "disagree* (favorable response)*." I'm not a lover of derivations, by any means; but I think that it can tell you a lot about linking two concepts together. It also creates the idea that science can be a unified thing, and that previous information is applicable to new information. And so, therefore, I think that it is — it's useful.*
>
> Leb: *I said disagree,* (favorable response) *because, once again, I think it helps to know where the formula came from and then to — helps you to know why it works and why you can apply it that way.*

However, some of the students' responses are illustrative of the complexities involved in student expectations. For example, two of the student responses were very clearly indicative of an apparent expectation, although both students gave an unfavorable response to item 2.

> Amy: *I do agree with that* (unfavorable response). *Okay*. … [Can you elaborate a little bit on that? Why you agree with that statement.] *When you get into the beginning, you're introduced some kind of equation. See, Dr. A won't give you the equation until he derives it. ... But if — You know, if he got up there and was — I don't know. I guess I probably would be negative towards it. You can use that proof, no matter what, anyway; because once you get experience with any kind of proof at all, you know you can plug and chug. And I mean that's fact. But what I get out of watching someone derive a derivation or a proof is it gets me closer to being that type of person who'll be able to do it, myself. I'm not, by nature, somebody who does that. I think Dr. A is and always has been.*

Krystal: *I said I agree* (unfavorable response)*, because I'm poor in mathematics, but not necessarily for everybody. I mean, those students — or a lot of students, I'm sure, would understand what they got and know where everything came from. But, again, I just — since I haven't had much math, usually, when he does all that on the chalkboard or on the overhead screen, it's Greek to me. I don't know where he got it from. I just write the final thing down and, you know, that's what I need.*

Amy recognizes that watching the professor model derivations is useful, but she doesn't see derivations as useful in this class or see herself as being able to do them. Krystal, sees derivations as useful for people who understand the mathematics but not her.

One thing that comes out clearly from the favorable responses is that the derivations are helping some students see the relationship between the concepts and the formulas. This is clearly seen in John's response above.

All but one of the eight student responses to item 6 are consistent with the intended interpretation. For example:

Charlie: *And I agreed with this* (favorable response)*. Whenever I -- If there's a formula presented in the text, or Dr. Wehrbein presents one on the board, I always try and understand where it came from. And if I don't right offhand -- like, if he writes something and goes through it too fast, I'll try and take all the notes I can. And later, I'll go over it again and see if I can make it all make sense in my head. And it's only at that point that I feel like I understand it.*

Krystal: *I said three, "neutral." Somewhat, again, because I'm a poor math student, so -- and it's also a phobia. Because when I see a derivation, I usually just turn the page and don't try to thumb through it.* [Okay. So, that tells me why you didn't agree with it. Why didn't you disagree with it?] *Well, because I also … I mean I don't always just flip the page, because I don't give up that easily. I usually will try and get partially through it. But, then, if I get lost, I usually give up then.*

John: *"Disagree* (unfavorable response)*." I don't spend too much time trying to figure them out. If you do it once, to show you where the formula came from. I know where it came from. I understand the concepts linking the two together. I usually don't try to go through it mathematically, because then it becomes more of a math exercise.*

It is interesting to note that while 5 of the nine students think derivations are generally useful for learning physics, three of these five students including John do not regularly work out derivations on their own. Only Kim's interpretation of item 6 is questionable since she spends a lot of time going over derivations with the instructor outside of class. In addition, Kim's response may indicate a difficulty that appears regularly in a small fraction of the students interviewed, usually in an interview near the beginning of the sequence. These students seem to confuse derivations with derivatives. Also note that although Krystal does do derivations but she does not spend a lot of time on them which is consistent with a neutral response.

Item 14 deals directly with the issue of what does it mean to learn physics. As we saw earlier, item 14 looks like a statement that students should agree with until you think about what you really want students to learn. Students who disagree with this item recognize their own role in building an understanding of physics that makes sense to them. For example,

> Hannah: *I think I should disagree with this* (favorable response), *because it's more than just acquiring knowledge. You also have to put it all together*. *It just can't be a bunch of facts*

Most students who agree with this item see learning physics as memorizing and applying the facts and formulas from lecture and the textbook. All nine students were asked this question. Of the nine, only Krystal and Ramsey misinterpreted the question although Ramsey changed his mind when asked to reconsider.

Both Krystal and Ramsey initially use the laboratory component of the course in their response and disagreed with item 14. For example, Hannah responded:

> *I said I disagree, because you need to be able to set up the experiment, too, which is something that isn't really given to you anywhere. Like -- Like you said, that was something that I had to learn first of all, where some people didn't. I don't know. Some people who really had a great high school curriculum, or just this natural instinct at how to set up circuits; but I didn't know how. So first I had to be taught that. Then I could continue to learn.*

The fact that both Ramsey and Krystal mention the laboratory as an important part of learning in the courses is not unusual since the students in a Workshop Physics class spend at least three quarters of their time on lab activities. Even so, this interpretation was unusual and not seen in other interviews, even at the other two schools using the Workshop Physics curriculum.

John and Leb gave neutral responses to item 14 which makes sense in light of their explanations which show that they have mixed expectations. This is particularly clear in John's response,

> John: *Three, "neutral." Again, I think it depends on the type of person you are. Learning physics as a whole – really learning physics – is more than just ... principles, equations, laws. It's ... in understanding and incorporating all this. But to pass a course, learning physics, I think that's all you really need to get by.*

John sees that there is more to learning physics than just the laws, principles and equations but he also sees that this view is not essential in this class. Despite the obviousness of item 14, almost all students who disagreed with this item in interviews had strong constructivist expectations. However, some constructivist students who express mixed expectations do not disagree with this statement like Leb and John.

The student responses to item 22 at Nebraska Wesleyan were unusually favorable. (The MPEX reality link survey responses overall for Nebraska Wesleyan were also unusually favorable. Please see the discussion in the workshop physics site visits section of chapter 10 to see why.) All the student responses indicate that the students are interpreting item 22 correctly. For example,

> Charlie: *And I disagree with this one* (favorable response). *I think any time you can relate it to the real world – the world around you, you can understand it a little bit better, because it familiarizes it with yourself a little bit more. And it seems to apply a little bit more, becomes a little more important.*

Several of the students like Ramsey emphasized the role of laboratory activities in connecting physics to the real world.

> Ramsey: *And I marked "disagree" for the reasons I stated before* (favorable response) – *that we're exposed to experiments or demonstrations that you can find in the real world. So I think that learning about those things are essential for what we're doing.*

Only John's response was questionable. His response is almost a false unfavorable since he sees the connection as beneficial, though not essential. However, this response and its categorization is consistent with his overall expectation profile discussed in chapter 10.

The interview results demonstrate that while there are occasional misinterpretations of survey items, the effect is small and tends to overstate the number of favorable student expectation responses.

VI. UNCERTAINTY & THE STATISTICS OF SHIFTS

Every finite set of data contains fluctuations, which have no real significance but arise from the variability of a particular sample. In this dissertation, my research questions involve comparisons of groups – experts and novices, novice students at different institutions, and students at the beginning and end of their first semester of physics. In order to compare these groups, we compare their averaged responses (agree vs. neutral vs. disagree). In order for us to understand whether two responses are significantly different, we have to have some model of the random variable in our sample.

Our interviews, our intuitions, and many discussions in the cognitive literature suggest that a human attitude is a highly complex object. As we noted above, some students gave clear evidence in interviews of being in two contradictory states at the same time. What this implies is that the random variable we should be averaging is itself a probability, rather than a set of well-defined values. Unfortunately, the average of probabilities may depend significantly on the structure of the constraints and parameterization of the probabilities, as is well know from quantum statistics. Since detailed models of student attitudes do not yet exist, we will estimate our shift significance by using a cruder model.

Let us assume that a class is drawn from a very large homogeneous[31] group of students and that in the large population, a percentage p_0 of responses to an item or cluster will be favorable and a percentage q_0 will be unfavorable with $p_0 + q_0 \approx 1$. (For now, we will ignore the possibility of neutral responses.[32]) In a finite sample of n students, we want to know what is the probability of finding n_1 favorable and n_2 unfavorable responses with $n_1 + n_2 \approx 1$. Using the Gaussian approximation to the binomial distribution, we get that the probability of finding fractions

$p = n_1/n$ and $q = n_2/n$ is $P(p) = Ae^{-(p-p_0)^2/2s^2}$

where A is a normalization constant and the standard deviation

$$s = \sqrt{\frac{p_0 q_0}{n}}$$

For this distribution, the probability that a sample will have a mean that falls within 1σ of the true mean, p_0, is 0.684 and the probability that a sample will fall within 2σ of the true mean is 0.954.

Since the fraction of neutral responses tends to be small, and since the binomial model is crude for this set of data, we treat our trinomial data as if it were approximately binomial by renormalizing the observed p and q into $p' = p/_{p+q}$ and $q' = q/_{p+q}$. We consider a difference or shift in means to be significant if it is at less than the 5% probability level, that is, if the difference or shift is greater than twice σ (where $s = \sqrt{\frac{p'q'}{n}}$). Because of the crudeness of this model, we consider differences of 2σ to be significant. For example, at values of p = 60%, q = 20% for N = 450, we get σ ~ 2%. We would therefore consider a 4% shift to be significant for the University of Maryland engineering students. For N = 50, those values of p and q give σ ~ 6%. We therefore consider a 12% shift to be significant for the Physics Olympics team and the college instructors attending the Dickinson Summer Seminar. Note that for a given sample size, σ doesn't change much over the typical values of p and q seen in Tables 10-3 and 10-4.

V. RELIABILITY

Reliability tests are measures of the random errors inherent in instrument measurement. In this section, I will describe the results of three types of reliability measurements: factor analysis, Cronbach alpha, and reproducibility of results. Since we have created clusters that are subsets of overall student expectations, I have used factor analysis to see if these clusters correspond to structures reflected in the student data.

The Cronbach alpha coefficient is a standard test of internal consistency, commonly used for Likert-scale attitude surveys.[33]

As any physics instructor preaches, the ultimate mark of a good measurement is that it is reproducible. In this case, this means reproducibility of the measurements of a class rather than the measurements of individual students. The reader is reminded that the MPEX survey was not designed to be used a diagnostic instrument for individual students, but rather to measure the distribution of student expectations in a class. However, reproducing the MPEX survey results from the same class sample twice presents at least two major difficulties. The first difficulty is the logistics in giving the survey twice to a class over a period of time long enough that the students do not remember how they responded before, but not so long that the student attitudes have changed. The second difficulty is that many students have mixed expectations, i.e. hold contradictory expectations simultaneously. These students can be thought of as existing in an expectation superposition state similar to the model of fragmented students' knowledge structures discussed in chapter 2. This implies that the measured expectation response can change is some circumstances. Unfortunately there are not enough survey items to accurately discriminate among various mixed expectation states.

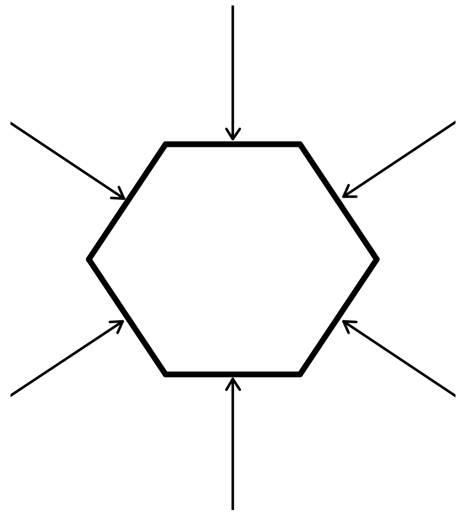
Instead of trying to compare two measurements from the same set of students, I compared results from pre-course surveys of multiple classes of the same course from the same school assuming that the incoming students' expectations do not change significantly from one year to the next. The survey results from a particular course at a particular school would then be considered reproducible if the distribution of the means of the survey results were consistent with the standard deviation σ given by the binomial probability theory in the previous section.

A. Factors & Clusters

Our seven MPEX clusters were written to make sense of students' epistemology and learning beliefs in way that could be used by an instructor to monitor different aspects of the hidden curriculum in the introductory physics course. The clusters are based on the previous work of Hammer and others as well as our own observations. While Hammer found that students' expectations could be categorized by a researcher, he did not believe that his categories necessarily represented how expectations were structured in the students' mind. Because so little is known about the cognitive expectations of undergraduate students in introductory calculus based-physics classes, there is little reason to expect that our seven dimensions represent dimensions of expectations in the students' minds.

The situation is analogous to the strain on a cubic crystal.[34] Consider the two dimensional case of a hexagonal piece of crystal as shown in Figure 5-2 below. Suppose 3 clamps are placed on the crystal so that external forces are exerted on all six sides. Although the internal structure of the crystal is easily described using two basis vectors, the external forces are most easily described in terms of those being exerted on the three pairs of parallel sides. The problem is most easily considered as two coordinate axes or three coordinate axes depending on your perspective and what

Figure 5-2. Two-dimensional view of hexagonal piece of cubic crystal.



you are trying to do. If I want to study how the overall crystal reacts to external forces, then I would want to use the (overcomplete) coordinate system that corresponds to the symmetry axes of the hexagonal shape of the crystal. If I want to study how individual crystal cells react to forces, then I would want to use a coordinate system based on the crystal structure.

Because so little is known about student expectations, the situation is like looking at the hexagonal block of cubic crystal but the internal structure is unknown. In making our measurements, we define our coordinate system or clusters along dimensions we can observe easily and which are easily related to external influences. In the crystal analogy, this coordinate system also makes it easier to relate the changes in the measurement of the crystal to the external forces being applied.

In a similar way, the MPEX clusters are easily related to various aspects and goals of instruction such as emphasis on linking physics to everyday life or on the coherence of physics. Like the external forces on the crystal, the seven MPEX expectation dimensions are interrelated. In addition, we can try to diagonalize the matrix obtained from our external measurements to learn more about the internal structure, i.e. how student expectations are structured in the student's mind.

One way to determine how student expectations are structured is to use factor analysis to look for underlying structure in the data. Factor analysis is a statistical method of using correlations to reduce a large number of variables to a smaller number of factors that might help to explain the data more easily. The procedure is similar to diagonalizing a matrix and finding the eigenvalues and eigenvectors.

However, to properly interpret the results of a factor analysis one must be aware of the assumptions and limitations of this type of analysis. Factor analysis assumes that the test items in question are linearly related to a set of uncorrelated, i.e. orthogonal and independent, factors.[35] While factor analysis is a powerful technique for looking for interrelationships in a data set, the resulting factors are purely mathematical constructs that may not have any real meaning if the linear relationship described above does not exist. Also, as the linear relationship may not be unique, one must have additional justification of the factors before trusting the results of a factor analysis.

It is common practice in the constructions of tests and surveys that use sub-scales like our clusters to use factor analysis to see if these sub-scales are reflected in the data. For a factor analysis of the MPEX survey data, I used my largest single sample of data, pre-course surveys from the University of Minnesota calculus-based introductory physics sequence. A course description can be found in chapter 9. I performed a principal-component analysis with varimax rotation using the SPSS statistics application.[36] A plot of the eigenvalues vs. number of factors shown in Figure 5-3 indicates that perhaps three, four, or five factors account for significant fractions of the variance. As can be seen in table 5-4, the percentage of variance accounted for by the fourth factor is marginal. Note that 3 factors would account for 29% of the total variance; four would account for 33%. For a typical survey instrument, the factors usually account for roughly 30% of the variance.[37]

An analysis was run on the data for three, four, and five factors. The full results are shown in Appendix H. A summary of the results is shown in Table 5-5. Note that

Figure 5-3. Scree plot: the # of Eigenvectors vs. the # of factors. The Scree plot is used to estimate how many factors should be extracted. The factor number at the knee should be within ±1 of the correct number of factors. This graph suggests there are 4 ± 1 factors.

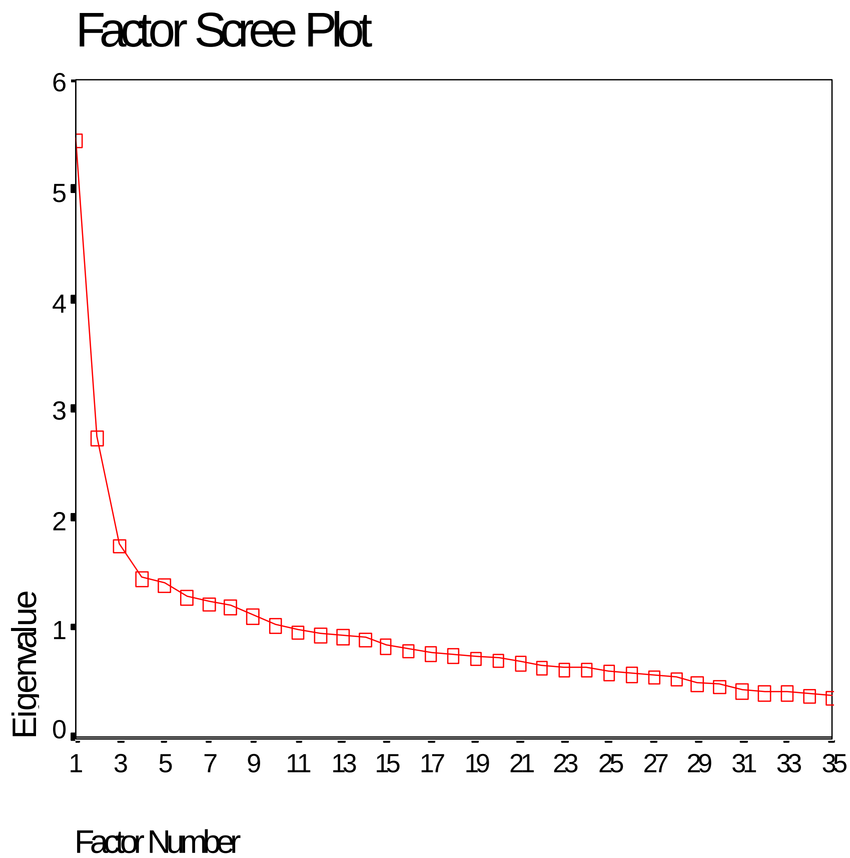


Table 5-4. Percentage of variance associated with the extracted factors. Ideally one wants to maximize the variance with the fewest factors.

Factor	Eigenvalue	Pct. of Variance	Cumulative Percentage
1	5.46	16.1	16.1
2	2.75	8.1	24.1
3	1.70	5.0	29.2
4	1.45	4.3	33.4
5	1.33	3.9	37.3
6	1.26	3.6	41.0

Table 5-5. Results of factor analysis for three-, four-, and five-factor extraction.[38]

3 factor extraction		4 factor extraction		5 factor extraction	
MPEX items	Factor loading	MPEX items	Factor loading	MPEX items	Factor loading
25	-0.66	25	0.71	8	0.62
18	-0.59	18	0.65	16	0.60
17	0.58	10	-0.60	20	0.59
10	0.57	30	0.58	2	0.58
20	0.52	31	0.47	15	0.52
8	0.52	22	-0.42	17	0.51
30	-0.51	11	0.38	21	0.42
31	-0.51	5	0.36	28	0.39
15	0.47	24	-0.29	29	0.37
22	0.47	26	0.27	—	—
29	0.40	—	—	25	0.74
11	-0.40	8	0.61	18	0.65
5	-0.36	16	0.60	10	-0.59
16	0.35	20	0.58	30	0.56
1	-0.35	2	0.57	22	-0.41
26	-0.34	15	0.52	11	0.39
24	0.29	17	0.49	5	0.37
—	—	21	0.42	26	0.28
19	0.64	9	0.38	—	—
4	0.56	28	0.38	4	0.62
14	0.52	29	0.36	14	0.61
27	0.52	—	—	19	0.58
23	0.52	4	0.63	23	0.56
2	0.51	14	0.61	27	0.47
13	0.45	19	0.58	1	0.42
12	0.40	23	0.56	13	0.39
28	0.35	27	0.49	—	—
21	0.31	1	0.40	6	0.74
9	0.24	13	0.39	7	0.56
—	—	—	—	34	0.47
6	0.75	6	0.75	3	0.47
7	0.59	7	0.58	12	0.41
34	0.49	34	0.49	33	-0.31
3	0.47	3	0.46	—	—
32	0.44	32	0.44	32	0.52
33	-0.28	12	0.38	9	0.50
		33	-0.27	31	0.42
				24	-0.34

four of the factors in the results from four factor extraction and five factor extraction have substantial overlap. Phenomenographic[39] examination of the survey items corresponding to the four-factor case by Wittmann,[40] Redish, and the author found themes in the first four factors:

1. The items in this factor are all related to student beliefs about building an intuitive conceptual understanding to learn physics.
2. The items in this factor all deal with how students think about the material they learn in introductory physics.
3. The items in this factor concern whether physics is mainly problem solving or building understanding.
4. The items in this factor ask about what students do to learn physics.

Not surprisingly, when I compared the factors and the clusters, I found that the four factors and the seven clusters do not correspond exactly. A comparison of the factors and the clusters is shown in Table 5-6. Note that in the four-factor table, several items from each cluster correspond to a single factor. The split in the effort cluster represents differences in students' perceptions. This split suggests that students see going over notes, derivations, and the text as different from going over graded homework and exams.

The fact that the structures of the factors and clusters do not correspond exactly does not invalidate either. Going back to the example of pairs of vertical forces applied to the cut crystal. The clusters correspond to the force pairs; the factors correspond to the crystal's internal structure. The two may be related, but they are two different ways of looking at the same situation. And while it is reasonably straightforward to design curriculum to deal with the students difficulties indicated by the clusters, designing curriculum in terms of the factors would be more difficult. Also, as with the case of the

Table 5-6. Comparison of factors and clusters

4 Factor extraction:

Clusters	Factor 1	Factor 2	Factor 3	Factor 4	No Factor
Independence	1,13,14,27	8,17	—	—	—
Coherence	—	15,16,21,29	—	—	12
Concepts	4,19,27	—	26	32	—
Reality Link	—	—	10,18,22,25	—	—
Math Link	—	2,8,16,20	—	6	—
Effort	—	—	24,31	3,6,7	—
No Cluster	23	9,28	5,11,30	33,34	—

3 Factor extraction:

Clusters	Factor 1	Factor 2	Factor 3	No Factor
Independence	13,14,27	1,8,17	—	—
Coherence	12,21	15,16,29	—	—
Concepts	4,19,27	26	32	—
Reality Link	—	10,18,22,25	—	—
Math Link	2	8,16,20	6	—
Effort	—	24,31	3,6,7	—
No Cluster	9,23,28	5,11,30	33,34	—

FCI factor analysis discussed in the last chapter, dimensions of expectations are not completely independent. They are very much interrelated. Furthermore, as discussed earlier, just as many students can hold two contradictory conceptual ideas in their minds simultaneously, they can also hold mixed expectations with regards to the physics class.

This analysis of the structure of students' cognitive expectations and their relation to the survey and the clusters is suggestive, but at this point our analysis is still preliminary.

B. Cronbach Alpha

One estimate of the reliability of a diagnostic or survey test instrument is to look at the internal consistency of the items. As mentioned earlier, the Cronbach alpha coefficient is a standard method for estimating the internal consistency of a test where the items are scored across a range of values like a Likert scale.[41] It measures how well the responses to the items in a test or a subset of questions in a test (items in a cluster or factor for example) contribute to measurements of a single construct or issue like students' understanding of Newtonian force, students general attitudes towards science, or students' epistemological beliefs with regards to science. The Cronbach alpha coefficient is commonly used to estimate the reliability of both the total test score and/or the scores from subsets of test items like the clusters or factors discussed above. Tests or subtests having an alpha coefficient ≥ 0.70 are considered to be reliable for group measurements.[42]

The coefficient is also used to determine if it is more appropriate to evaluate test measurements in terms of the overall test score or the subset scores.[43] If the alpha coefficient for the overall score is smaller than the coefficients for the subtests, the test items within the subtests correlate more strongly with the subtests than with the overall score. This is an indication that the test is measuring more than one variable or issue. In this case it is not appropriate to construct an overall score and the test results should only be discussed in terms of the subtest scores.

The Cronbach alpha coefficient is based on the idea that the responses to a group of questions addressing a single issue should correlate strongly with one another. The formula for calculating the Cronbach alpha coefficient[44] is

$$a = \left(\frac{k}{k-1}\right)\left(1 - \frac{\Sigma\, s_i^{\,2}}{s_t^{\,2}}\right) \quad \text{where}$$

k = the number of test items, Σ = the sum over all test items,

$s_t^{\,2}$ = the variance[45] of the total test scores, and

$s_i^{\,2}$ = the variance of the ith item on the test.

If the responses to the test items are substantially intercorrelated, then $s_t^{\,2}$ will be considerably larger than when they are not intercorrelated. However, if the responses to the items are uncorrelated, the variance for each item is independent. This means the summed variance term is unaffected by the correlation of the test item responses. Thus alpha approaches zero when the item responses are not intercorrelated and becomes larger as the item responses become more and more correlated. If the standard deviation of the test items is one, the Cronbach alpha reflects the average inter-item correlation, i.e. how well the test data items inter-correlate with one another and the overall score. If the standard deviation of the items is not one, then the Cronbach alpha reflects the average inter-item covariance.[46] Note that since the average of the standard deviations of the MPEX survey items for the data set used to calculate the Cronbach alpha coefficient is 0.93 ± 0.10, the result can be interpreted as the inter-item correlation.

Note that negative inter-item correlations can yield a reduced or negative alpha coefficient. So, data from surveys like the MPEX that have a mixture of negative and positive items (items where people with a particular expectation should disagree in some cases and agree in others) need to be adjusted so that all the inter-item correlations are positive. In this case, the MPEX data used in the factor analysis described above was transformed so that 5 was the extreme favorable response and 1 was the extreme unfavorable response for all items. Items where the extreme favorable response had been 1 were transformed by the equation $x' = 6 - x$. The full five point Likert scale was used for this analysis. The SPSS program was used to calculate the alpha coefficients for the overall survey score, the clusters, and the factors. The results are shown in Table 5-7. In some cases, the alpha coefficient for a group of items could be increased by removing some items from the grouping in question. The maximum obtainable alpha's from the selected items are also shown in Table 5-7.

The table shows that while the clusters and factors have coefficients ranging from 0.47 to 0.78, the overall survey has an alpha coefficient of 0.81. This indicates that the overall survey score is estimated to be a more reliable measurement of expectations than the clusters or factors. Since its alpha coefficient is > 0.70, the overall MPEX score is therefore a useful and reliable measure. (Item groupings with alpha coefficients of ≥ 0.70 are considered to be well intercorrelated and reliable.)

Table 5-7. Cronbach Alpha for the overall, cluster, and factor MPEX results

i.) Cronbach Alpha for overall MPEX survey result = 0.806

Number of students = 417 Number of survey items = 34
Fall 1995 University of Minnesota pre course data

ii.) Cronbach Alpha for the n = 3 and n = 4 factors using the same data

3 Factor Extraction

Factor	Items	Alpha
1	1, 5, 8, 10, 11, 15, 16, 17, 18, 20, 22, 24, 25, 26, 29, 30, & 31	0.78
2	2, 4, 9, 12, 13, 14, 19, 21, 23, 27, & 28	0.69
3	3, 6, 7, 32, 33, & 34	0.54

4 Factor Extraction

Factor	Items	Alpha
1	5, 10, 11, 18, 22, 24, 25, 26, 30, & 31	0.71
2	2, 8, 9, 15, 16, 17, 20, 21, 28, & 29	0.72
3	1, 4, 13, 14, 19, 23, & 27	0.65
4	3, 6, 7, 32, 33, & 34	0.54

iii.) Cronbach Alpha for the MPEX clusters using the same data

Cluster	Items	Alpha		Alpha
Independence	1, 8, 13, 14, 17, 27	0.48	8, 13, 14, 17, 27	0.57
Coherence	12, 15, 16, 21, 29	0.49		
Concepts	4, 19, 26, 27, 32	0.49	4, 19, 27	0.58
Reality Link	10, 18, 22, 25	0.67		
Math Link	2, 6, 8, 16, 20	0.66		
Effort	3, 6, 7, 24, 31	0.47	3, 6, 7, 31	0.56

Still, this does not mean that MPEX clusters with alpha coefficient < 0.70 are unreliable. The Cronbach alpha analysis assumes that all the items in a grouping are measuring the same construct in a similar way and therefore all items should correlate. Unfortunately, this is not consistent with the overall survey design, particularly in the clusters. The survey was designed so that as a student's expectations increase in sophistication, the student will give more favorable responses for the items in a given cluster. The responses of students in a mixed state for some items in a given cluster are expected to negatively correlate. For example, some items are designed so that only students with extremely favorable expectations will respond favorably to these items. Because there are only a few items like this in the survey, the effect of these items on the overall MPEX result is small. The effect on the clusters is greater because of the smaller number of items in the clusters.

Also, as we discussed earlier, the clusters were intended to give instructors a way of interpreting the results in terms of student response to instruction. That does not mean that the clusters represent the way the students associate the issues associated with the items. For example, while student response to the items in the effort cluster that concern studying for the class are correlated, the items regarding learning from mistakes on exams and homeworks do not correlate with these items. These items correlate better with items dealing with how students construct their understanding.

However, the construction of the factors is based on the intercorrelations of the test items. Therefore, factors like factors 4 and 5 with alpha coefficients < 0.65 should not be considered reliable and should not be used for interpreting test results.

C. Repeatability

For any experimental measurement in physics, a measurement is not considered reliable unless it is reproducible. Since the MPEX survey was designed to measure the distribution of student expectation in a class, one method for evaluating repeatability is to give the survey to a class twice over a period of a few weeks. While the data I took at the end of one quarter and the beginning of the next quarter could be used for this purpose, there is a problem. Between the two measurements, the students study and prepare for the final exam. This can have a strong effect on student expectations. In addition, we have noticed an another effect that makes a pre/post comparison of this kind questionable. First, students tend to respond more favorably to the MPEX survey items at the beginning of the second course in the sequence than at the end of first. This effect is particularly noticeable in the effort cluster. This is an example of students' responding according to what they think and hope they will do compared to their evaluation of what they have really done.

A better test of repeatability is to compare MPEX results of the initial state of the students coming into the introductory physics sequence from several classes for the same sequence at the same school over a two to three year period of time. In this case, the student responses will be classified as favorable, unfavorable, or neutral. Recall that a favorable response is one that agrees with a majority of our experts.

For this comparison I wanted a data set that met the following conditions so that the class populations would be as similar as possible:

1. At least six classes of one course from one school were surveyed. (The sample needs to include enough classes to get a reasonable estimate of the distribution of mean responses.)

2. The classes are all from the main sequence over a period of two years and none are evening extension classes. (Off sequence classes often have varying mixtures of students who are taking the course early, students who delayed, and students who are repeating the course. Thus it is more difficult to say that any given class taught off-sequence is typical or has a population similar to a main sequence course. Evening extension classes often have populations that are significantly different from normal day classes. These are often older, returning students with different views of learning and motivation.)
3. The same version of the MPEX survey was used in all the classes (This ensures that the wording of all the items was the same for all classes and that none of differences are due to the change over from a pencil and paper survey to the scantron version.)
4. The average class size is a large enough sample to account for local fluctuations (The relatively small number of classes surveyed, N < 13, does not adequately account for the differences due to small sample size, i.e. classes with fewer than 50 students. In this study I have observed large significant fluctuations in both the initial MPEX results on some items and the initial FCI scores between physics classes taught in the same semester in the same school when average class size was < 50 students.)

Only data from four of the schools participating in the study met the first condition of pre-course data from at least 6 classes being given the MPEX survey at the beginning of the introductory physics sequence: University of Maryland (9 classes), University of Minnesota (10 classes), Dickinson College (6 classes), and Nebraska Wesleyan University (6 classes). Descriptions of the courses and schools can be found in chapter 8. However, of these four schools, only the data from University of Minnesota met all four conditions. University of Maryland and Dickinson College failed to meet conditions two and three. Both Dickinson College and Nebraska Wesleyan University have small classes and failed condition four. In addition, the six classes at Dickinson were surveyed over three years using versions 3.0 and 3.5 of the survey. Although versions 3.0 through 4.0 of the MPEX survey are very similar, there are some small wording changes on some items. At University of Maryland, classes from on and off sequence were surveyed with MPEX versions 3.0, 3.2, and 3.5. The data from all of these schools were analyzed to see how weakening the conditions would affect the repeatability of the measurements.

The MPEX item results for these four schools were translated to a binomial distribution by renormalizing the observed p and q into $p' = p/(p+q)$ and $q' = q/(p+q)$ where p is the observed percentage of favorable responses and q is the observed percentage of unfavorable responses. The overall MPEX results and the cluster results were recalculated from p' and q'. Then the measured averages and standard deviations for the individual survey items, the overall survey, and the seven clusters for each school were calculated from the normalized results for each class from that school. An estimated standard deviation σ' calculated for a Gaussian distribution from binomial probability theory was calculated for each item, each cluster, and the overall survey. The overall survey and cluster results for the four schools are shown in Table 5-8 with both the measured σ and the estimated σ'.

The results from the University of Minnesota data for the survey items, the overall survey, and the clusters show that the distribution of actual results in all three cases is consistent with the distribution spread estimated using the binomial Gaussian distribution. The results for Nebraska Wesleyan University and Dickinson College for the overall survey and the clusters are also consistent with the distribution spread predicted using the binomial Gaussian distribution but for some of the individual items $\sigma' < \sigma \leq 2\sigma'$. The measured overall and cluster results for University of Maryland from courses both on and off sequence using up to three different versions of some survey.

questions were roughly consistent with the predicted distribution spread. While the spread in some of the items is consistent with the predicted spread, on six of the survey items $\sigma > 2\sigma'$.

These results indicate that if the four conditions for comparing classes with similar populations are met, the distribution of all normalized survey results are consistent with the distribution spread estimated using binomial probability theory. If the class sizes are small, there may be small differences in the spread of the distribution for some of the survey items, but not in the distribution of the overall survey and cluster results. The classes from University of Maryland, where the distribution included data from both on and off sequence classes and results from versions 3.0-3.5 of the survey, had larger differences on some of the survey items, but the distribution of overall survey and cluster results were roughly consistent with the predicted standard deviation. Thus, the MPEX survey results of classes at the beginning of the introductory physics sequence are reproducible and reliable. The overall and cluster results are more robust than the results of individual survey items.

VII. SUMMARY

We saw in chapter 2 that students' expectations can have significant influence on what they learn in an introductory physics class. To better understand the role of student expectations in introductory physics, we developed the MPEX survey to study the distribution and evolution of student expectations through the introductory sequence. The MPEX survey items look at many expectation issues that affect student learning, but mainly focus on the following six dimensions described in more detail in Table 5-1:

independence, coherence, concepts, reality link, math link, and effort. The MPEX survey can determine the pre-course distribution of student expectations in a class both overall and for each of the six dimensions listed above. Measurements taken later in the sequence can show how the distribution changes during the sequence.

The MPEX survey was tested for both validity and reliability. Validity was demonstrated by studying the results of calibration groups and over 120 hours of student interviews. The results of these two tests indicate that both faculty and students agree with our interpretation of the MPEX survey items and what we consider to be an expert response. Although the interview results indicated that a few students occasionally misinterpreted some items, the interview study showed that the misinterpretation error is small and generally tends to overstate the favorable nature of the measured expectations. The survey was tested for reliability by comparing pre-course results for similar main sequence classes at a particular school. The measured distribution was comparable or less than the estimated distribution width for the overall survey results and each of the seven dimensions listed above.

---

[1] D. Hestenes, M. Wells, and G. Swackhamer, "Force Concept Inventory," *Phys. Teach*. **30** (3), 141-158 (1992).

[2] R.K. Thorton and D.R. Sokoloff, "Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the evaluation of active learning laboratory and lecture curricula," *Am. J. Phys*. **66** (4), 338-351 (1998).

[3] E. Mazur, *Peer Instruction A Users Manual* (Prentice Hall, New Jersey 1997).

[4] D. Hammer, "Two approaches to learning physics," *Phys. Teacher* **27** (9) 664-670 (1989).

[5] S. Tobias, *They're Not Dumb, They're Different: Stalking the Second Tier* (Research Corporation, Tucson AZ, 1990).

[6] D. M. Hammer, *Defying Common Sense: Epistemological Beliefs in an Introductory Physics Course*, Ph.D. Dissertation, University of California at Berkeley, 1991 (unpublished).

[7] E.F. Redish, J.M. Saul, and R.N. Steinberg, "Student Expectations in Introductory Physics," *Am. J. Phys*. **66** (3), 212-224 (1998).

[8] I. Halloun, "Views about science and physics achievement: The VASS story," in *AIP Conference Proceedings No. 399 The Changing Role of Physics Departments in Modern Universities: Proceedings of the International Conference on Undergraduate Physics Education*, edited by E.F. Redish and J.S. Rigden (AIP, Woodbury NY, 1997), 605-614; M. Prosser, P. Walker, and R. Millar, "Differences in students' perceptions of learning physics," *Phys. Educ*. **31** (1), 43-48 (1996); J.M. Schober, *Assessing the Effectiveness of Tutorials in Introductory Physics and Developing an Electricity and Magnetism Concept Inventory*, M.S. & M.A. Thesis, Miami University, 1996 (unpublished); I. Novodvorsky, *Development of an Instrument to Assess Attitudes Toward Science*, Ph.D. Dissertation, University of Arizona, 1993 (unpublished).

[9] To really determine what a person believes regarding a particular issue, current testing practice in the social science suggests that is necessary to ask at least five or six questions on that issue. Questions on related issues may not be sufficient to fully understand the belief in question. This is one reason why the reliability coefficients for sub-scales of survey instruments tend to be smaller (less reliable) than those for sub-scales of diagnostic instruments. See D.J. Mueller, *Measuring Social Attitudes: A Handbook for Researchers and Practioners* (Teachers College, Columbia University, New York, 1986).

[10] W. F. Perry, *Forms of Intellectual and Ethical Development in the College Years* (Holt, Rinehart, & Wilson, NY, 1970).

[11] M. F. Belenky, B. M. Clinchy, N. R. Goldberger, and J. M. Tarule, *Women's Ways of Knowing* (Basic Books, New York NY, 1986).

[12] R. Likert, "A technique for the measurement of attitudes," *Archives of Psychology* No. 140 (1932).

[13] We refers to Edward R. Redish, Richard N. Steinberg, and the author.

[14] A copy of the current version of both the pre and post versions of the survey can be found in Appendix C.

[15] M.C. Linn and N.B. Songer, "Congitive and conceptual change in adolescence," *Am. J. of Educ*., 379-417 (August, 1991). N.B. Songer and M.C. Linn, "How do students' views of science influence knowledge integration," *Jour. Res. Sci. Teaching* **28** (9), 761-784 (1991).

[16] See Refs. 10 & 11.

[17] See Ref. 6.

[18] See Ref. 4.

[19] See Ref. 5.

[20] Indeed, some student comments lead us to suspect that formula sheets may have the tendency of *confirming* student expectations that formulas dominate physics. Their interpretation is that although memorizing lots of formulas is important for professionals, they do not need to do so for the current course. Thus, many faculty may be encouraging precisely that attitude they hope to discourage when they permit the use of formula sheets on exams. Redish, Steinberg, and I are not aware of any research that shows the effect of formula sheets on student perceptions of the coherence of the material.

[21] Note that this is an area where students' beliefs about their abilities may surpass their actual abilities. More detailed investigations will require direct observation of student behavior on solving physics problems.

[22] Redish has referred to this failure as a lack of *parsing* skills. These students, when faced with a complex sentence that they do not understand, will try reading it over and over again until it becomes familiar -- but they still may not understand it. They seem to lack the ability to decompose a complex sentence into its constituent parts in order to make sense of it. See E. F. Redish, "Is the computer appropriate for teaching physics?", *Computers in Physics* **7**, 613 (December 1993).

[23] M.T.H. Chi, P.S. Feltovich, and R. Glaser, “Categorization and representation of physics problems by experts and novices,” *Cog. Sci*. **5**, 121-152 (1982).

[24] See Ref. 7.

[25] See Ref. 12.

[26] The device of plotting three numbers whose sum is fixed in a triangle is well known in elementary particle physics as a Dalitz plot. In our case, the percentage responding agree, disagree, and neutral must add up to 100%.

[27] Note that we have included all items, including those marked with parentheses in Table 3. As remarked above, even though the agreement on these items is not as strong, there is still a strong plurality of our experts in favor of the indicated responses. The shift in the position of the overall items resulting from removing these items is on the order of a few percent and the relative order of the groups is not modified.

[28] Private communication with E.F. Redish and R.N. Steinberg, Summer 1995-96.

[29] See Ref. 6

[30] The ability of an individual to hold conflicting views depending on circumstances is a fundamental tenet of our learning model. See ref. 3 and R. Steinberg and M. Sabella, "Student performance on multiple choice questions vs. open-ended exam problems", *Phys. Teach*. **35** (3) 150-155 (1997) for more discussion of this point.

[31] "Homogeneous" in this case does not of course mean that we assume the students are identical. Rather, it means that the students are "equivalent" — that they are characteristic of the students who are to be found in "that type of class in that type of school."

[32] We choose this reduction from two independent variables to one because the primary variations we observe tend to maintain a fairly constant proportion of neutral responses.

[33] D. J. Mueller, *Measuring Social Attitudes* (Teacher's College, Columbia University, New York, 1986).

[34] This crystal analogy was developed by E.F. Redish to describe the relation between students' mental models and beliefs to what is measured by assessment, private communication.

[35] There is a method of factor analysis to study non-orthogonal factors that uses oblique rotations.

[36] SPSS for Windows, Release 7.0 standard version (Dec. 19, 1995).

[37] Private communication with Gilley Shama (Spring 1997).

[38] Factor loading is the correlation of the item with the factor.

[39] In phenomenographic analysis, the researcher looks at the data obtained and develops categories from the data to classify the results. For more information on phenomenographic analysis, see F. Marton, "Phenomenography — a research approach into investigating different understandings of reality," *J. Thought* **21** (3), 28-49 (1986).

[40] Michael Wittmann is another senior graduate student in the University of Maryland Physics Education Research Group.

[41] See Ref. 33.

[42] R.J. Beichner, "Testing student understanding of kinematics graphs," *Am. J. Phys*. **62** (8), 750-762 (1994).

[43] P.L. Gardner, "The dimensionality of attitude scales: A widely misunderstood idea," *Int. J. Sci. Educ*. **18** (8), 913-919 (1996).

[44] See Ref. 33.

[45] Variance is the square of the standard deviation.

[46] See Ref. 36.