CHAPTER I
INTRODUCTION TO THE STUDY

1.1 Overview

Students come into the classroom with preconceptions about how the world works (Bransford, Brown, & Cocking, 1999). It is claimed that engaging these preconceptions during the teaching practice is necessary to grasp new concepts, which requires that teachers are prepared to draw out their students' existing understanding and help to shape it into scientifically accepted knowledge (Bransford et al., 1999; Donovan, Bransford, & Pellegrino, 1999).

Among many different types of students' preconceptions and difficulties, of special interest to physics educators and researchers are those that originate from some structured cognitive concept or mental model. "The term mental model is frequently used today in science education research to describe the way students understand various scientific concepts and ideas" (Zollman, 1999). Students' mental models may contain contradictory elements (Redish, 1994) and are generally different from scientific models. Spontaneous concepts that result from these mental models today are commonly called alternative conceptions (Wandersee, Mintzes, & Novak, 1994). We will address these terms later in greater detail.

During the teaching process we want, in a sense, to "replace" spontaneous mental models with scientific models that are accepted as valid if they are coherent, stable and experimentally verified. Being familiar with the common mental models and related alternative conceptions, a physics instructor can much more effectively lead the class discussion, particularly before and after a demonstration or an experiment.

Recently, an analytical method for analyzing students' understanding of scientific models was developed by Bao (1999). This method enables quantitative analysis of students' mental models in "real time" (during the lecture or research). To utilize this tool, model analysis inventories of various topics in science must be constructed. The

aim of this research is the construction of a model inventory in one physics domain --
sound propagation.

1.2 Model Analysis

Researchers observed that when the learning of a particular physics topic is explored
through systematic qualitative research, usually a small finite set of commonly
recognized models is identified (Marton, 1986). This finding is a basis for model
analysis (Bao, 1999; Bao & Redish, 2001; Bao, Zollman, Hogg, & Redish, 2000) -- an
analytical tool that gives information about the ability of an individual student and the
class as a whole in order to correctly apply the relevant concept.

Model analysis provides information relevant for instruction in a more
comprehensive way than score-based analysis. By knowing what models students use
and to what extent they use them and by tracing the changes in their model dynamics
instructors can more easily identify possible causes of students' difficulties and develop
better instructional strategies to address them (Bao & Redish, 2000; Bao & Redish, 2001;
Zollman, 1999).

This approach "assumes that the most commonly used mental models are
identified through extensive qualitative research. These known models can then be
mapped onto the choices of an appropriately designed multiple-choice test" (Bao &
Redish, 2001, p.3). This test is called a mental model inventory. The results that it
provides give explicit information about the students' state of understanding. Computer
software that is specially designed for the inventory analyzes and displays the results so
that a teacher gets immediate, real-time feedback during the lesson. This formative type
of assessment is a crucial benefit arising from this method.

Although the more prevalent student models can be reliably identified through
qualitative research, it is possible that some students have models that are significantly
different from the models discovered through the qualitative research, albeit not as
prevalent. For this reason in the model analysis Bao and Redish (2001) also employ a so-
called "null model." It serves to make sure that possible less common, irrelevant and/or
not identified models are also included in the analysis. "With the null model included,

2

the set of models becomes a complete set, i.e. any student responses can be categorized" (Bao & Redish, 2001, p.9).

Based on our earlier (Hrepic, 2002; Hrepic, Zollman, & Rebello, 2002) and current research findings, we advanced some of these principles and consequently adopted testing and analysis procedures somewhat different from those described by Bao and Redish (2001). As a consequence, we employ a different testing format, use different data analysis procedures and display the results in a different way. However, our approach is built on the same basic principles that Model Analysis was built on and which were previously described.

1.3 Why investigate students' understanding of sound?

As a physics topic sound is typically glossed over at all levels as a straightforward example of wave phenomena. Although research related to students' difficulties associated with sound is not as abundant as those related to some other physics topics, numerous difficulties in students' understanding of sound have been identified (Barman, Barman, & Miller, 1996; Hrepic, 1998, 2002; Linder, 1987, 1992, 1993; Linder & Erickson, 1989; Maurines, 1992, 1993; Merino, 1998a, 1998b; Wittmann, 1998; Wittmann, 2001; Wittmann, Steinberg, & Redish, 1999; Wittmann, Steinberg, & Redish, 2002).

Because of sound's virtual omnipresence in our daily life, this topic deserves our attention. Within sound as a topic, we decided to concentrate on its propagation because students' understanding of sound propagation seems to govern their conceptions related to other aspects of sound. Several studies (Hrepic, 1998, 2002; Linder, 1987, 1992, 1993; Linder & Erickson, 1989; Maurines, 1992, 1993; Wittmann, 1998; Wittmann, 2001; Wittmann et al., 1999; Wittmann et al., 2002) suggested that a naive mental model is associated with sound propagation. These authors generally referred to it as a "Particle Model" of sound propagation. In this model, sound travels as a particle-like object. We investigated students' mental models of sound propagation (Hrepic, 2002; Hrepic et al., 2002) and confirmed the existence of this model as a student's common initial alternative

model. We described its variations and more refined aspects and named it the "Entity" Model.

Another important reason to study sound as a topic was that research into students' understanding of sound propagation might give us some insights into how to teach waves better in general based on ontological categories students tend to use while thinking about abstract physical quantities like waves. Further, waves as a topic is a crucial domain not only in classical but also in quantum physics. As an everyday and commonly known phenomenon sound may be an optimal introductory topic to waves.

1.4 Research goal

The goal of the broad research project (Zollman, 1999) of which this study is a part, is to develop tools that will measure students' states of understanding and also trace changes in those states during instruction (ranging over a period from one class section to several weeks of instruction). Once developed, these tools should provide real-time feedback on instruction and promote research in ongoing classes ranging from small seminars to large lectures (Zollman, 1999). The study described in this dissertation is a final part of one of the segments of this research project and it is related to sound propagation. Sound is one of a variety of introductory physics topics with which this major project deals. The specific problems of this study and its research questions are defined below.

1.5 Problem definition

The purpose of this study is to develop a multiple-choice test that can elicit students' mental models of sound propagation during the lecture while using a classroom response system and appropriate software.

Benefits of the real-time, in-class assessment are that it engages students and therefore facilitates interactive learning and peer instruction regardless of the class size. It also provides immediate feedback to the teacher so s/he can adjust the teaching before an exam rather than afterwards it and according to the specific needs of his/her students. Finally it allows for a detailed post-lecture analysis of students' responses and appropriate interventions in subsequent classes and for future generations.

1.6 Research questions

Main question:

> What is the optimal multiple-choice test that can elicit students' mental models of sound propagation in real time during the instruction?

Sub questions:

> What are the common fundamental mechanisms of sound propagation that can be drawn from students' mental models of sound propagation identified in earlier studies? The fundamental mechanism in this context refers to a set of features that different mental models might have in common.

> What are the optimal test questions and optimal answer choices that can elicit students' mental models of sound propagation in real time?

> Is model analysis the optimal analytical tool for analysis of students' responses in this test? If not, what is the analytical method that best fits the particulars of the nature of models of sound propagation?

> How reliable is the test?

> How valid is the test?

> How do we represent the data so that the display provides a variety of instruction guiding information?

> How do we adapt this test and the accompanying analysis software to make it commercially available via the students' response systems?

1.7 Problem evaluation – contribution to science education

A tool for analyzing the state of students' understanding and their mental models was recently developed (Bao, 1999). This study is a part of the broader research effort that strives to construct model analysis inventories primarily for different topics in physics taught at the introductory college level. As a part of this project, this dissertation research was specifically concerned with sound propagation as one of those topics. Together with contemporary classroom technology (classroom response system and on-line homework), the constructed model analysis inventories will permit the extraction and quantitative display of the effects of instruction on a class's knowledge during the

teaching process. This process will allow for instant adjustments in teaching approaches as opposed to an ongoing praxis of post-course assessments and adjustments (Zollman, 1999).

CHAPTER II
REVIEW OF THE LITERATURE

2.1 Introduction

This chapter introduces existing literature related to constructivism (as a theoretical framework under which we operate) to the theory of mental models (as analytical framework that we employ) and to previous investigations of students' understanding of sound, which is the area of physics about which we are concerned in this study.

2.2 Constructivism

Constructivism is an educational philosophy, according to which learners construct knowledge for themselves (Asynchronous Learning Networks, 1998). It differs from the traditional view that knowledge exists independently of the individual and that the mind is a "tabula rasa," (a blank tablet), on which knowledge is to be imprinted.

To the objectivists "knowledge and truth exist outside the mind of the individual and are therefore objective" (Runes, 1962, p.217). According to this assumption, knowledge is "true" if it corresponds to reality and "false" if it does not (Bodner, Klobuchar, & Geelan, 2001). Constructivist theories on the other hand, are based on the assumption that "knowledge is constructed in the mind of the learner" and it "results from a more or less continual process in which it is both built and continually tested" (Bodner et al., 2001).

However, according to the constructivist view the knowledge we are "allowed" to construct is the only useful knowledge that that "works" (Bodner et al., 2001). Consequently, knowledge should be judged in terms of its viability, rather than in terms of whether it is true or false. Similarly, Polkinghorne (1992) argues that constructivist theories require a shift "from metaphors of correctness to those of utility." Constructivism is not considered another epistemology or a way of knowing. Rather, it is said to be a way of thinking about knowing. As such, it serves as a reference for building models of teaching, learning and curriculum (Tam, 2000; Tobin & Tippins, 1993).

2.2.1 Forms of constructivism

Many different schools within this theory fall within the same basic assumption about learning (Chen, 2001). The main two are cognitive constructivism and social constructivism. Cognitive constructivism has two main versions. One is 'mild' (or 'trivial') (Boudourides, 1998) and it is based on the work of Piaget (1972). Knowledge is actively constructed by the learner and is not passively transmitted by the educator. In the radical constructivism of von Glasersfeld (1990), cognition is considered adaptive in the sense that it is based on the learner's experience and constantly modified by it. Social constructivism is generally attributed to Vygotsky (1978), who challenged Piaget's ideas by stressing the primary role of communication and social life in formation of meaning and cognition. As a result of their complexity, Philips (1995) sees various forms of constructivism spread out along several different dimensions or continua. Two of the most significant axes or dimensions are "individual psychology versus public discipline" and "humans, the creators, versus nature, the instructor" (Phillips, 1995). According to Phillips, (1995) this second dimension is crucial because at a point somewhere along this dimension one ceases to be a constructivist.

2.2.2 Problems with the constructivist theory

Due to its many generally appealing characteristics, constructivism has enormous influence on contemporary science education thought, research and practice (Phillips, 1995) and this trend has lasted for the past few decades. However, the critics of the theory are also abundant (Matthews, 1993; Osborne, 1996) and some urge caution in its adoption (Millar, 1989; Solomon, 1994).

Three main objections are raised about classroom applications of constructivist theories of knowledge (Bodner et al., 2001). Constructivism is being accused that (1) it questions whether a real world actually exists, (2) it prevents us from saying that a student is wrong and (3) that by concentrating on the process of learning it ignores the role of those who influence the learning. Bodner et al. (2001) argue that all these objections arose because most of the constructivist or radical constructivist theories have been based on the work of Jean Piaget which emphasizes the role of the individual in knowledge construction. Bodner et al. (2001) suggest that an alternative view of the

construction of knowledge proposed by the clinical psychologist George Kelly (Kelly, 1955) addresses these objections. Kelly, in his theory of personality, similarly to Piaget, emphasizes the role of the individual in the construction of knowledge, but he also provides a basis for thinking about the kinds of interactions between people that can facilitate this construction (Bodner et al., 2001). According to social constructivism, (Vygotsky, 1978) the interaction between learners is a primary mechanism through which the learning occurs. Therefore, incorporation of social aspects of learning into the constructivist theory addresses the aforementioned objections, especially the second and the third ones.

In spite of these objections to some constructivists' views, even the authors who relatively critical of the constructivist theory admit that "there is a very broad and loose sense in which all of us these days are constructivists" (Phillips, 1995).

2.2.3 Constructivism and ideological perspectives of this research

The premises of constructivism are baselines of philosophical assumptions or ideological perspectives employed in this research. Constructivists view learning as the result of mental construction.

From the constructivists' point of view, mental models can be defined as internal schemes for understanding both the tools with which knowledge is constructed and the foundation upon which knowledge is constructed (Brandt, 2002). Further, it is the constructivist's view that students learn by fitting new information together with what they already know. This concept links constructivism to this research because we are trying to find out what students "already know" in the domain of sound propagation. This information is necessary because teaching of physics cannot be effective, in general, if a presentation does not take into account the students' existing alternative conceptions (Bransford et al., 1999; Donovan et al., 1999). Also, in the constructivist's perception learning is affected by the context in which it occurs and in this research we want to explore if, and in what way(s), the students' models depend on the context in which sound propagates.

2.3 Mental models

2.3.1 Definition and nature of mental models

Wider studies of mental model definitions show that no consensus exists about the definition of the term mental model and "some definitions of the concept are even contradictory" (Van der Veer, 2000). According to Cañas and Antolí (1998) the main reason for disagreement in the definition of the mental model is that the term has been used by researchers who work in different fields and who focused on its different aspects. According to Van der Veer, (2000) although there is no agreement about the exact definition of the concept in general the them "mental model" refers to the internal representations that people form about the environment through their interaction with it.

The notion of the mental model as a "small-scale model" of reality can be traced to the work of Kenneth Craik (1943) who stated that mental models can be constructed from perception, imagination or from comprehension of the discourse.

According to Johnson-Laird, (1983) while reasoning people construct working cognitive representations of phenomena with which they interact. They build mental representations by associating the incoming information with their existing knowledge. In this sense, while reasoning people construct a mental model. With respect to real-world phenomena, mental models are similar in structure but simpler, and they serve to provide explanation (Johnson-Laird, 1983).

Norman (1983) defines the mental model as the mental representation constructed through interaction with the target system and constantly modified throughout this interaction. Listed below are Norman's general observations related to mental models (Norman, 1983, p.8):

a) Mental models are incomplete.

b) People's abilities to "run" [employ] their models are severely limited.

c) Mental models are unstable over time (due to forgetting and mixing of old and new incoming information).

d) Mental models do not have firm boundaries.

e) Mental models are parsimonious. Users tend to do extra physical actions rather than the mental planning that would allow them to avoid those actions.

f) People often feel uncertain of their own knowledge, even when it is in fact complete and correct (Norman, 1983, p.8).

With the term mental model, Vosniadou (1994) refers to "a special kind of mental representation, an analog representation, which individuals generate during cognitive functioning and which has the special characteristic that it preserves the structure of the thing it is supposed to represent." Vosniadou (1994) introduced the notion of a "synthetic model," which is constructed as a combination of the aspects of a student's initial model (one based on everyday experience) and the culturally accepted, scientific model.

Young (1983) uses the term "user's conceptual model," which is "a more or less definite representation or metaphor that a user adopts to guide his actions and help him interpret the device's behavior" (Young, 1983, p.35). Young states that it is possible to have different mental models about a system representing different kinds of information.

Minsky in his book *Society of Mind* (Minsky, 1986, p.303) writes that, "Jack knows about A means that there is a 'Model' M of A inside Jack's head." For our purpose, this statement is too broad to be considered a useful definition of a mental model. However, his notion of model usefulness is applicable, "Jack considers M to be a good model of A to the extent that he finds M useful for answering questions about A" (Minsky, 1986, p.303).

Holland et al. (1989) emphasize the dynamic nature of mental models. For these authors mental models are partially based in static prior knowledge, but "they are themselves transient, dynamic representations of particular unique situations" (Holland et al., 1989, p.14). Therefore, mental models are changed and most of the time refined as additional information is acquired.

Through the set of principles related to mental modes and their implications, Redish (1994) summarizes what he calls a framework for understanding students' learning. His fundamental hypothesis about how the mind works is that people tend to organize their experiences and observations into patterns or mental models. Redish

(1994, p.797) builds largely on Norman's work (Norman, 1983) and defines that mental models have the following properties:

> ➤ "They consist of propositions, images, rules of procedure and statements as to when and how they are to be used.
>
> ➤ They may contain contradictory elements.
>
> ➤ They may be incomplete.
>
> ➤ People may not know how to 'run' [employ] the procedures present in their mental models.
>
> ➤ Elements of a mental model do not have firm boundaries. Similar elements may get confused.
>
> ➤ Mental models tend to minimize expenditure of mental energy. People will often do extra physical activities - sometimes very time consuming and difficult - in order to avoid a little bit of serious thinking...
>
> ➤ Students may hold contradictory elements in their minds without being aware that they contradict" (Redish, 1994, p.797).

diSessa (1996) defines mental models as "frequently instructed knowledge forms that...can be the basis for extended and articulate arguments in the course of developing or displaying explanations or in problem solving" (diSessa, 1996, p. 12). Mental models rely on elaborate and well-developed descriptive components – spatial configurations and causal events.

Witmann, et al. (1999) define mental models as patterns of associations (i.e. rules, images, maps or analogies) used to guide spontaneous reasoning. According to these authors, students' mental models are often incomplete, self-contradictory and inconsistent with experimental data.

In applying the concept of mental models to human-computer interaction, Van der Veer (2000) considers mental models "any type of mental representation that enables and facilitates the interaction with the system and that develops during the interaction with the system" (Van der Veer, 2000).

Taber (2000) claims that it is possible for a learner to hold several different, yet stable and coherent explanatory schemes that are applied to the same concept area. "This is a

significant claim as research evidence that learners apply several different conceptions to a concept area that has been interpreted as implying that their thinking is not theory-like, but incoherent, fragmentary and closely context-bound" (Taber, 2000, p.399). This paper argues that, at least in some cases, multiple frameworks are genuine evidence for the manifold of learners' conceptualizations.

Bao and Redish (2001) state they use the term mental model in a broad and inclusive sense and define it as "a robust and coherent knowledge element or strongly associated set of knowledge elements. A mental model may be simple or complex, correct or incorrect, recalled as a whole or generated spontaneously in response to a situation" (Bao & Redish, 2001, p.2).

Brandt (2002) claims that from the constructivists' point of view, mental models can be defined as "internal schemes for understanding that are both are the tools with which knowledge is constructed and the foundation upon which knowledge is constructed."

According to Johnson-Laird and Byrne (2002) "mental models are representations in the mind of real or imaginary situations…Mental models underlie visual images, but they can also be abstract, representing situations that cannot be visualized." This statement is important for understanding the mental models of sound propagation as we found them in our study.

Greca and Moreira (2002) provide an operable account from the physics education research (PER) perspective: "A mental model is an internal representation, that acts out as a structural analogue of situations or processes. Its role is to account for the individuals' reasoning both when they try to understand discourse and when they try to explain and predict the physical world behavior" (Greca & Moreira, 2002, p. 108). They also state that the understanding of a scientific theory would require the construction of mental models in the mind of the one who wants to understand it. From Johnson-Laird's work these authors stress his belief that the core of understanding lies in existence of working models in the mind of the individual. Greca and Moreira also state that "it would seem that students recursively generate mental models based on their initial ones, in an attempt to fit into them or to give meaning to the different contents of the subject

matter" (Greca & Moreira, 2002, p. 116). These "bifurcated" models that appear as a product of these successive reformulations are called hybrid models by these authors.

Building on his article from 1996, diSessa (2002b) states, "To my mind, mental models should (1) involve a strong, well developed "substrate" knowledge system, such as spatial reasoning, (2) allow explicit hypothetical reasoning, and (3) involve only a small, well defined class of causal inferences" (diSessa, 2002, p.27).

In personal correspondence diSessa (2002a) told the author, "My definition of a mental model entails:

1. Strong "base descriptive vocabulary" - e.g. spatial configuration of identifiable kinds of things.

2. Localized causality - i.e. just a few principles (e.g. "gears work by conveying motion via contact" or 'resistors work by Ohm's law').

3. Explicit hypothetical reasoning - e.g. "if this gear moves that way then connected gears move ..."

2.3.2 Definition of the mental model employed in this study

We understand the mental model in a way proposed by Greca and Moreira, (2002) i.e. as "an internal representation, which acts out as a structural analogue of situations or processes. Its role is to account for the individuals' reasoning both when they try to understand discourse and when they try to explain and predict the physical world behavior" (Greca & Moreira, 2002, p.108).

In addition to this we assign properties proposed by Redish (1994, p.797) to mental models, along with several properties defined by other authors:

➢ Mental models are dynamic, evolving systems (Holland et al., 1989; Johnson-Laird, 1983; Norman, 1983).

➢ Mental models underlie visual images, but they can also be abstract, representing situations that cannot be visualized (Johnson-Laird & Byrne, 2002).

➢ Mental models "bifurcate" (Greca & Moreira, 2002)

➢ Mental models "synthesize" (Vosniadou, 1994) i.e. "hybridize" (Hrepic, 2002; Hrepic et al., 2002).

➢ Mental models can be mixed (Bao & Redish, 2001; Taber, 2000; Young, 1983).

Regarding mental models of sound propagation identified and described earlier (Hrepic, 2002; Hrepic et al., 2002) and the references therein, in this study we will analyze those that fulfilled diSessa's (2002a, 2002b) requirements for mental models in addition to the aforementioned definition as well. diSessa requires that mental models:

1. Involve the strong "base descriptive vocabulary" e.g., spatial configuration of identifiable kinds of things,

2. Involve only a small, well defined class of causal inferences i.e., just a few principles (e.g., "gears work by conveying motion via contact" or "resistors work by Ohm's law".

3. Allow explicit hypothetical reasoning e.g. "if this gear moves that way then the connected gears move ..."

When talking about "identifiable kinds of things" diSessa did not restrict them on "correct" things and neither do we. We also do not restrict mental models to concrete "ingredients" (those that can be visualized) (Johnson-Laird & Byrne, 2002), but recognize abstract ones as valid too, whether they are "correct abstracts" (like the electric field) or incorrect abstracts (like the ether).

2.4 Mental model states

It is well documented that in introductory college physics, students often do not recognize relevant conditions in which to use their mental models appropriately (Bao & Redish, 2001). The way students use mental models in different contexts (i.e. problem situation) define their mental model state.

2.4.1 Context dependence of mental models

Different models are often activated by the presentation of a new situation or problem. Research reveals significant inconsistency of student responses in apparently different situations that an expert would consider equivalent (Clough, 1986; Maloney & Siegler, 1993).

As an example, while analyzing the Force Concept Inventory, which is a tool for understanding students' models in dynamics, Schecker and Gerdes (1999) were looking for the possible dependence of students' models in the context of different questions. In the two different questions students were asked about forces on a golf ball and on a soccer ball after these have been hit and while they were flying through the air. In the golf ball context, 42 of 87 participants included a force in the direction of motion in their answer.

However in the context of the soccer ball, 23 of these 42 students omitted this non-existing force and included in their answers either gravity alone, or gravity together with air resistance. (The participants in the study were students from Europe and therefore were more familiar with soccer than with golf.) After obtaining a similar result in another question, authors concluded that the models students apply are context dependent.

Consequently, in our study probing the context dependence of students' models we must determine the scope and limitations of the students' understanding and strength of the transfer of their knowledge. Different contexts are generally considered apparently different situations that an expert would consider equivalent (and would treat identically), and which are at the same time perceived as essentially different by a novice. For the purpose of this study, two different contexts can be defined in an alternative, less outcome-based way. Different contexts are situations that are different enough so there is no single non-zero number that might relate them to each other. Instead, the difference needs to be conceptually or verbally described. For example, according to this definition a billiard ball hitting the wall and a bowling ball kicking it are the same context because the only difference is in the mass of the two objects that kick something and there is a single numerical factor that can relate those two masses.

Within the same context, there may be different aspects of the phenomena in explaining which students will employ different models. Therefore, different models may be used in different instances within the same context as well. For this reason we will probe students' models from several aspects within the same contexts.

2.4.2 Defining mental model states

According to Bao and Redish (2001) one should keep in mind that "the mental states of the individual students tend to be mixed, especially when they are making a transition from an initial state dominated by a naive incorrect model to an expert state" (Bao & Redish, 2001, p.3). Bao and Redish (2001) state that "if a student always uses a particular mental model in a reasonably coherent way in response to a set of expert-equivalent questions we say they are in a pure model state. If the student uses a mixture of distinct mental models in response to the set of questions we say the student is in a mixed model state" (Bao & Redish, 2001, p.8). According to these authors, a student in mixed model state simultaneously occupies a number of different models with different probabilities. Applied over a period of time, model analysis traces the change of model from the old one, through the mixed state, to a new one.

Another important model state is a hybrid model state in which only one, so called hybrid model is used. A hybrid model is a composite mental model that unifies different features of a common initial alternative model and the scientifically accepted mental model. A hybrid model is at the same time inconsistent (by one or more features) with both models from which it was derived (Hrepic, 2002; Hrepic et al., 2002).

2.4.3 Hybrid models and hybrid model state

Our understanding of mental model dynamics is in accordance with an imperfect mental model view. This view assumes that "self-explaining is the process of revising (and updating) one's own mental model, which is imperfect in some ways" (Chi, 2000, p.196). According to this view, a majority of students do not generate a similar explanation and each student may have in some ways a unique naive model (Chi, 2000). Greca and Moreira (2002) further state that it seems that "students recursively generate mental models based on their initial ones, in an attempt to fit into them or to give meaning to the different contents of the subject matter" (p. 116). Models that appear as products of successive reformulations are called hybrid models by Greca and Moreira (2002). These perspectives (Chi, 2000; Greca & Moreira, 2002) do not agree with Bao's (1999) view that "the set of possible models is bounded". Vosniadou (1994) does not impose any restrictions on the possibility of creation of what she calls synthetic models as well.

Our previous results (Hrepic, 2002; Hrepic et al., 2002) support Chi's (2000) and Greca and Moreira's (2002) claims although we found that certain overall structure in the dynamics of model upgrading and repairing exist in the case of models related to sound propagation. In this content domain the process of improvement of the mental models generally begins with the Entity Model, which is later upgraded with the features of the Wave Model. However, this "overall structure" does not limit the number of possible models. The possible outcomes of model restructuring are bounded only by the student's imagination. In the mentioned study, (Hrepic, 2002; Hrepic et al., 2002) students' demonstrated a great deal of inventiveness in reshaping their models when discrepancies were pointed out between models and their experience.

We build on Greca's (2002) term "hybrid model," and Vosniadu's (1994) notion of a synthetic model and define a hybrid model as a mental model that contains the combination of features of the dominant initial alternative model and the scientifically accepted mental model. We add the requirement that the hybrid model is at the same time inconsistent (in one or more features) with both models from which it derived. In the case of sound, these essential models are the Entity Model (initial alternative model) and the Wave Model (scientifically accepted model). If a student consistently applies a hybrid model across the situations, we say he/she is in a hybrid model state. In our view it is an important special case of the pure model state.

2.4.4 Hybrid model state and mixed model state

If a hybrid model is the only model that a student uses (e.g. during the interview) and if it is applied in more than one instance (e.g. question or context), we call the associated model state a hybrid model state. Unlike in the hybrid model state, in the mixed model state the student applies more than one model. Models combined in the mixed model state can be also one or more hybrid models. So, a hybrid model state is a single model state, and a mixed state is a multiple model state. This is the reason why a hybrid model state is just a special case of a pure model state.

The last of the possible model states that we need to mention is a "no model state" in which a student incoherently uses different, isolated and incoherent conceptual schemes or resources (in a wide sense of that term) to provide explanations of

18

phenomena. This classification is primarily useful in domains of physics where only one dominant alternative model exists. This study indicates that sound is one such domain. Figure 2.1 represents different model states pictorially. The model features in the figure can be any knowledge structure that is simpler or more fundamental than a mental model (e.g. p-prim, conceptual resource, facet of knowledge and so on).   These simpler knowledge structures are described in the section that follows.
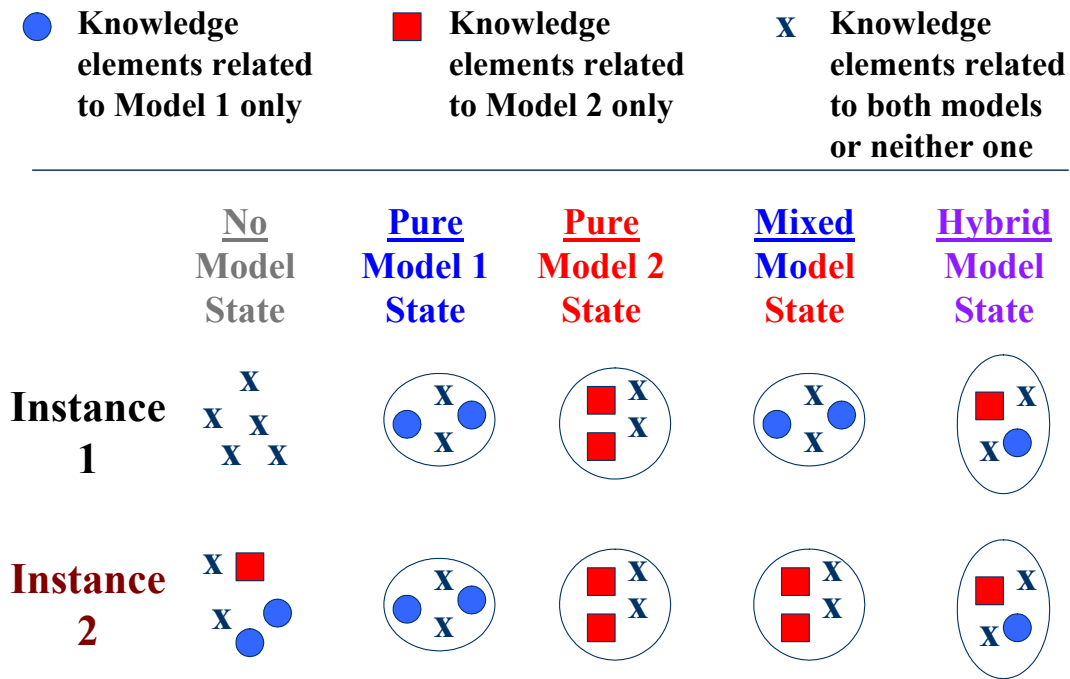

Figure 2.1. Representation of the different model states

2.5   Knowledge structures at a smaller scale than a mental model

Physics education researchers today operate with a variety of mental structures or modes of reasoning (Wittmann, 2001) that are considered more fundamental than the mental model. Of these, we will define here several that are most widely accepted and used.

P-Prims

diSessa (1993) introduced the phenomenological primitive or p-prim as a hypothetical knowledge structure that often originates as a minimal abstraction of everyday phenomena. P-prims are self-explanatory. They are used as if they need no justification – something happens "because that's the way things are" (diSessa, 1993). "They have predicate logic but this logic is intended only as a familiar example of the reasoning process " (diSessa, 1993, p.116).

Conceptual resources

The concept of the resource as the mental structure was introduced by Hammer (1996; 2000). He defines the resource as "a unit of mind-code" (Hammer, 2002). To explain it, he uses the analogy with a computer program: The resource would be analogous to a sub-routine – one or more functions put together to perform a single useful operation. In some cases the resource and the p-prim can be the same, but Hammer (2000) distinguishes the resource from the p-prim (phenomenological primitive) as the resource does not have to be either phenomenological (can be epistemological, procedural…) or primitive. (In a sense, that a resource is not necessarily the smallest meaningful unit, but rather, the smallest practically useful unit of mind processes.)

Alternative conceptions

The term alternative conception refers to "experience-based explanations constructed by a learner to make a range of natural phenomena and objects intelligible" (Wandersee et al., 1994, p.178). As a synonym for alternative conception, many authors today use a new-old term "misconception" (Bao & Redish, 2001; Clerk & Rutherford, 2000) and some also differentiate among them. Examples of the latter ones are Abimbola and Baba (1996) who, for the purpose of their study, defined "misconception" as an idea that is clearly in conflict with scientific conceptions and is therefore wrong. They defined an alternative conception as an idea which is neither clearly conflicting nor clearly compatible with scientific conceptions but which has its own value and is therefore not necessarily wrong (Abimbola & Baba, 1996). Wandersee et al. (1994) consider these two terms synonyms, but also suggest the term alternative conception is more appropriate.

Clerk and Rutherford (2000) define that "a misconception exists if the model constructed by an individual fails to match the model accepted by the mainstream science community in a given situation" (Clerk & Rutherford, 2000, p.704).

While putting the misconception into relation with a mental model, Bao and Redish (2001) define that misconceptions can be viewed as "reasoning involving mental models that have problematic elements for the student's creation of an experts view and that appear in a given population with significant probabilities" (Bao & Redish, 2001, p.2).

Holding misconceptions theoretically ambiguous, Wittmann (2001) uses term reasoning resources in general fashion to describe any of the smaller grain size modes of reasoning (p-prims, facets of knowledge, intuitive rules, etc). He also distinguishes these from a higher-level concept – a coordination class.

Coordination class

diSessa and Sherin (1998) introduced "coordination class" as the type of concept that is relevant for science education research and teaching. They define it as "systematically connected ways of getting information from the world" (diSessa & Sherin, 1998, p.1171). It is characterized by "an accumulation of a complex and broad set of strategies and understandings" (diSessa & Sherin, 1998, p.1173). So, unlike the other mentioned mental constructions, a coordination class is a mixture of both knowledge obtaining strategies and knowledge constructs. Examples of coordination classes are "an object" and "an event" (Wittmann, 2001). Depending on the actual example, coordination class may or may not be of a smaller grain size than a mental model.

Facets of students' knowledge

In his description of students' knowledge, Minstrell (1992) is defining and cataloging the pieces of knowledge or reasoning that students seem to be applying in problem situations. He calls these pieces the "facets". We will address this concept later in much more detail because this study contributed to this aspect of knowledge structuring and to the corresponding ways of teaching.

2.6 Model analysis inventory creation process

There are several basic steps in the creation of the model analysis inventory (Bao & Redish, 2000; Bao & Redish, 2001; Zollman, 1999):

1. Common student models are identified and validated in a particular physics domain. If previous research is not enough for this purpose, additional research through in-depth interviews is conducted. As an example, the following models are commonly found in the domain of dynamics:

    o Newtonian model - acceleration of the body is proportional to force applied to it (a~F).

    o Aristotelian model - velocity of the body is proportional to force applied on it (v~F).

    o Impetus model – the body moves until an impetus (a sort of moving agent) given to it is exhausted. This model "explains" why the body does not stop when the force stops acting.

    o Null model - represents irrelevant (marginally present) and/or possibly not identified models.

2. Multiple-choice inventory questions are designed to track and measure the development of students' models.

3. The effectiveness of the questions is validated through the research.

4. Responses are analyzed simultaneously with the contexts in which they are given. The results obtained this way provide explicit information about the mental models that students use.

5. Finally, computer software specially designed for a particular inventory is used to obtain immediate feedback in real-time teaching.

After constructing the test, we will administer it to determine its validity and reliability in a way that will be described in Chapter III that deals with methodology.

2.7 Previous work on difficulties in the understanding of sound

In their pioneering work on this subject, Linder and Erickson (1987) identified a number of difficulties that students express in understanding sound. In this early stage, they also realized that students apply different reasoning in different contexts. They conducted a phenomenographic study and interviewed 10 students who completed a baccalaureate degree with physics as a major subject and were enrolled in a teacher education program. The same authors later (Linder & Erickson, 1989) structured their findings into two qualitatively different ways in which students describe the phenomena of sound:

"The microscopic perspective:

➢ Sound is an entity that is carried by individual molecules through a medium.

➢ Sound is an entity that is transferred from one molecule to another through a medium.

The macroscopic perspective:

➢ Sound is a traveling bounded substance with impetus, usually in the form of the flowing air.

➢ Sound is a bounded substance in the form of some traveling pattern" (Linder & Erickson, 1989, p.494,496).

In his review article, Linder (1992) listed observed difficulties in students' understanding of sound:

➢ "Sound is an entity that is carried by individual molecules as they move through a medium.

➢ Sound is an entity that is transferred from one molecule to another through a medium.

➢ Sound is a traveling bounded substance with impetus, usually in the form of the flowing air.

➢ Sound is a bounded substance in the form of some traveling pattern.

- Sound is linked to the concept of waves as part of a mathematical physics modeling system (and in this context could not be distinguished from light: the wave equations look identical)" (Linder, 1992, p.258).

In the same paper, Linder proposed several possible reasons for these difficulties:

- Some students seem to be comfortable with conceptualizing the physics in one way and knowing it in another.
- Teachers sometimes use inappropriate analogies (for example, water waves are often used as an analogical example of transverse waves with sinusoidal wave profile).
- Terminology related to sound is often poorly understood by students and sometimes it is also poorly defined in literature.
- Some common oversimplifications in the topics' presentations in the literature, that have historical roots, may cause problematic understanding.
- Explanations and visual representations in introductory physics textbooks are often misleading (Linder, 1992).

A year later, Linder (1993) identified three qualitatively different ways of describing sound propagation:

- "Conceptualization No. 1: the speed of sound is a function of the physical obstruction that molecules present to the sound as it navigates its way through a medium" (p.656).

Linder claims that this conceptualization is based on what he calls a sound-resistance factor: "Conceptualized as a physical thing, sound is slowed down by physical obstacles as it travels through a medium" (Linder, 1993, p.656). The two types of the obstacles for propagation are: physical size of the molecules and the density of the molecules in the medium. The resistance is smallest in a vacuum – so speed is the greatest.

- "Conceptualization No. 2: the speed of sound is a function of molecular separation".

Here, sound is conceptualized "as an entity that is carried by molecules for a certain distance and then transferred to other ongoing molecules" (Linder, 1993, p.658).

Consequently the speed of the sound is determined by the separation of individual sound carrying molecules and therefore it is greater in a denser media.

> ”Conceptualization No. 3: the speed of sound is a function of the compressibility of a medium (the more compressible the medium, the faster the propagation and vice versa)” (Linder, 1993, p.658).

Linder concludes that it seems that all three conceptualizations are the result of the students being taught that certain factors (such as density, pressure and temperature) affect the speed of the sound without any explanation of how these factors affect it. Another important contribution to this subject was made by Maurines (1993). He reported results of his preliminary inquiry conducted in the form of conceptual paper and pencil questionnaires that were administered to nearly 600 sixteen-year-old French students before any lessons about the sound. Summarized, his findings related to students' understanding of the sound are:

> Sound velocity depends on the source, on the signal amplitude (proportionally) and can decrease with time.

> The medium is the passive support or even useless for sound propagation.

> Sound can propagate through the vacuum.

> Propagation is especially difficult when the medium is dense.

> “The supply, a mixture of energy, intensity and speed is given by the source to the medium and is materialized in the 'sound particle' ” (Maurines, 1993, p.201).

Comparing these findings with his study about propagation of the visible mechanical signals, Maurines concludes that the same mechanistic rationale observed for the signal on the rope can be seen in the case of sound propagation. The signal is again the material object created and set in motion by the source. The signal is materialized in the "supply" (a mixture of a force, energy and speed) given by the source.

Barman et al. (1996) used sound as a topic to compare two teaching methods – traditional and learning cycle. Thirty-four fifth grade students were randomly selected from a pool of 51 students and assigned to the two treatment groups (learning cycle and textbook/demonstration method). The same instructor taught both classes for the two-

week unit.   Related to students' understanding of sound, the authors found that: the students

- generally, viewed sound as an "object" that moved from one place to another,
- thought that sound could be produced without any material objects,
- concluded that sound is a transversal wave that travels similar to the way water waves and light waves do, and
- believed that when waves interact with a solid surface they are being destroyed.

Hrepic (1998) investigated students' understanding of sound using a written survey with open-ended, mostly original questions that covered a wide array of sound-related phenomena and situations.   He compared results obtained from 8th graders (middle school), high school juniors and college seniors (most of whom were physics majors).   This author concluded that almost all observed alternative conceptions can be found at all of these levels.

The author calls the students' conception that "sound propagates as a particle-like object" the first "law" of spontaneous acoustics.   He claims that several other alternative conceptions are consequences of the particle conception about propagation of sound. These are:

- Material obstacles slow down propagation of sound,
- If louder, sound travels faster,
- The speed of sound depends on the movement of the sound source., and
- Sound can be perceived in distance, like a distant object.

Another set of alternative conceptions the author classified as generated by inappropriate knowledge transfer:

- Not all the materials can propagate the sound.
- Electric insulators propagate the sound poorly.
- Sound energy is not generally transformable.

Finally, three alternative concepts were claimed to be generated or enforced by school knowledge.

- The denser the medium, the faster sound propagates.
- Speed of sound depends on its frequency.

➢ Wind influences the frequency of received sound.

In two consecutive articles, Merino (1998a; 1998b) presented a set of observations related to common mistaken ideas about sound at the college level. The author derived these observations from personal teaching experience. The first article (Merino, 1998b) deals with the relation of sound loudness and its intensity. Problems he observed related to these are that students often wrongly assume that:

➢ Intensity and loudness are the same thing,

➢ Doubling the intensity of the acoustic wave doubles the acoustic level,

➢ If a frequency is halved, the corresponding pitch is also halved,

➢ Regardless of the frequency, a similar acoustic energy always produces the same loudness,

➢ The timbre of a complex sound is a mere overlap of the partials.

Students are also often not aware of the existence of virtual pitch and the fact that loudness, pitch and timbre are interdependent properties (Merino, 1998b). Merino's second article (1998a) deals with concepts of sound pitch and timbre. An understanding of the concept of pitch presents problems since this sensation is commonly linked in a simplistic manner to the fundamental frequency. Although it is the main component in the perception of a tone, the fundamental frequency is not the only one. The tone that we perceive depends also on intensity, spectral composition, the duration of the stimulus, the amplitude envelope and the presence of other sounds. Also, to perceive the sensation of a tone, the brain needs certain minimal stimulation threshold time. If the stimulus duration is shorter than this time, the sound is described as a "click" with undefined pitch (Merino, 1998a).

The timbre is a property of the auditory sensation, which allows two sounds of equal loudness and pitch to be distinguished. Therefore, timbre is the subjective correlation of all the properties that do not directly intervene in loudness and pitch (such as its spectral power distribution, temporal envelope and degree of anharmonicity (Merino, 1998a). Merino concludes that any understanding of the nature of sound invariably involves the study of the intrinsic properties of loudness, pitch and timbre.

However, a serious didactic problem in teaching arises from the complex structure of those three concepts. A common mistake is identifying the subjective sensory properties, loudness, pitch and timbre, with the physical magnitudes. Another widespread mistake is the association of loudness with wave amplitude only, pitch with frequency only and timbre with the mere overlapping of two higher partials. Merino suggests that the problem might be solved if teachers improve their knowledge of the three acoustic sensations. Together with practical demonstrations, important didactic help is available in the form of instruments (dB-meters, MIDI synthesizers, etc) and software.

Beaty (2000) compiled a list of children's misconceptions about science as a result of the AIP Operation Physics Project. The section on sound contains the following list (Beaty, 2000):

1. Loudness and pitch of sounds are confused with each other.
2. You can see and hear a distant event at the same moment.
3. The more mass in a pendulum bob, the faster it swings.
4. Hitting an object harder changes its pitch.
5. In a telephone, actual sounds are carried through the wire rather than electrical pulses.
6. Human voice sounds are produced by a large number of vocal chords.
7. Sound moves faster in air than in solids (air is "thinner" and forms less of a barrier).
8. Sound moves between particles of matter (in empty space) rather than matter.
9. In wind instruments, the instrument itself vibrates while the internal air column does not.
10. As waves move, matter moves along with them.
11. The pitch of whistles or sirens on moving vehicles is changed by the driver as the vehicle passes.
12. The pitch of a tuning fork will change as it "slows down," (i.e. "runs" out of energy)".

Wittmann et al. (1999) report on research done in the second semester of a three-semester university physics course at the University of Maryland. In the investigation, a

variety of probes were used, including videotaped individual demonstration interviews, pretests, (short, ungraded quizzes that accompany tutorials) examination questions, free-response as well as Multiple-Choice, Multiple-Response (MCMR) questions and specially designed diagnostic tests. The authors have found that while describing the physics of waves, many students use analogies with Newtonian particle mechanics and related ideas of force, energy and collisions between the objects. To describe students' difficulties with mechanical waves, the authors organized the data in terms of a mental model. Mental models are defined here as patterns of associations (rules, images, maps, or analogies), used to guide spontaneous answers and reasoning in unfamiliar situations. In their data, the authors see evidence of what they call the "Particle Pulses Mental Model" of waves. Analogies that students make using this model are typical for mechanical, particle-physics models (Wittmann et al., 1999). To illustrate students' descriptions of the ways in which the sound waves affect the air, the authors use the metaphor of a "surfer on a wave" where the surfer is a medium pushed by the wave. In accordance with the surfer metaphor, students were often describing sinusoidal waves as a succession of pulses, each exerting a force on particles of the medium and in the direction of wave propagation. Another metaphor that authors used for "particle" description of the wave propagation was the "ball-toss" analogy where the ball was an analogue for a wave.

In his dissertation, (Wittmann, 1998) which was the basis for this paper, Wittmann uses the term "particle pulse pattern of associations" and states that it can be loosely referred to as a particle model. In the same research, Wittmann also realized that some students understand the wave as propagating air although he did not report this as one of the possible mental models in this domain.

Wittmann (2001) later reported results of reanalysis of his data using diSessa and Sherin's concept of coordination class (diSessa & Sherin, 1998). This approach suggests that students' use of the specific reasoning resources is guided by possibly unconscious clues. In this paper Wittmann defines a coordination class as "one of many different possible types of concepts in which nets of simple and reasonable pieces of information are chosen and linked together" (Wittmann, 2001). The author also uses the term

"reasoning resources" broadly to describe any of the smaller grain size modes of reasoning (p-prims, facets of knowledge, intuitive rules, etc).

By considering students' understanding of the mechanical waves while using the coordination classes as those made up of reasoning resources, Wittmann shows that students' associations are primarily built around the concept of a particle and not around a series of events. Students "treat wave pulses as cohesive objects rather than as extended propagating disturbances of the medium" (Wittmann, 2001). Wittmann calls this reasoning strategy the application of the object coordination class to wave pulses. In another paper (Wittmann et al., 2002) this group of authors reports on their investigation of how students distinguish between the motion of the wave and the medium through which it travels. The researchers posed two different questions related to sound waves. Some students described the motion of a dust particle sitting motionlessly in front of a previously silent loudspeaker after the speaker was turned on. Others described the motion of a candle flame placed in front of the loudspeaker. More than 25 students were interviewed and over 200 answered the questions in a pretest. 137 students answered these questions in a test that was administered six weeks after students completed the instructions on waves. The findings were as follows:

➢ While describing the motion of a candle flame and the motion of a dust particle, students generally had the same difficulties.

➢ Although students think about waves both in terms of objects and a series of events, they primarily focus on the object-like properties of the system.

➢ The great difficulty for most of the students is distinguishing between the propagation of the sound wave and the motion of the medium through which sound travels.

➢ Traditional lecture instruction with the associated homework problems had little effect on student understanding.

➢ A tutorial that researchers designed to address this topic showed better results than a traditional lecture but the gain was not as large as desired.

➢ The students seemed to think about sound waves as exerting a force on the medium through which they travel in the direction of sound propagation.

"In summary, the students:

> ➢ Map object-like properties onto sound waves,
> ➢ Treat them as solid and pushing through a medium, and
> ➢ Do not correctly interpret the event-like properties that are more appropriate in this setting" (Wittmann et al., 2002).

It is significant for this study that (based primarily on Wittmann's work) the Physics Education Group at University of Maryland assembled a "Wave diagnostic test" which addresses students' difficulties in understanding of waves and can be found online (University of Maryland, 1999). However, this test does not address their mental models of sound propagation, which is the purpose of the test we developed in this study.

2.8 Identified models of sound propagation

Most of the research studies mentioned thus far had several disadvantages for the purpose of the creation of the Model Analysis inventory. In many of the studies the researchers were primarily interested in whether or not the student applies a correct model. Consequently there is a lack of analysis of incorrect answers and alternative models. Researchers also do not seem to consider the context in which the model was presented i.e. the question: Do students apply the correct model in some situations and incorrect in others?

These issues were addressed in our earlier study (Hrepic, 2002; Hrepic et al., 2002). This previous research was specifically designed as an introductory study for construction of a model analysis inventory related to sound propagation. The aim of this study was to address the following questions through in-depth interviews.

> ➢ What models of sound propagation can be drawn from students' reasoning?
> ➢ How do these models depend on context?
> ➢ Do students' mental models change after the instruction?

Researchers used a semi-structured protocol to interview 16 students enrolled in a conceptual physics class, before and after the instruction.

In order to probe for context dependence of mental models our interview protocol consisted of five broad contextual settings:

Context 1: Propagation of a human voice through air and its impact on air particles.

Context 2: Propagation of a human voice and its impact on a dust particle in the air.

Context 3: Propagation of a constant sound and a rhythmic, beating sound from a loud speaker and the impact of these sounds on a dust particle in the air.

Context 4: Propagation of a human voice through the wall at a macroscopic and microscopic levels and the impact of this sound on wall particles.

Context 5: We performed an experiment with propagation of sound through a tight string with cans attached to its ends. We compared propagation of human voice through the tight string vs. air and through the tight string vs. the loose string.

Because this study was specifically aimed at addressing students' mental models of sound propagation, as a review of the findings of this study, I list and define models of sound propagation as they were identified. In this study the students' mental models were determined in two ways:

1. Through the definitions that authors constructed from students' descriptions of sound propagation.

2. Through the sound properties in cases where the authors recognized some of them as uniquely associated with the respective model.

Using the criteria listed above we identified a dominant alternative model that we call the "Entity" model.

2.8.1 Entity Model

According to the "Entity" model, sound is a self-standing independent entity different from the medium through which it propagates. Along with a model definition, as we established identification of an Entity Model, four sound properties were identified that authors considered uniquely associated with that model. These are:

➢ Sound is independent – sound propagates through the vacuum (does not need a medium).

➢ Sound is material - sound is a material unit (of substance) or has mass.

➢ Sound passes through the empty spaces between the medium particles (seeping).

➢ Sound is propagation of sound particles that are different from medium particles.

In addition to model definition and model defining properties, authors also define properties of sound that are inconsistent with each of the models. Properties of sound that are consistent with a model together with those that are not consistent enable unambiguous defining of hybrid models. In this overview we will omit aspects of models that are incompatible with respective models and mention only their definitions and model defining features.

2.8.2 Wave Model

The Wave Model is the scientifically accepted model. Operational definitions of the Wave Model were:

a) Sound is a traveling disturbance of medium particles.

b) Sound is the (longitudinal) vibration of medium particles.

2.8.3 Hybrid models

A common feature of all models that were identified in this study besides entity and Wave Model is that they unify some characteristics of each of these models and form a new composite model. At the same time, by one or more features, these compound models are inconsistent with both the Entity and Wave model. We call this class of composite models hybrid models.

2.8.3.1 Shaking Model

Definition: Sound is a self-standing entity different from the medium. When it propagates through the medium it causes vibration of the particles of the medium (air particles, wall particles) and particles in the medium (dust particles). These particles of/in the medium vibrate on the spot.

Besides the model definition, the following combination of sound properties is uniquely associated with the Shaking Model:

➢ Sound is intrusive i.e. particles of/in the medium vibrate, and

➢ any sound property uniquely associated with the Entity Model.

2.8.3.2 Longitudinally Shaking Model

The Longitudinally Shaking Model is a special case of the Shaking Model as the type of vibration here is specified as the longitudinal vibration.

Definition: Sound is a self-standing entity different from the medium. When it propagates through the medium it causes longitudinal vibration of the particles of the medium (air, wall particles) and particles in the medium (dust particles). These particles of/in the medium vibrate longitudinally on the spot.

Besides the definition, a particular combination of sound properties identifies the model.

➢ Sound is intrusive, i.e. (particles of/in the medium) vibrate longitudinally and

➢ any sound property uniquely associated with the Entity Model.

2.8.3.3 Propagating Air Model

Definition: Sound propagates so that the air particles travel from the source to the listener.

This definition is at the same time the only identifying property of Propagating Air Model.

2.8.3.4 Vibrating Air Model

Definition: Sound is an entity different from the medium that propagates through the air, which constantly vibrates back and forth horizontally. Vibration of the air particles is identical with and without sound. When the source produces sound, this motion of medium molecules transfers the sound forward.

The sound property is uniquely associated with a Vibrating Air Model: Sound is transferred by vibration of medium molecules that vibrate longitudinally with and without sound.

2.8.3.5 Ether and Compression Model

Definition: Sound is propagation of the disturbance created by a longitudinal vibration of etheric particles that are different from particles of any physical medium. These etheric particles are called sound, sound waves or sound particles. To propagate, sound needs compressions and rarefactions of the physical medium through which it propagates. However, compressions and rarefactions always exist in the medium regardless of sound propagation and sound itself has nothing to do with their formation.

Other implications of this model (as stated by a student who described it): Sound is carried by the compressions and rarefactions of the physical medium and it also travels through them. The speed of sound is different from the speed of compressions and rarefactions and sound is not the cause of their creation or movement. Compressions and rarefactions do not move relative to each other but alongside, staying in a phase.

Although sound does not create the compressions and rarefactions in the air, the air is always arranged so that it has some more or less dense spots that will serve the purpose and transmit the sound. Solids that sound encounters serve as compressions – spots of higher density. Sound travels faster through compressions than through rarefactions so it travels faster through solids than through gases. But, compressions in air (gasses) can move and fixed solid objects are static compressions. Also, sound diminishes faster while traveling through static compressions (of solids) than through moving compressions (of gases). This explains why sound goes faster though the wall than through air and yet it diminishes faster while going through the wall.

Besides the definition, the following combination of properties is uniquely associated with the Ether and compression model:

➢ Sound moves back and forth, and

➢ sound travels through (the air compressions and rarefactions).

All of these models also fulfill diSessa's (2002) requirements for a mental model: They have (1) spatial configuration of identifiable kinds of things, (2) (few) principles of how system works and (3) (certain) predictive power.

In addition to these models there was one model that was identified as consistent with Greca and Moriera's (2002) definition of a mental model but which did not fulfill more restrictive requirements of diSessa's definition mentioned earlier. However,

features of this particular model are a subset of features of the Ether and Compression Model. Therefore, it has no additional features that we might be concerned about while developing the model inventory.

2.8.4 Ear-born sound

In addition to models of sound propagation, there is a specific understanding of what the sound is that may be associated with different models of propagation. This understanding is that the sound is what we hear i.e. exclusively what we hear. The dilemma of whether or not there would be a sound if a tree falls down in the middle of a jungle where there is noone to hear it is well known and can be even found as a textbook "problem" (e.g. (Hewitt, 2002)).

In a study that aimed at eliciting students' models of sound propagation, several students expressed ideas that may be interpreted along the lines of "Ear-born" sound. In this earlier study, the research questions were focused on sound propagation while sound production and reception were omitted. For this reason if students made statements about sound perception they were not followed up and on their own they may be interpreted in different ways. For example it is not clear if the two students whose statements presented below believe sound is exclusively something in the ear or that what we hear is only one aspect of what we call sound.

Example 1:

S: [When the speaker speaks] vibrations…the force of vibrating air molecules...it's going across [the air] and then vibrates…the air and the ear. Then [it] vibrates through the inner ear, then it'll make sound, she'll recognize that as sound.

Example 2:

I: What is a sound wave? How do you perceive it?

S: (Pause)…Umm, I am not sure like if you would mean like…like definition of sound wave or (laughs) whatever?

I: Yeah, if you want…

S: It's just what you hear, it's like…

I: If you want to say a definition, that's all right.

S: It's kind of like the radio wave. It's the signal and so the sound wave would be the signal that she is hearing.

Although the issue of the ear-born sound has been raised to the level of a textbook anecdote, in earlier studies on students understanding of sound it was not investigated enough. So, while developing the test in the present study, researchers "gave a shot" to the ear-born idea to determine if it has an audience. And it turned out it does, especially at lower educational levels.

2.8.5 Dynamics of medium particles

Authors of the study on the mental models of sound propagation (Hrepic, 2002) also investigated students' ideas related to the dynamics of particles of the medium that occur while sound propagates. A variety of different movements were identified - from no movement to various vibrations to random travel in any direction.

In many cases students believed that while sound propagates particles of the medium undergo composite motion. This composite motion often consisted of vibration along the particular direction and traveling of the particles in a certain direction (usually away from the source) at the same time. Before instruction, students most frequent belief was that particles of the medium travel away from the source. After the instruction the most frequent answer was longitudinal vibration of the particles, but in both cases students also expressed a variety of other movements.

2.9 Conclusions and implications of previous research for this study

According to constructivists' theory, people ultimately construct their own knowledge (Asynchronous Learning Networks, 1998) and organize their experiences into mental models (Redish, 1994). "A mental model is an internal representation, which serves as a structural analogue of situations or processes and enables individual to explain and predict the physical world behavior" (Greca & Moreira, 2002, p.108). Bao (1999) developed the tool for analyzing students' mental models called model analysis, which

can be employed after common students' models in a particular physics topic are identified and model analysis inventories are constructed.

Sound is a topic in which students exhibit a series of difficulties. (Barman et al., 1996; Beaty, 2000; Hrepic, 1998; Hrepic et al., 2002; Linder, 1987, 1992, 1993; Linder & Erickson, 1989; Maurines, 1992, 1993; Merino, 1998a, 1998b; Wittmann, 1998; Wittmann et al., 1999; Wittmann et al., 2002) The lack of understanding of sound and its propagation are not appropriately addressed by science educators and researchers (Linder, 1993; Maurines, 1993; Merino, 1998a).

One of the first prerequisites for creation of the model analysis inventory is to determine the prevailing mental models in a targeted student population and to describe those models qualitatively. A crucial step toward meeting this goal was our earlier research (Hrepic, 2002; Hrepic et al., 2002). We have found that in order to describe sound propagation students use two models that are fundamentally different -- the Wave Model and the Entity Model. All other identified models are hybrid models that share some but not all of the features of each of the fundamental models.

CHAPTER III
METHODOLOGY

3.1 Introduction

The purpose of this research is to create a test that will elicit students' mental models of sound propagation during a class. In the preceding chapter those models were described as identified in earlier studies. In this study we developed and validated a test that probes for these models. For this purpose we used both qualitative and quantitative research methods. This chapter presents the nature of research methods and details of the research design employed in the study.

3.2 Research methods

To address our research questions we employed both qualitative and quantitative methods. Both methodological traditions add to the body of knowledge but from different aspects. They differ in the kinds of research questions they investigate as well as in the goals, methods, procedures and description of findings. However, they share some basic features such as concern with logical plausibility and rigor in the research design.

Qualitative research "focuses on the experiences interpretations, impressions or motivations of an individual or individuals and seeks to describe how people view things and why" (CIREM, 2002). Qualitative researchers describe occurrences and objects of interest narratively, but they also give the range and frequency of observed perspectives (Erickson, 1998). Krathwohl (1998) considers qualitative methods "particularly useful in understanding how individuals understand their world, in showing how individuals' perceptions and intentions in situations determine their behavior, in exploring phenomena to find explanations and in providing concrete and detailed illustrations of phenomena" (Krathwohl, 1998, p.225). Qualitative research methods utilize a variety of data collection techniques. Those most frequently used are interviews, observations, documents and audiovisual materials (Creswell, 1994).

Quantitative research on the other hand "focuses on measuring and counting facts and the relationships among variables and seeks to describe observations through statistical analysis of data. It includes experimental and non-experimental research and descriptive research (research that attempts to describe the characteristics of a sample or population)" (CIREM, 2002). Quantitative research, unlike the qualitative, describes phenomena in numbers and measures instead of words (Krathwohl, 1998).

While contrasting the two approaches Krathwohl (1998) stresses the following differences:

➢ The explanation guides the development of the quantitative study and in the case of qualitative study, the explanation grows out of its data.

➢ In the case of quantitative study, creative work precedes the data collection while in quantitative study it occurs after the data collection.

➢ The aim of the quantitative researcher is to investigate and describe an objective reality that is "out there," and a qualitative researcher aims to describe the reality as perceived by each individual.

➢ Quantitative studies describe behaviors against measurement scales and qualitative studies describe behaviors verbally.

Both qualitative and quantitative methods have their respective strengths and weaknesses and researchers often combine methods in order to investigate a problem from different aspects. In our study we needed both methodologies in order to validate the test we developed. We needed qualitative methods to probe in depth how students understand and perceive the test questions and the answer choices. For this purpose we used semi-structured interviews with open-ended questions. Interviews were tape-recorded (audio only). The disadvantage of interviewing (as well as the other qualitative methods) is the limited number of students that can be economically studied. Another disadvantage is that the interview setting differs from a natural settings to some extent in which this test will be used (classroom setting or any setting in which students do their homework). So data collection on a large sample and in the setting in which the test will later be used was another crucial component of the test validation.

3.3 Validity and reliability verification of the testing instrument

The most important purpose of educational measurement (test, inventory, questionnaire, etc.) is to help in decision making. This is true not only for testing in education but also in psychology and any other fields of human endeavor. Moreover, the test should not be even administered unless its results will be used to improve decision making (Hanna, 1993). Our aim is to create a test that will facilitate the teaching process in a way that it will inform a teacher what the students' problems are so she or he can accordingly adjust the lecture. Therefore, the primary purpose of the test developed in this study is formative assessment and only with reservations that will be explained letter (Section 5.2.1), can it also be used for the summative purposes. In either case, in order to be a useful instrument that enables appropriate decisions, a test has to be valid and reliable.

3.3.1 Test validity

Validity pertains to the extent to which a test measures what it is supposed to measure and nothing else (Oosterhof, 2001). "Moreover, validity concerns the appropriateness of inferences and actions that are based on a test's scores" (Hanna, 1993, p. 8). Validity is not an attribute of the test, but "of the interaction of a test with a situation in which the test is used to make decisions" (Hanna, 1993 p. 382).

Validity is considered a unitary concept with three aspects so we distinguish content-related, criterion-related and construct-related validity evidences (Hanna, 1993). While validating this test against these different aspects of validity we need to keep in mind the object of the testing. We want to probe what mental models of sound propagation students use (among those most commonly observed), how frequently and consistently they use them and how students' use of mental models depends on a context.

3.3.1.1 Content-related evidence of validity

Content-related evidence of validity indicates "the degree to which the sample of test questions or tasks is representative of the content domain of interest" (Hanna, 1993). It indicates how well the content of a test corresponds to the student performance that we

want to observe (Oosterhof, 2001). Content-related validity is particularly important for achievement tests and it is judgmental, not quantitative. One way we demonstrated content validity is through the "table of specifications" (e.g. Oosterhof, 2001) that lists the things we want to measure and how the test addresses those items. Another way in which we demonstrated content-related validity was through experts' reviews of the content and correctness of the answer choices. The experts' review also demonstrated also the face validity of the test. Face validity is a component of content validity that refers to what a test superficially appears to measure. So, although face validity is not validity in a true sense, it is nonetheless an important test attribute because if the test appears irrelevant or silly, the result can be poor cooperation and weak rapport with test takers and the public concerned with the test regardless of the actual validity of the test (Hanna, 1993).

3.3.1.2 Criterion-related evidence of validity

"Criterion-related evidence of validity indicates how well performance on a test correlates with performance on relevant criterion measures external to the test" (Oosterhof, 2001, p.55) where the criterion is the variable of primary interest. Hanna (1993) distinguishes two aspects of criterion-related evidence of validity: predictive validation (that compares the scores on the original test with scores on one or more criterion measures obtained in a follow-up testing) and concurrent validation (that compares the test results with results obtained through a parallel, substitute measure). We demonstrated criterion-related evidence of validity primarily through its aspect of concurrent validity. For this purpose we employed think aloud interview protocols and compared students' free responses with their answers on the test. Predictive validity, another aspect of criterion-related aspect of validity was more difficult to measure because we are the first to develop the test for this particular purpose so there is no other equivalent measure that we can use at this point on a larger sample to determine its validity in a predictive manner. This is not a critical issue because we do not propose this test as a predictive tool like the SAT or similar tests. The primary purpose of the test we developed is to serve as a formative assessment and its secondary purpose is to serve as a summative assessment tool. Predictive validity is tightly related to the measurement of

so called occasion sampling errors, which pertains to the reliability of the test. For this reason these two concepts will be further addressed later in the section on reliability.

3.3.1.3 Construct-related evidence of validity

Construct-related evidence of validity "establishes a link between the underlying psychological construct we wish to measure and the visible performance we choose to observe" (Oosterhof, 2001, p.46). "Construct validation consists of building a strong logical case based on circumstantial evidence that a test measures the construct it is intended to measure (Hanna, 1993 p.402). Kinds of evidence that are acceptable span wide range of psychomotor, affective and cognitive domains. The psychological construct that we measure is knowledge. Knowledge is divided into declarative (information) and procedural (concepts, rules and complex skills) (Oosterhof, 2001). In this categorization, the knowledge that we are testing corresponds to the concepts. More specifically it is a kind of a conceptual knowledge called a mental model as defined earlier. Therefore, we will primarily use findings of the previous research to document that we are assessing mental models. Further, according to Oosterhof (2001) concepts involve more than one characteristic that classify physical objects or abstractions. And "because a concept involves a class of things, it should be assessed under a variety of conditions" (Oosterhof, 2001, p.32). In the case of our test this translates into a need to assess different aspects of the concept that is scientifically called a Longitudinal mechanical wave and also to assess these aspects in different contexts. We will use previous research findings and a construct-related adaptation of the "table of specifications" to make the evidence of the construct validity of this test.

There is an important point to note here. We can ask two different questions related to the knowledge we actually test i.e. mental models of sound propagation. These questions are (1) what kind of knowledge is this and (2) what is the content of this knowledge? The answer to the first question is conceptual knowledge or a "mental model" and this refers to an aspect of construct-related validity. The answer to the second question is "sound propagation" which refers to aspect of content-related validity. This point illustrates that different aspects of validity are interrelated and they overlap. So, although these rubrics are convenient, their use does not imply that distinct types of

validity exist or that any of these three aspects is sufficient to determine the test validity (Hanna, 1993). "Recent scholarly thinking has stressed that validity is a unitary concept" (Hanna, 1993, p.380).

Listed below is a brief overview of the types of validity verifications that we performed for our test. Here we classify them as belonging primarily to one of the three groups although, as we stated, this classification has weak boundaries:

- Primarily content-related validity verifications
  - Table of (content) specifications
  - Experts' review of the content and correctness of the answer choices
  - Instructional sensitivity
- Primarily criterion-related validity verifications
  - Correlation analysis of answer choices
  - Validation through the interviews
    - Think aloud interview protocols
    - Comparisons of students free answers in interview setting with their results on the test
  - Role-playing validation
- Primarily construct-related validity verifications
  - Built on previous research
  - Table of (construct) specifications

3.3.2 Test reliability

Reliability pertains to the degree to which a test consistently measures what it is supposed to measure (Oosterhof, 2001). Reliability is not related to truthfulness but only to consistency of scores (Hanna, 1993). To further explain the concept of reliability I will use an analogy that will be useful later. To measure the electric current we use an instrument called an ammeter. If we have constant current (in magnitude and direction) in our circuit, our ammeter will demonstrate its reliability by showing the same value every time we perform the measurement.

One of two main approaches to the reliability focuses on inter-individual variability i.e. the consistency with which individuals maintain their position in the group

(Hanna, 1993). In this approach, reliability is assessed through different correlations that reflect various consistencies. Another perspective "is based on intra-individual variability or the consistency among repeated measures of the same person". (Hanna, 1993, p.434) While making the case for the reliability of our test, we will use some aspects of both views.

Major sources of measurement error which compromise reliability are:

1. Content sampling error
2. Occasion sampling error
3. Examiner error
4. Scorer error

3.3.2.1 Content sampling error

Content sampling error occurs because students may or may not be lucky with how the test items correspond to things they know. As a consequence, some of them are overrated by test results and some are underrated. A way to reduce content sampling error is to test more of the content. For this reason three to five item tests may be highly unreliable and educators should be cautious about commercial instruments that are claimed to be conveniently short yet highly reliable at the same time (Hanna, 1993). To measure content sampling error one needs two parallel forms of the test that are administered to the same examinee. These two sets of scores are then correlated (Hanna, 1993).

Measuring content sampling error for this test was not a straightforward procedure. This test is one-of-a-kind. We could not administer parallel forms to large number of students and perform any statistically relevant comparisons. So, several issues need to be addressed here. The small number of items (six questions) in the test is not one of the problems but actually strengthens the test. As stated earlier, we reduce content sampling error by increasing the number of test items so that we cover as much of the content as possible. This test has only six questions but unlike in most tests, all of the questions are addressing a single concept. The question/topic that the test deals with is: What are students' mental models concerning the mechanism for the sound propagation?

With this test we want to elicit students' mental models states, i.e. we want to determine what model(s) students use and how frequently and consistently they use them. To probe this consistency we need multiple probes. The test covers the same target "topic" six times from various aspects. Thus, a "small" number of questions is not a problem.

A more complex issue related to content sampling error and possibly pertaining to our test is that based on the previous research we can expect many students will not use only one model to describe the sound phenomena. We can expect that a number of students might not select a set of answer choices (in different questions) that we would consider corresponding to the same model because more than one model might appeal to them. Above all, this test is supposed to be used primarily before and during the instruction so it is natural to expect that students in this stage of learning will not have a firm or developed model. This situation will yield to inconsistency in their answers with respect to different models of sound propagation. Thus, we can not use different questions of this test as each other's control for reliability in a straightforward way. In addition, we can not use two tests that pertain to different contexts to serve as alternatives to each other. This is because based on the previous research we do not expect that students will use the same set of models and with the same frequency in two different contexts.

So, the question is: how do we know if students' inconsistencies arise because of their mixed model states or because of the lack of reliability of the test? In addressing this question we combined two procedures. The first one is a correlation analysis of all combinations of the answer choices in the test and from the large sample. The rationale for this approach is based on the assumption that although many students might not be self consistent, if the test is reliable (and if data are obtained from a large and diverse sample) correlation coefficients between all pairs of answer choices that correspond to the same model should be positive. In addition there should be no significant positive correlations between the choices that do not correspond to the same model.

The second procedure that we employed to distinguish a mixed model state from the test unreliability is validation through an interview protocol with a small sample of students. In these interviews we compared students' answers on the test with their

answers to open-ended interview questions. After that we also asked them to determine which of the pictorial representationsof the models of sound propagation, if any, best corresponds to the model they described. We constructed these pictorial representations to represent each of the four fundamental mechanisms of propagation that we identified (Section 4.3). These open-ended questions and pictorial representations served as alternate forms of test.

3.3.2.2 Occasion sampling error

Occasion sampling error occurs because students can be more or less lucky with respect to the time when the test was administered. This error is not an issue with motivated students who work consistently, but it may be significant with students who are in the habit of procrastinating and cramming right before a test. A way to reduce occasion sampling error is to test on more occasions. Occasion sampling error is measured so that a sample of examinees is given the same form of the test on two occasions and these two sets of scores are then correlated (Hanna, 1993). Economical restrictions did not allow us to probe occasion sampling error in this direct manner so we did it indirectly in several ways.

1. We took data from the same kinds of classes at different educational institutions and we expected that (if occasion sampling error is negligible) the overall pattern of results in each of these instances is similar.

2. We took data from students enrolled at different course levels at the same university. In these cases we expected that students in more demanding course levels would perform to some extent better than those at lower level courses.
   Note: Observing the difference in an expected way would contribute to the case that the test is resistive to occasion sampling error. Observing a result opposite of the expected one would raise questions although that result would not clearly indicate the problem.

3. Finally since we were administering the tests before and after the lecture we could not expect the same results in these two instances, but we do expect a definite direction in the shift of the answers. This shift should occur toward (1) the

scientifically accepted side of the model spectrum, (2) the increase in self-consistency and (3) the bigger percentage of correct answers in absolute terms.

Note 1: Item No. 3 does not necessarily correspond to the positive model shift in No. 1)

Note 2: The same note mentioned in the point 2 applies here as well.

The reason we were not able to administer the test multiple times in a row to determine this error directly was due to the economics of class time. We needed multiple testing primarily to determine the pre-post instruction dynamics that pertains to the criterion-related validity of the test. So, in cases where the instructors agreed to administer the test we would ask if they could do it twice. But, because each testing interrupted the instructors' schedule and took some of the class time, not all of them were able to do it twice. Asking for three administrations of the test was not feasible. In those situations in which we could do it twice we used the opportunity to determine pre- and post-instruction differences in various instances. However, the pre- and post-instruction data was only indirectly usable to probe for occasion sampling error because due to the instruction, the sample was not the same in these two instances. We expected the test results to be different but in a definite way (i.e. better after the instruction).

The occasion sampling error was therefore addressed indirectly, in the aforementioned ways and because of the reasons described.

3.3.2.3 Examiner error

Examiner error occurs because of the differences in examiners. Examiner error is not a very worrisome source of unreliability when examiners follow standardized instructions conscientiously. Therefore, the primary way to minimize examiner error is to follow the administration instructions to the letter. Examiner error can be estimated by recording and then critiquing the examiner's performance. However, no direct research procedure would isolate the contribution of examiner error to the test (un)reliability (Hanna, 1993).

A possible threat to the examiner error was that we could not control that part of the students' motivation to take the test that is related to their interaction and rapport with the examiner. The rapport level between students, their teacher and the test administrator (if not the teacher) might have differed in several instances. Another source of

differences in motivation related to the examiners was that some students were offered extra credit for correct answers in the test and others were not. This is something we could not control, but we had to let to each of the instructors decide for his/her class. A primary way of reducing the examiner error was writing identical instructions for students on the top of the each test (although there is no guarantee that all of the students read the instruction). Another way in which we reduced this error was by collecting data from a variety of sources (different institutions, instructors, class levels) in various ways (in class and online) with various incentives for students and by taking data from the same educational levels all together in order to check the correlations relevant for reliability. This way we averaged out possible different contributions to examiner error.

In addition, it is not likely that our inability to fully control these threats caused our results to be different in any way from the results that teachers who use the test in the future will be obtaining. This is because thes exact same differences between examiners and students that existed between our samples will also exist when the test is applied in classrooms.

3.3.2.4 Scorer error

Scorer error occurs if students' scores depend on the individual who happened to review their work. Normally in order to reduce this error multiple graders are used and the error is measured by comparisons of their results (Hanna, 1993). This source of error is not an issue with multiple-choice tests in general.

As written earlier, reliability is about consistency. If we observe expected kinds of similarities and differences, together they will make a solid case for the test reliability.

3.3.2.5 Reliability as a precondition for validity

A test can not be valid unless it is reliable because a student's performance at some point in time is not measured validly if the measurement data is inconsistent. Going back to our analogy, given that our electric current is constant, the ammeter has to always show the same value. This consistency proves it to be reliable. This condition is necessary, but not sufficient to show it is valid. We still do not know if this measurement represents the amount of current that we want to measure. If the circuit is a simple with two parallel

resistors, the ammeter could be connected in the wrong branch of the circuit. In that case, if it always shows the same value, the measurement is reliable but not valid. Also, we may read the values from the wrong scale, which will again yield reliable but invalid results. For this reason we performed all the procedures described in the sections on validity and reliability in order to establish that the test is valid and applicable for the purpose for which it is intended.

3.4 Procedure

The target population for this test is primarily college students, but this does not exclude the possibility that the test is applicable at lower educational levels as well. We did develop the test primarily on the basis of previous research on college students' understanding of sound propagation (Hrepic et al., 2002; Hrepic, Zollman, & Rebello, 2003; Linder, 1987, 1993; Linder & Erickson, 1989; Merino, 1998a, 1998b; Wittmann, 1998; Wittmann et al., 2002). However "results from research on student understanding in physics indicate that certain incorrect ideas about the physical world are common among students of a wide variety of national backgrounds, educational levels and ages. There is considerable evidence that university students often have many of the same conceptual and reasoning difficulties that are common among younger students" (McDermott, 1998). In particular, several studies showed this is the case in the area of students' understanding of sound (Barman et al., 1996; Hrepic, 1998, 1999; Maurines, 1992, 1993). These studies showed that most of the difficulties that students have in understanding of sound at the college level are very similar or the same as those that exist at the pre-college level. These findings have two important implications for our study:

1) We can expect that content-wise our test will be applicable to pre-college educational levels as well. Because of this reason, although our validation process revolved primarily around the college level, we probed whether the test is applicable at the high school and middle school levels as well.

2) To add to the diversity of the sample we can include in the analysis the results obtained at the corresponding educational levels from the test sites not only in US but abroad too. Therefore, to add to the diversity of the sample (and therefore to the

50

external validity of the findings) we collected data at all possible sites (nationally and internationally) where instructors agreed to administer the test.

We accomplished our research goal through several steps of test development and validation verifications that consisted of individual interviews with 30 students enrolled in different introductory courses at KSU along with several rounds of surveys on more than 2,000 students at different schools and colleges in the USA and Croatia. In the section below I will describe the major steps of the test development, their rationale, the main procedures and the participants involved in these different stages. However, details of each of these procedures will be described in Chapter IV along with the data analysis. Our findings in each of the steps dictated that the step that followed and the rationale for the subsequent steps could be explained only on the basis of the findings related to the step that preceded it.

We divided the research procedure into four major steps that we labeled pilot testing, pre-survey testing, survey testing and post-survey testing. Each of these steps is described in the following sections.

3.4.1 Pilot testing

3.4.1.1 Rationale

Pilot testing was a bridge between qualitative data on students' mental models of sound propagation and the multiple-choice test that we were developing. The qualitative study (Hrepic, 2002) that probed these models had two disadvantages when the construction of a multiple-choice questionnaire was concerned. First, the number of participants was limited (23). This disadvantage is not specific to the study, rather it is associated with the qualitative studies in general. Second, although the set of interview questions in this research (Hrepic, 2002) covered five broad contexts and although the questions were specifically targeted toward models of propagation, some other situations or questions might also be useful in eliciting students' models of propagation. This possibility was certainly worth probing, but we did not expect to discover any new models because of the battery of tests with a broad range of questions related to sound phenomena were used in

earlier studies (e.g. (Hrepic, 1999; Maurines, 1993)). Based on these studies, contexts and questions were selected for the study on mental models of sound propagation (Hrepic, 2002) in the first place. However, the matter was worth addressing as a part of our test construction.

The purpose of pilot testing was to address the aforementioned issues by administering a larger number of questions to a larger number of students than was possible through the interviews in the earlier phenomenographic study (Hrepic, 2002). Tests that we used for this purpose in the stage of pilot testing had open-ended and semi-open-ended questions.

3.4.1.2 Description and participants

We administered two tests as a part of the pilot testing. The first one had open-ended questions that were similar to part of the earlier interview protocol (Hrepic, 2002). The test was given to students enrolled in a concept-based introductory physics course at Kansas State University (from now on KSU). The instructor agreed to administer the test and students were receiving extra credit not only for their correct answers but for any answers accompanied with a logical explanation. This incentive caused students to accompany their answers with explanations which was the primary interest.

Another pilot-stage test was a semi-structured test that covered a broad range of situations and concepts associated with the propagation of sound. We provided some structure to the answers because we were not interested in the correctness of the answers but rather what the students' rationale was for them. So, to avoid simplistic "yes-no" or "more-less-equal" types of answers, we offered those as choices in multiple options and then required explanations of their choices. This test was administered to students in another large enrollment (139 students) concept-based introductory physics course at KSU a year after the open-ended version.

In this study no student ever refused to take a test when it was administered in the classroom whether or not they took it for extra credit. So if there is some difference between students that come to lectures and students enrolled in class as a whole, then our sample is more representative of that portion of students who do attend the class. This is

not a disadvantage (if a difference exists) because the test will later be used with those students that attend the classes.

3.4.2 Pre-survey testing

3.4.2.1 Rationale

Based on the results of earlier studies and from our pilot testing, we chose three contextual situations and a set of five test questions that we considered optimal for the purpose of eliciting students' mental models of sound propagation. The test questions were related to each of the two contextual combinations (Air-Vacuum and Wall-Vacuum). Answer choices were made so they corresponded to the identified models. Based on pre-survey results, the five initial questions were later expanded to six).

Before finalizing sets of multiple-choices for each of the questions we verified whether the offered choices are all that (the majority of) students would want to pick. This was investigated so that all questions in the first version of the multiple-choice test contained an additional option that allowed a student to write in an answer on his/her own. It was also possible that some students prefered more than one of the choices that we offered. To investigate this possibility, all questions contained also a choice that allowed for choosing more than one of the offered answers.

In addition to administering surveys, at this stage of the research we made an initial validation of the test through the students' interviews and through experts' review of the test. The purpose of interviewing students was to verify whwther their understanding of the questions and answers corresponded to their intended meanings.

3.4.2.2 Description and participants

During this stage of research we were changing the test as we gathered more data and feedback from the students. We also interchangeably used interviews and surveys at a larger scale (N>30 each time) to see if our improvements were making an impact on the test results. The particularities and rationale for these changes will be described in detail in Chapter IV because the rationale for the changes is meaningful only if accompanied by findings that preceded and dictated those changes. Students who participated in the study

at this stage were enrolled in all three levels of introductory courses at KSU (concept-based, algebra-based and calculus-based).

Experts that validated the test were members of the physics faculty at KSU who did not participate in this project. They were asked to determine if:

1. ... in the correct answer choices there is nothing that is incompatible with the Longitudinal Mechanical Wave Model (at the introductory physics level) although the answer may not be developed enough to tell the entire story.

2. ... in the incorrect answer choices there is at least something that is clearly incompatible with the Longitudinal Mechanical Wave Model?

3. ... the questions and answers are clearly formulated.

They were also asked to provide any other suggestions they felt might improve the test.

3.4.3 Survey

3.4.3.1 Rationale

After the detailed preparation of the test questions, answer choices and their wording, we administered the test to a large number of students (>2000) at 13 different educational institutions in the US and Croatia. The purpose of this was to determine whether the answer choices are correlated in a meaningful way when data is collected from a large number of students in a variety of institutions and educational settings. Together with the second round of validation through interviews, this procedure was a crucial piece in determining whether students use multiple models in the test because they really are in the mixed model state or if this happens because the test was not reliable. The rationale for this procedure was that a particular student who is not firm about his/her model may select choices that correspond to different models in different questions. However, if answers related to the same models have significant correlations these correlations add weight to making a case that the test is reliable. To show the validity of the test we coupled correlation analysis of the survey data with another round of test validation through the interviews.

Another purpose of the survey was to determine the test's instructional sensitivity (whether the test indicates differences in students' knowledge after they learned about the

topic) and make a judgmental (non-quantitative) estimation of the effect. At this stage of the research we also wanted to determine whwther the test recognizes similarities or patterns that are expected at different educational levels (middle, high school and college level) and if it recognizes differences that are expected at different levels of introductory physics at the same institutions. Finally, we wanted to see if patterns and similarities exist when similar courses at different institutions are compared. This comparison addresses the issue of the data generalizability.

3.4.3.2 Description and participants

At this stage of our research we administered the same test to as many students as possible and at as many levels as possible. We distinguish educational levels (middle school, high school and college level) and levels of collegiate introductory physics courses (concept-based, algebra-based and calculus-based introductory physics). Our sample selection was not really a selective process because we administered the test at all of the sites where the instructors agreed to participate. We sent out requests to physics instructors all over the US through an e-mail list with subscribers interested in issues related to physics education. In addition we sent the same request to physics instructors that the author knew in his native country of Croatia and several of them administered the test in their classes. Therefore, the test was administered in all classes whose instructors agreed to participate in the study regardless of the character and level of the institution or the class size and its instructional setting.

Interview validation, which is another equally important segment of this research stage, was done by interviewing KSU students enrolled at three different levels of introductory physics courses.

3.4.4 Post - survey modifications

3.4.4.1 Rationale

In the survey phase we established a case that the test is valid and can serve the purpose for which it is intended. Results at the college level indicated this assumption was clearly true. However the large scale survey pinpointed two problematic answer choices that had

been unnoticed in earlier validation procedures. Validation through the interviews confirmed and explained these problematic items. The next step was to address these issues before finalizing the test.

3.4.4.2 Description and participants

Based on the inputs from the survey part of the study we improved the test yet another time. The changes were primarily cosmetic except in one of the questions (Q6) where a question was modified so that it asks about the pictorially presented situation. After making those changes we did yet another survey procedure to see if the correlations improved in a favorable way in the problematic choices (while not deteriorating in others). We administered the test primarily to KSU students this time and performed the correlation analysis of the answer choices with the same purpose as in the survey phase of our research.

Again, we did an expert validation of this new version of the test. A new group of KSU physics faculty members were asked the same questions as in the pre-survey faculty review.

Finally, to add to our case of the test validity we performed role-playing validation in which participants with a Ph.D. degree in physics played the role of students "having" different mental models of sound propagation. Based on "their" models, participants were supposed to pick the answers in the test that corresponded to their models.

3.5 Research timeline

The study was accomplished in the following timeframe:

1) Pilot testing was finished in the fall semester of 2002. Open-ended testing that was a part of the pilot testing was administered in December of 2001 and the semi-open-ended survey was given in October (pre-instruction) and November (post-instruction) of 2002.

2) Pre survey testing took place from the late fall semester of 2002 to the middle of spring semester 2003. A multiple-choice version of the test with seven options was

administered in December of 2002 as the first item in this stage of research and to students enrolled in a concept-based course at KSU. A five-option multiple-choice version followed and it was administered to students enrolled in different courses at KSU. Interviews were conducted throughout February of 2003 and a final probing version of the test was administered in April of the same year.

3) The survey phase had two major rounds of data collection in the late spring semester and the late fall semester of 2003. The interviewing part of what we call the survey phase was done in the fall semester of 2003.

4) Post-survey improvements took place during the spring semester of 2004.

3.6 Data collection and organization

In the different phases of research data were collected in various ways. The collection method depended primarily on the requirements for the different purposes. When interviews were conducted, the same interviewer interviewed all participants individually. As the interviewer, the author was trying to be not only a concentrated listener, but also an empathetic one and to set aside my presuppositions as suggested by Ashworth and Lucas (2000). To build a rapport with the participants, before the interview I openly explained to each of them the background of the study and how their participation would contribute to the research goal. Using more than three years of teaching experience, I also tried to establish an atmosphere that was not intimidating or judgmental to enable and facilitate the openness in the students' answering. All interviews were audio taped with the permission of each of the participants. Interview protocols were aimed primarily at students' verbal expressions so although the students were encouraged to draw if they wanted to, a video camera was not necessary to record the data. Videotaping would have also recorded the students' facial gestures and body movements, but the intrusiveness of the video camera would not justify its use in this case. Interviews were transcribed verbatim except some segments that were not relevant to the research questions were omitted. In the transcript writing we have followed good interviewing and transcribing methods (APA Publication Manual, 2001; Taber, 2000; Wolcott, 1994).

Multiple-choice surveys were administered in most of the cases as paper and pencil tests. However, in a few cases data were collected via web assignments (as a part of the students' homework). Web-based data collection is one of the proposed uses of this test and thus added to its generalizability.

Unlike in the interviews, multiple-choice tests were administered by a number of different instructors. We were not able to control the level of established rapport with students unless we (researchers) administered the test on our own. However this does not mean that the rapport was necessarily better when the researchers administered the survey than in other cases. It just means that most likely the rapport level was more diverse in different instances during survey administration than during data collection through the interviews. On the positive side, we can assume that on a large sample these contributions averaged out. In addition, we reduced the examiner error related to survey test results by providing the standard introduction at the beginning of the each test.

3.7 Data analysis

With respect to qualitative research traditions, interviews that we conducted could be classified as a set of relatively simple case studies oriented primarily toward within-case analysis. In addition we were looking for patterns in the cross-case analysis. At a smaller scale our analysis was concerned with describing (1) students understanding of sound propagation and (2) students' interaction with the test i.e. interaction of students' understanding of sound propagation and the test. These descriptions were done in the tradition of phenomenographic qualitative analysis in which the researcher is primarily interested in describing the phenomenon as individuals (learners) perceive or experience it (Marton, 1978). While analyzing students' free verbal responses, we ensured that our categories emerged from the data (Marton & Saljo, 1984) and were not imposed on them. Mental models were identified by following the analysis approaches specified earlier in studies on students mental models of sound propagation (Hrepic, 2002; Hrepic et al., 2002).

Survey data were analyzed through a combination of quantitative methods and judgmental (non-quantitative) estimations of the observed effects. For example,

correlation coefficients between answer choices and their statistical significance were both determined through standard (quantitative) statistical procedures. Although the improvement of students' knowledge after instruction is a quantifiable, an estimation of how satisfactory the improvement is belongs in the category of subjective judgment, mainly because it is not only one model whose change we measure, but rather a set of models, each of which has a different (and not measurable) "degree of correctness".

3.8 Ethical considerations

Our study was exempt from a full review by the Institutional Review Board (IRB), but we still decided to comprehensively inform the participants about the research, its background and the purpose. We also asked the participants to sign a consent form. Although no particularly sensitive ethical issues were involved in the study, we have guaranteed the anonymity of all participants (those who were interviewed as well as those who took the survey). The original documents are kept in a safe, locked place and the computer data are on a secure server.

3.9 Limitations and potential biases of the study

We differentiate between the potential biases (possible sources of error for which we were able to compensate to some degree) and the limitations of the study (possible sources of error for which we could not compensate). Potential biases and their resolutions are listed below so that their respective numbers correspond in two lists.

The potential biases of the study are:

1. The sample was not randomly selected.
2. The test did not have an existing counterpart against which we could validate our newly developed instrument.
3. Different instructors were administering the survey and we could not control the rapport established in each of these situations, which might have affected students' motivation.

The corresponding resolutions are:

1. The sample was chosen in such a way that an invitation for participation was sent out to the largest number of instructors objectively possible and all of those who were willing to participate were included in the study. So, although it was not randomly selected, our sample can not be straightforwardly labeled as a sample of convenience either.

2. We validated the test through interview procedures in which we contrasted students' free responses with their answer choices on the test and their selection of the model amongst different pictorial representations of the model. In addition we performed role-playing validation with experts. Thus, this potential bias affected primarily a predictive aspect and not the concurrent aspect of the criterion-related validity. Of these two, concurrent validity is much more relevant for the proposed use of the test.

3. By taking data from a variety of instructors, classes and institutions we can assume that these contributions averaged out. Also, a standard set of instructions was written as an introduction to each of the tests.

Limitations of the study:
1. Not all of the participating instructors had the same number of students.

2. Not all of the participants had the same incentive to take the test and this could have affected their motivation. However, results from students that were taking the test for extra credit are not necessarily more credible than those who were taking it without such an incentive. We observed during our interviews that test taking strategies may distort the way in which students map their reasoning onto question choices. For example, if a students doubts between the two options, he or she may decide to pick one of them in one of the questions and then another in different question to allow for a 50-50 chance of being right in at least one of them (versus risking to pick the incorrect choice in both questions). This strategy is of no use if the test is not taken for extra credit. However, in the latter case we can not be sure if the students were motivated to take the test seriously. A positive aspect of this feature is that in the proposed use of the test the instructor may or may not decide to give extra credit. Therefore, this polarization of our

sample (with respect to the extra credit incentive) did not make the sample different from the target population.

3.10 Reflection on the process

To construct and validate the test that would serve to elicit students' reasoning about sound propagation, we used both quantitative and qualitative methods. These methods were intertwined because the advantages of one of them are in general disadvantageous of the other one and vice versa. By using them in parallel and in numerous steps we were able to create an optimal testing instrument. In several instances data that we collected dictated new routes in the study and new approaches with respect to the test construction and analysis of the results. Although not originally planned, these changes significantly contributed to the test quality and usefulness. The research had several twists of this kind and I would not label it as a straightforward or smooth process. However, we managed to stick with the timeline relatively closely and finish with the product that serves its purpose more than well. Altogether, the research was a thought-provoking, challenging and enjoyable experience.

CHAPTER IV
RESULTS

4.1 Introduction

The process of construction and validation of the test we methodologically divided into four major steps:

1. Pilot testing
2. Pre-survey testing
3. Survey testing
4. Post-survey testing

We gave an overview of these steps in the Chapter III and explained the rationale for each of the research protocols within them. In each of them we had a specific goal and we used different methods to meet the goals. In this chapter we will elaborate further on each of the research steps and show results obtained during those steps.

In presentation of our findings, these steps will be given a different amount of emphasis in order to present results clearly and coherently. As with any other research, this study had its blind roads and bumpy segments. Not all these deserve special attention from the perspective of the final product, so their descriptions will be balanced accordingly.

4.2 Pilot testing

Pilot testing had two purposes. The first one was to determine if in the earlier research anything of significance was omitted in terms of elicited students' mental models. To answer this question we administered an open-ended questionnaire to a large enrollment concept based introductory class.

The second purpose of the pilot testing was to determine the contextual situations optimal for eliciting of students models of sound propagation. Our question was if some contexts or questions different from those used in earlier research to determine students' models of sound propagation, might be productive for this purpose. To address this

question, researchers administered a battery of semi-structured conceptual questions related to sound as a wave phenomena and in variety of situations.

"Surprises" were not expected with respect to this second research question related to pilot testing phase. Namely questions that were probed for this purpose were already used in various studies related to the sound propagation (Hrepic, 1998; Maurines, 1993; Wittmann, 1998). On the basis of these earlier studies the contextual situations were selected and further developed to specifically address the mental models of propagation (Hrepic, 2002; Hrepic et al., 2002).

4.2.1 Survey with open-ended questions

An open-ended test was administered to a large enrollment introductory, concept-based class. All students who were in the class on the day when the survey was administered took the test (158 out of 183 students enrolled in the course). The test can be found in the Appendix A.

4.2.1.2 Findings

Several findings of this survey are relevant for the construction of the final version of the test. First is that all ideas that were expressed in open-ended answers fit into mechanisms of sound propagation that were identified earlier and no new ideas were found. The second finding is that students express a vast variety of movements when asked how particles of/in the medium are affected when sound propagates. The list of these movements is in the Appendix B and their frequency, in the Appendix C. The most frequent are: longitudinal vibration, traveling in the direction of the sound propagation, vibration (with unspecified direction), transversal vibration, sinusoidal movement away from the source and random dispersive movement of air particles in the direction of sound propagation. Although very rare, "backward movement" (movement of the particles of/in the medium toward the source of the sound) was identified in this and previous studies (Hrepic, 2002).

Another finding is that the percentage of students (at this introductory, concept based level) who answer this type of questions correctly even after the instruction is rather low. In this particular sample (N=158) 4.4 % of the students answered all the

questions in the test correctly and 6.3% answered correctly all questions related to the movements of the particles of/in the medium.

Students are also frequently inconsistent in their answers related to dynamics of the particles in/of the medium when this dynamics is probed in different contexts. In this sample, 26 students were consistent and 132 were not with respect to the movements alone although we employed very inclusive definition of consistency.

Finally it is noteworthy that in this sample 11 (of 158) students used the term "sound particles" while explaining the nature of sound propagation. We associate this term to Entity Model. However, not all students who expressed the Entity Model used the term "sound particle" to describe it.

4.2.1.2 Implications

When asked about the dynamics of the particles of/in the medium, students' answers consist of a variety of motions including the motion toward the source of the sound. Because of this diversity, two questions related to the dynamics of the air particles of the medium are needed in the test to allow for all the motions that students' responses require.

Further, based on results from the open-ended test in which no new ideas of propagation appeared, we can conclude that the majority of students use models that were earlier identified. These models should be probed in the test. Although some models that may differ from those identified may appear, these instances are likely to be rare and isolated.

We can not expect that a large percentage of students will give correct answers in the multiple-choice test at this level. The situation may be better at higher introductory college physics level.

The usage of the term "sound particles" should be optimized in the test because some students use it to denote the sound entity different from the medium or the medium motion, however, not all of the students who have this same idea like the term "sound particle".

4.2.2 Survey with broad range of questions with semi-structured answers

The next task was to probe if some questions or contextual situations which differ from those that were used in interviews (Hrepic, 2002) and open-ended test -can be useful for the purpose of eliciting students models in the multiple-choice test. A semi-structured test with a large number of questions was used for this purpose. The answers were semi-structured to reduce number of simple "Yes-No" or "More-Less" answers. In this way the questions were concentrating on the explanation of the rationale of the answers. The semi-structured survey covered the following areas related to sound phenomena:

> ➢ Mechanism of sound wave propagation
> ➢ Wave properties of sound
> ➢ Doppler effect
> ➢ Relation sound propagation – light propagation
> ➢ Propagation of sound in mediums/vacuum

The test was administered in Fall 2002 to students who were enrolled in a concept based introductory physics course at KSU and who were not exposed to sound surveys earlier. The students took the test both before and after the instruction on sound. In pre instruction testing they earned 5 points (0.5% of the total grade) just for answering all questions. After instruction they earned maximum 8 points for answers with logical explanations. Out of 139 students enrolled in the class, 128 took the pre-instruction test and 115 took the post instruction test. The pre instruction test can be found in the Appendix D-1. The post instruction test had a set of additional questions with respect to the pre-instruction test and these additional questions are separately listed in the Appendix D-2.

4.2.2.1 Findings

The test and the grading procedure were set up so they aimed at eliciting of the rationale of the students' answers rather than the answers themselves. For this reason we gave the answers as multiple options and asked for the rationale in each of the questions. We hoped that students would in some of these questions describe on their own the

mechanism of sound propagation as they understood it in order to give the rationale for their answer choice. Unfortunately, this situation did not happen. None of the answers described the nature of sound propagation in a way that the usage of that question seemed promising for elicitation of the mental models of the mechanism of sound propagation.

4.2.2.2 Implications

Implication of this part of the pilot testing was that the contextual situations that are optimal for elicitation of mental models of sound propagation are simple propagation in the air, propagation through a barrier (wall) and "propagation" through the vacuum.

4.3 Boiling mental models down for instructional use

After verifying the mental models of sound propagation that the test should be aiming at and after verifying the optimal contextual situations for that purpose, the next step in the test construction was determining if the list of identified models can be narrowed down based on the possible commonalities that they might have. As it turns out, several models that were described in earlier studies (and were recapitulated in the Section 2.8. of this dissertation) have some common basic features. Accordingly, in this step they were clustered according to their "common denominators". Models (as well as sub models) can be distinguished according to the answers that they give to the four questions below.

1. What is sound?
2. What happens to the sound without the medium?
3. What is the dynamics of the particles of the medium during the sound propagation?
4. How is this dynamics related to the sound propagation?

According to these criteria, four fundamental models of sound propagation can be distinguished:

➢ Wave Model
➢ Intrinsic Model
➢ Dependent Entity Model
➢ Independent Entity Model

In addition to these four mechanisms, there is a particular understanding of what the sound is, that may be associated with different mechanisms of propagation. We call it "Ear-born" sound and it was described in the Section 2.8.4. Each of these models is summarized below.

4.3.1 Wave Model

According to Wave Model:

➢ Sound is a vibrational motion of particles of the medium caused by the source of sound.

➢ Without the medium, sound can not exist and can not propagate.

➢ When sound propagates, particles of the medium vibrate around the same point longitudinally (along the direction of sound propagation). Transversal and circular vibrations are (incorrect) wave sub models.

➢ This particular motion of particles of the medium is the sound.

The model as described above corresponds to the "Wave model" that was identified earlier.

4.3.2 Intrinsic Model

According to Intrinsic Model:

➢ Sound is a translational motion of particles of the medium caused by the source of sound.

➢ Without the medium sound can not exist and can not propagate.

➢ When the sound propagates, particles of the medium travel away from the source in the direction of sound propagation. At the same time and in addition to this motion, particles of the medium may or may not vibrate.

➢ This particular motion of particles of the medium away from the source toward the listener is the sound.

The model as described above corresponds to "Propagating air – intrinsic model" that was identified earlier. In accordance to this nomenclature, the Wave Model could be labeled "Vibrating air – intrinsic model".

4.3.3 Ear-born Model

According to Ear-born Model:

- ➢ Sound is exclusively what listener hears. Sound exists in the listener's ear-brain system and not elsewhere and not before the moving particles of matter hit the listener's eardrum.
- ➢ Without the medium sound can not be created.
- ➢ Particles of the medium may either vibrate or travel away from the source toward the listener.
- ➢ This motion creates the sound in the listener's ear.

The Ear-born sound model is different from the other four in that it is not a mechanism of the propagation but rather a definition of what the sound is and can be associated with more than one nature of propagation. Another feature of the Ear-born sound is that it is a partially correct idea and is well aligned with our daily definition of the sound.

4.3.4 Dependent Entity Model

According to Dependent Entity Model:

- ➢ Sound is a self-standing entity different from the medium through which it propagates. However, to propagate sound needs the motion of the particles of the medium. Due to this motion of the medium particles, sound propagates through the empty spaces in between them.
- ➢ Without the medium sound can exist but can not propagate.
- ➢ When a source creates the sound, it also sets the particles of the medium into motion so they either (a) travel away from the source toward the listener, (b) vibrate around the same point, (c) do both (a) and (b). In another version,

particles of the medium move in a specific way constantly and this motion is not affected by the sound propagation.

➢ The motion of the particles of the medium enables the sound to travel through the empty spaces in between them.

The Dependent Entity Model as described above corresponds to "Vibrating air" and "Ether and Compression" models described before.

4.3.5 Independent Entity Model

According to Independent Entity Model:

➢ Sound or sound particle is a self-standing entity different from the medium through which it propagates. Sound does not need the medium to propagate. It propagates independently through the empty spaces in between the medium particles.

➢ Without the medium sound can exist and can propagate.

➢ Particles of the medium either (a) travel away from the source toward the listener, (b) vibrate around the same point, (c) do both (a) and (b). In another version, particles of the medium move in a specific way constantly. This motion is not affected by the sound propagation.

➢ Sound propagates independently of the particles of the medium. The particles move this way because sound affects their motion.

The Independent Entity Model as described above corresponds to "Entity," "Shaking" and "Longitudinally Shaking Models described before"

Figure 4.1. represents these models and the differences between them in a cartoon-like manner. Human characters here represent air particles and footballs represent sound entities. These models are:

(A) Wave Model which is scientifically accepted model

(B) Propagating Air Model (hybrid model)

(C) Dependent Entity Model (hybrid model)

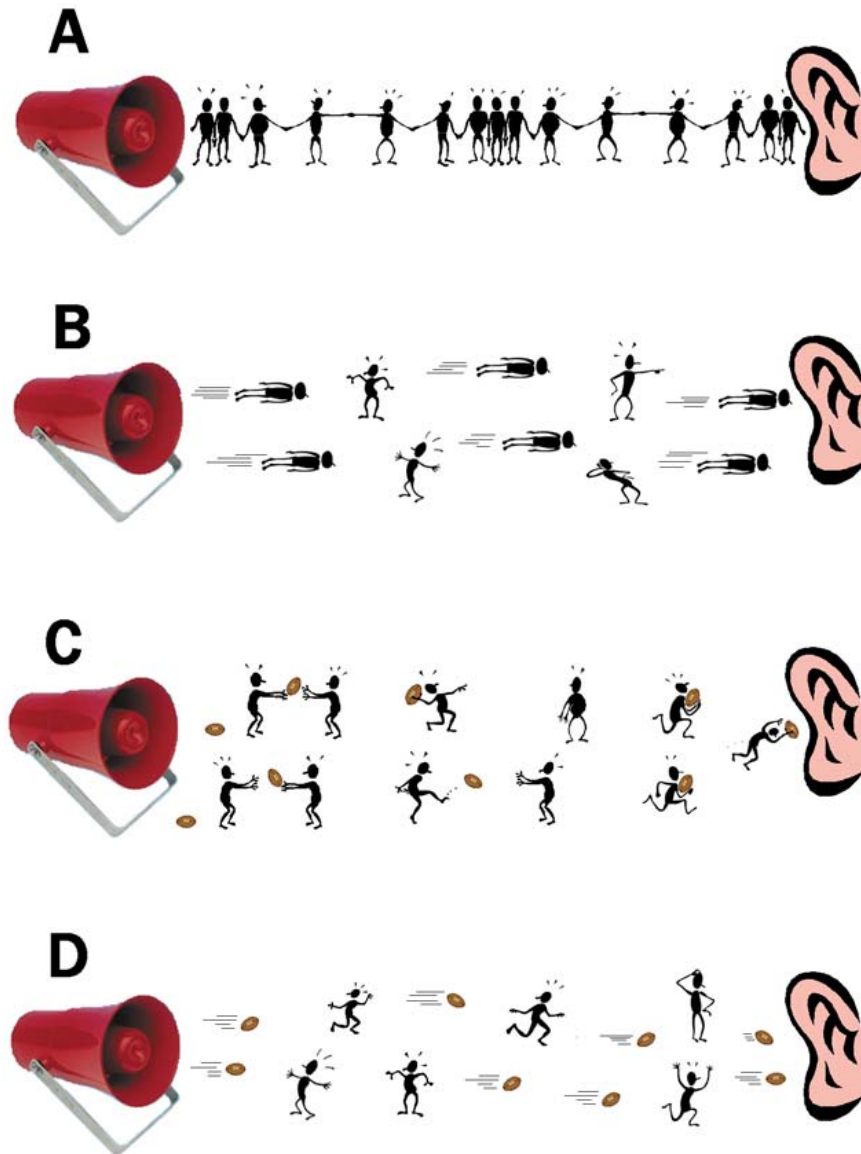(D) Independent Entity Model which is a dominant alternative model

Figure 4.1. Pictorial representation of mental models of sound propagation

In the Appendix I-1 and I-2 analogous representations without human characters are given for the both (air and wall) contexts of sound propagation.

An advantage of "boiling models down" or finding their "common denominators" is that we have now fewer and better defined models to deal with and at the same time

these reduced models comprise all the features of the models identified in the interviews. An additional advantage is the shift of the focus in the model classification from how medium particles move to what the sound is. For instance, consider the idea of the mechanism of sound propagation as air particles moving from the speaker to the listener. From the perspective of sound definition, this mechanism can be associated with all four different definitions which are: (1) Intrinsic Model (this movement is the sound), (2) Dependent Entity Model (sound is an entity that propagates due to this motion of the air particles), (3) Ear-born sound (this motion causes sound in the ear only (4) Independent entity (sound is an entity different from the medium that propagates with and without the medium and when it propagates through the medium it pushes the air particles this way. The same example applies also to the vibrational dynamics of the particles of the medium.

For this reason the test that elicits the models of sound propagation has to probe both dynamics of the particles of the medium and the rationale for this dynamics. That is: we need to elicit how this movement is associated with the sound. But, models are defined and will be sorted out according to the sound definition, not particle dynamics.

4.4 Pre-survey

Once models were known as well as optimal contextual situations, the next step in test creation was to map defined mental models onto answer choices of the multiple-choice test. The main difficulty in this mapping was that mental models of sound propagation are not simple knowledge elements even in their reductionistic version. They are stories that may not fit a single answer choice. Each of the described models also has its sub-models which is the reason that sometimes more than one choice in the same question may correspond to the same model. In addition, because of the nature of hybrid models these models may not map onto answer choices so they match one-on-one (that one choice corresponds to one model). Instead, in some instances more than one model may correspond to the same choice. Hybrid models also cause overlaps in multiple-choice answers so that some choices pertaining to the same question may have substantial commonalities. In the case of sound more than one question was needed to differentiate between models. For example longitudinal movement of medium particles during the

sound propagation is consistent with Wave, Independent Entity, Dependent Entity and Ear-born Model. Thus, two questions were needed to determine the movement alone. Finally, a student who uses a hybrid model can give a variety of correct answers on standard questions. The goal of this test was to avoid this situation.

Our first attempt to solve this long list of problems was creating a test that would, as a whole, probe students' models. We mapped answer choices so that particular combinations of answers throughout the test corresponded with a particular mental model.

4.4.1 Pre survey testing

In the pre-survey phase, the first version of the multiple-choice test was probed and then improved and refined. For this purpose we initially utilized surveys with seven multiple-choice options in each question. Later, we utilized a combination of 5-option surveys with semi-structured interviews in order to refine the test. First surveys had 7 answer choices because in addition to the model related answers each question had following options:

f) If more than one of the answers is correct, list them here. _____

g) None of the above. The correct answer is:

This survey can be found in the Appendix E. These additional choices were included to determine the need for possible adjustments of the offered choices and to determine the possible need to include new choices. Option g) was somewhat different than in other questions only in the first question that deals with the general mechanism of sound propagation . For this question it was:

1g) If you do not agree with any of statements above, or if you agree partially with some of them, please write the correct statement that describes sound propagation through the air.

The specific instruction pertaining to the choice g) of the question 1, aimed at possible ideas of the nature of sound propagation that were not identified earlier.

The 7-option survey was administered to an introductory physics class at Kansas State University in the late Fall semester 2002. Out of 100 students who took the test,

only four wrote their own answer and each of them only in one question. All of these four written answers pertained to the question 1. However none of them was related to the mechanism of sound propagation, but rather they described the sequence of events that occur when sound is created, propagated and heard more broadly than the question was asking about. Specifically, in the context of the propagation of the sound through the air one student referred to the role of the source in the sound propagation and the other three students addressed the role of the air in the sound propagation in the wall context. Multiple answers that students were choosing were mainly related to the combination of the longitudinal and transversal movement, to the combination of Dependent and Independent entity and to the combination of the Ear-born sound with some of the other models.

All these issues were addressed in the survey version of the test to the maximum possible extent. Results related to the wording of the test are not elaborated in detail here because they are reflected in the differences that were made from the pre-survey to the survey test. Also since the survey version of the test was probed for the reliability and validity, the reasons and the rationale for each of these changes is not worth detailed elaboration. The test we used for the survey on the large scale is in the Appendix F (Air context) and Appendix G (Wall context) and will be described in detail in the next section.

4.4.2 The issue of student's consistency

Besides improvement of the wording of the test choices, the pre-survey phase of the research revealed another issue which is related to students' (in)consistency and had profound implications on the nature of the final test. Namely, after administering the 7-choice version we realized that very few of the students picked throughout the test the combinations of the answers that corresponded to what was mapped out as a consistent model. This situation did not improve much in the 5-choice test version that we administered next. The 5-choice version was administered to the two lab sections of each of the calculus-based and concept-based introductory physics courses. The percentage of students that gave consistent answers in all questions when first multiple-choice tests were administered ranged from 0% in the concept based class (N=79) to 10.5 % in the

73

calculus based class (N=38). And if loose criteria are applied, this percentage ranged from 9% to 24 % respectively.

Contrary to the researchers' expectations, the level of inconsistency that was observed in the pilot testing (with the open-ended version of the test) did not decrease in the multiple-choice questionnaire.

The approach of determining a model from the answer combination of all questions in the test failed because of the high degree of students' inconsistency that occurs when students are not sure about the answer. By interviewing students in the pre-survey phase of research, it become obvious that most of them do not have clear model at the start. They develop their ideas based not only on their experiences and previous related knowledge but also – to a large degree – based on the choices offered in the test and based on the questions themselves. Because answer choices were mapped according to ideas that researchers earlier heard from the students themselves, test takers often like more than one of the choices in each of the questions.

An additional problem that increased the inconsistencies was that students apply different test taking strategies when they are not sure about the answer. For example, they may distribute their answers over different models that they like in order to prevent being incorrect in all instances (or in order to be correct in at least some of them). In addition, students seem to concentrate on one question at the time and they pick what "makes sense" in each particular instance (question).

For all of these reasons the attempt to probe students' models based on the test as a whole did not work. In retrospect, previous research supports this conclusion because it showed that students often do not have clear model even after the instruction. The proposed use for this test is to be a diagnostic tool that will be administered before the instruction. At this time students are even less likely to have well formed ideas. What students normally bring in the classroom when sound is concerned (as well as many other physics and scientific topics) is a vast everyday experience and a set of a vague ideas and fractioned pieces of knowledge. The task of this test is to identify those pieces of students' knowledge before the instruction in order to build on them so students achieve stable scientifically accepted understanding. For this purpose, the approach to the testing that was initially applied required a change. Number of items needed to identify a model

had to be minimized (rather than spread out over the whole test). Mapping a model on a single question was not possible because of above mentioned issues in dealing with mental models and hybrid models in particular. But the reduction of necessary items was possible.

Because of the variety of movements of the particles of the medium related to the sound propagation that students express in open-ended questions, two questions were needed to elicit this dynamics. One more question was necessary to associate this motion to sound propagation. Thus, a minimum of three different questions was needed for one probe of the student's model. In this new approach, additional questions could have been used to probe if student's model is stable or other models are also attractive to him or her. This new approach therefore made it possible not only to determine whether a student is in a pure or in a mixed model state but also what models a student uses and how often he or she uses each of them. The hope was that this change in methodology will be sufficient to bring down the percentage of the models that cannot be identified to less than 10 %. The main advantage of the new approach was its "partial credit" grading approach which is different from the initially used system. The old system in a way treated only students that got 100% correct answers as worthwhile grading. This new approach to model analysis in which the complete meaning of the particular answer choice is determined by answers given in other (sometimes all other) test questions we call Linked Item Model Analysis (LIMA).

4.5 Survey

Based on all previously described results we created a test that was then administered to a large sample at a variety of different institutions. While creating the test choices, we followed standard procedures in writing quality items (Haladyna, 1999) such as:
- Using typical errors of students to write distractors and making all distractors plausible,
- Keeping all choices about the same length,
- Avoiding overly specific or overly general wordings,

➢ Minimizing the amount of reading in each item while making sure that all necessary ingredients of the model are present,

➢ Avoiding giving clues to the right answer,

➢ Making sure that only one of the choices is the correct answer and

➢ Keeping vocabulary as simple as possible without compromising accuracy. (A problem with this request was the word "propagate" with which many students are not familiar. One of expert reviewers believed that some other simpler words such as "traveling" were "biased" toward Entity Model in a sense that they seem to state that "some thing" moves in the case of the sound propagation.)

We limited the number of answer choices to 5 options in each of the questions because this test will be primarily used in the classroom with some sort of the commercially available class response systems and some of these systems can not handle more than 5 answer options. In addition, and more importantly, a greater number of answer choices causes difficulties for a student in keeping record of the earlier choices by the time he or she reaches the last ones.

4.5.1 Describing the surveyed test

The test as it was given throughout the 2003 year is described below. After each of the answer options the model (or models) that the choice corresponds to is denoted in brackets. Exceptions are the answers in questions 2 and 3 because the combination of the answers determines whether or not the movement is compatible with any of the models. The full list of models, sub-models corresponding answer choices, program codes and sub-model descriptions can be found in the appendices J-1, 2 and K-1, 2. For simplicity, from now onwards "Independent Entity Model" may be also referred to as "Independent" and "Dependent Entity Model" as "Dependent". Similarly the attribute "model" might be omitted when Wave, Intrinsic and Ear-born Models are mentioned.

Paraphrased, test questions can be simplified to following inquiries:

1. What is the **mechanism of sound propagation** in the air/wall?

2. How do particles of the medium **vibrate**, if at all while the sound propagates?

3. How do particles of the medium **travel**, if at all, while the sound propagates?

4. What does this motion have to do with sound propagation – **cause and effect relationship**?

5. What does this motion have to do with sound propagation – **time relationship**?

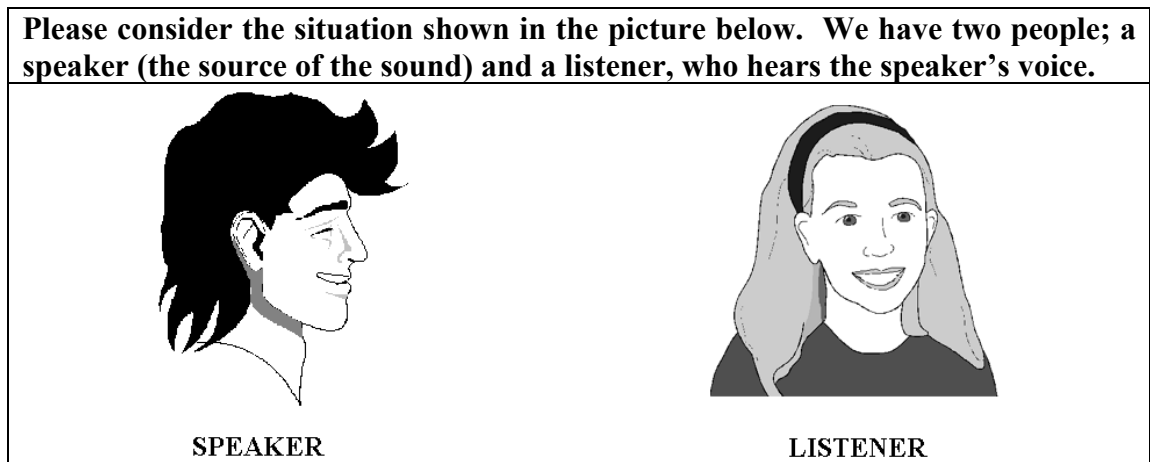6. What happens with sound propagation **in the vacuum**?

Out of these six questions two define dynamics of the particles of the medium (Q2 and Q3). We will call these two questions dynamics defining questions. Remaining four questions serve to define the relationship of the sound propagation with this dynamics, ach from it's different perspective. These four questions (Q1, Q4, Q5 and Q6) will be together called relationship defining questions.

In the following section two contextual versions of the test will be described. These versions include (1) a version that includes propagation through the air and through the vacuum (for short "air context") and (2) a version that includes propagation through the wall and (again) through the vacuum (for short "wall context"). In addition to two contextual versions of the test we will distinguish different temporal versions of the test i.e. the test versions that were employed at different phases of the research. Temporal test versions will be labeled numerically or according to the research phase in which they were employed. When changes related to different temporal versions were made, air and wall test contexts were changed simultaneously. In the next section we will describe the test version employed in the survey phase of the research which will be given a numerical code 8.9.

4.5.1.1 Survey test: Air context

The version of the test as pertaining to the Air-Vacuum or air context went as follows:

Please read all of the choices for each question before choosing one of them. Also, please answer *all* questions to the best of your ability. The following 6 questions together will give you an opportunity to fully describe the propagation of sound through air.

Please consider the situation shown in the picture below. We have two people; a speaker (the source of the sound) and a listener, who hears the speaker's voice.

SPEAKER   LISTENER

1. Which of the following statements best describes the propagation of sound through the air?
(Propagate = to spread out, to travel through, to transmit).

  a) *Sound moves* through the empty spaces *in between the air particles and affects their motion* in a specific manner. [Independent]
  b) The *air particles move* in a specific manner*. Sound is this motion of air particles*. [Intrinsic, Wave]
  c) The *air particles move* in a specific manner. This motion *enables the sound to travel through the empty spaces in between them*. [Dependent]
  d) The *air particles move* in a specific manner. This motion creates the sound after hitting the listener's eardrum. *Sound does not exist before the eardrum is hit*. [Ear-born]
  e) The *sound particles move* in a specific manner. The moving *sound particles propagate throughout the air*. [Independent, Dependent, Phonon*]

*Phonon Model will be explained in detail in the next section.

The following two questions (2 and 3) together (in combination) will give you the opportunity to completely describe the motion of air particles while the sound propagates through the air.

2. **Do air particles *VIBRATE* while the sound propagates through the air?**

   a)   No. They ***do not vibrate*** in any way.
   b)   Yes. They vibrate ***randomly*** in all directions.
   c)   Yes. They vibrate primarily ***back and forth*** along the direction in which sound propagates.
   d)   Yes. They vibrate primarily ***up and down*** perpendicular to the direction in which sound propagates**.**
   e)   ***Both c) and d)*** are correct. They vibrate equally along and perpendicular to the direction in which sound propagates.

3. **Do air particles TRAVEL in a certain direction (move away from one place to another) while the sound propagates through the air?**

   a) No. They ***stand motionlessly*** at their original point (without either traveling or vibration).
   b) No. They ***only vibrate*** around the same point (without traveling).
   c) Yes. They ***travel in the direction*** of sound propagation (away from the source).
   d) Yes. They ***generally travel in the direction*** of sound propagation (away from the source) but ***at different angles*** due to scattering.
   e) Yes. They ***may travel equally in any direction*** (with respect to one another and with respect to direction of sound propagation).

Comments:

Combination 2a-3a is not compatible with dependent entity because dependent entity in the model as we defined it can not move without motion of air particles.

4. **Complete the following sentence: The motion (or lack of motion) of the air particles that you described in previous questions…**

   a) …***is not affected by the propagation*** of the sound. [Independent (secondary choice), Dependent (secondary choice)]
   b) …***is caused by the propagation of the sound*** through spaces ***in between the air particles.*** [Independent]
   c) …***enables the propagation of the sound*** through spaces ***in between the air particles.*** [Dependent]
   d) …creates the sound in the listener's ear because ***sound does not exist before air particles hit the listener's eardrum.*** [Ear-born]
   e) …***is*** the sound. [Intrinsic, Wave, Phonon]

5. **Complete the following sentence: The motion (or lack of motion) of the air particles that you described in previous questions…**

   a) …occurs *before the sound can propagate* through spaces *in between the air particles, as a precondition* for propagation. [Dependent]
   b) …occurs before the sound is created. *Sound is created when air particles hit the listener's ear and it does not exist earlier*. [Ear-born]
   c) …occurs *when sound encounters the air particles while passing* through spaces *in between them.* [Independent]
   d) …occurs *at the same time as sound propagates because the described motion of air particles is sound*. [Intrinsic, Wave, Phonon]
   e) …exists *all the time the same way*, with or without the sound propagation. [Independent (secondary choice), Dependent (secondary choice)]

6. **Can sound propagate through a vacuum (empty space without matter)?**

   a) *No. Sound can exist in empty space without particles of matter*, *but it needs the motion of those particles to be carried* to another place. (A vacuum has no matter so this is not possible). [Dependent]
   b) *No*. *Sound is the motion of particles of matter* caused by the source of sound. (A vacuum has no matter so this is not possible). [Intrinsic, Wave, Phonon]
   c) *No*. Sound is created when moving particles of matter hit the listener's eardrum. *Sound does not exist before the listener's eardrum is hit*. (A vacuum has no matter so this is not possible). [Ear-born]
   d) *Yes. Sound particles move in a vacuum* as freely as, or more freely than in matter, **because** *a vacuum has no matter to obstruct their motion*. [Independent]
   e) *Yes. Sound propagates through a vacuum* as easily as, or more easily than through the matter, *because a vacuum has no matter to resist the sound propagation*. [Independent]

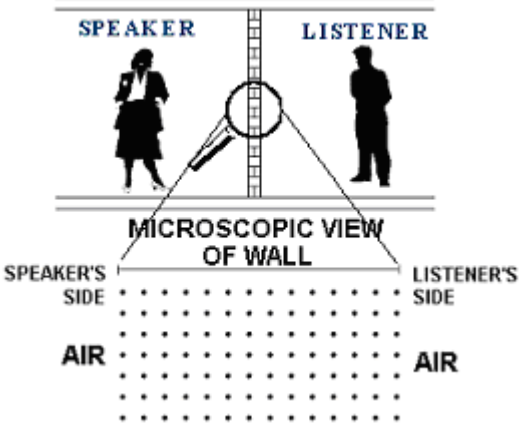### 4.5.1.2 Survey test: Wall context

The difference between the wall and the air context is that the wall context has three kinds of "identifiable kinds of things" --wall particles, air particles and the sound -- whereas the air context has only two. One of the mechanisms that must be probed in the wall context is the "propagating air" mechanism.   This model was identified through the questions Q2 and Q3 in the test version pertaining to the air context. In the air context, those two questions defined if the air moved away from the source and toward the listener.

However, in the wall context, questions Q2 and Q3 ask about the dynamics of the wall particles. Thus, the "propagating air" idea has to be probed through one of the choices pertaining to other questions. The optimal way to ensure that the Propagating Air Model is probed in other questions (Q1, Q4 and Q5), without increasing the total number of choices in these questions, was to replace the choice corresponding to the Ear-born Model in these questions with a choice corresponding to the 'Propagating Air' Model. The Ear-born Model was therefore probed only in Q6. Thus a student could potentially select the Ear-born Model choice in Q6 along with choices corresponding to any of the propagation mechanisms in Q1 through Q5. Because of this arrangement with ear-born choice offered only in question No. 6, the wall context test can not be divided into smaller tests although air context can. This will be discussed in the Chapter V in a greater detail.

For the reasons described above, questions and answers in the wall context are the same as those in the air context except in the introductory statement, three of the answer choices and "type" of the medium particles. The full versions of the survey tests can be found in the appendices F (Air context) and G (wall context). The differences of the wall context with respect to the air context are described below. Introductory statement and the corresponding picture were as follows:

Please consider the situation shown in the picture on the right. We have two people; a speaker (the source of the sound) and a listener, who hears the speaker's voice. They are in two rooms separated by a solid brick wall.

It is our common experience that if the wall is relatively thin and speaker is loud, the listener can hear the sound in the other room. The figure on the right shows a microscopic view of a wall. The particles of the wall (shown as dots) are arranged as shown in the figure.

The answer choices that are different in the wall context with respect to the air context are 1d), 4d), 5b). Their text in the wall context is give below.

Choice 1d) in wall context:

The *air particles move* in a specific manner through spaces *in between the wall particles from the speaker's side to the listener's side of the wall*.

Choice 4d) in wall context:

…*is caused by the propagation of the air particles* while they pass through spaces *in between the wall particles* to another side of the wall.

Choice 5b) in wall context:

…occurs *when air particles encounter the wall particles while passing* through spaces *in between them to another side of the wall*.

4.5.2 Experts' review of the content and correctness of the answer choices

A panel of experts (Ph.Ds in Physics) reviewed the test in two phases of its development. The first time was at the end of the pre-survey phase before we administered the test to a large sample. The second time was in the post survey phase after we made modifications based on the results in the survey phase. Each time four experts reviewed the test to determine if choices that we consider correct are correct and the only correct answers and to give us the feedback on the clarity of the sentence formulations in the test. This procedure was described in greater detail earlier in Section 3.3.2.2.

In the first round of the experts' reviews only one major issue was raised. Namely an expert believed that in the case of the answer 1e, an incorrect answer could be interpreted as a correct one. The specific answer in the wall context version that this expert reviewed was: "Sound propagates so that sound particles vibrate throughout the wall".

According to his opinion "sound particle" could be interpreted as a phonon, which is a quantum of acoustic energy. Phonon is used in physics to describe the mechanical excitations at atomic level. It represents an acoustical/thermal analog to the quantum of electromagnetic radiation, the photon. Later, we changed this answer to an extent but this expert felt the same way about the revised answer: "I cannot get beyond thinking 'phonon' when I see sound particle. In either case, there is a more correct choice." This expert believed that although there is evidently a more correct answer related to Q1, the

choice that contains the phrase "sound particle" could be interpreted by a reader as a phonon, which would make the choice 1e) a correct answer. However, according to our scheme this student's response would be categorized as incorrect.

It is highly unlikely that students in our target population (introductory college or lower levels) could misunderstand "sound" particle as a phonon. The concept of phonon is typically introduced at upper level undergraduate physics. In earlier studies students spontaneously used the term "sound particles" to denote a sound unit completely independent from the medium. Also, none of the other seven experts brought up this issue, but we nevertheless addressed this expert's concern. The way we resolved it was that we (i.e. analysis program) specifically probe if a students picks the choice 1e while answering all other questions correctly. If this is the case he or she is assigned to a "Phonon Model" which is separated in a detailed table and in graphical representations it is summed up together with the correct Longitudinal Wave Model.

Also, based on an expert's input we added the word "primarily" into choices 2c and 2d of the Q2. These choices now go as follows:

2c) Yes. They vibrate *primarily* back and forth along (parallel to) the direction in which sound propagates.

2d) Yes. They vibrate *primarily* up and down perpendicular to the direction in which sound propagates**.**

"Primarily" was added because this vibrational motion of the sound propagation is superimposed on the thermal motion of the particles of the medium. In gasses like air this is especially the issue because of the long random walk paths of the particles when compared to the solids (like a wall).

These remarks were related to validity issues and were fixed as described. Other issues that experts brought up were different suggestions related to improvement of the clarity of the wording. Experts' comments that were made in the post-survey phase will be described later.

4.5.3 Data analysis

In introduction to the Section 4.4. we explained why it is not possible in this test to map different models onto answer choices so that one choice corresponds to one model (one-on-one). This restriction was the primary reason that model analysis analytical method as described by Bao (2000) was not useful in this case. Instead, the LIMA (Linked Item Model Analysis) approach was developed and than utilized in a following way: Results of the tests are analyzed in a way that a program compares student's set of 6 answers with sets of answer combinations associated with the models as defined in appendices K-1 and K-2. If a match is found, the student is consistent or in a pure model state. If no match is found, student is in a mixed model state (inconsistent). In that case, the program looks for triplet combinations and compares them with those in the database.

The test has four triplets (Q1-Q2-Q3; Q2-Q3-Q4; Q2-Q3-Q5; Q2-Q3-Q6). Each of these is a single probe to see what model(s) a student uses. If the student is consistent so that the corresponding model combination is found in a database, this particular model is assigned to all four triplets. If student is not consistent, program determines the model that was used in each of the triplets.

A model is not necessarily ascribed to any triplet. If students picks an incompatible motion combination (E.g. "Yes, medium particles vibrate" in question 2 and then "No, they are motionless, without vibration and without movement" this will cause that all four triplets will be inconclusive or sorted into "Other". Besides, not all of the models are compatible with all of the movements so some triplet combinations may not be recognized as a model although movements are consistent by themselves. E.g. "Random vibration" is not compatible with the Wave Model and dependent entity can not be associated with motionless particles. These combinations and others of this kind are classified by the program into "other" category. Further details of how the program functions can be found in Appendix L.

Although the 6-question model combinations are all unique, they may have overlapping triplets. In these cases a triplet is assigned to a "lower level model". For example, the Dependent Model is consistent with "constant motion" and "preconditioned motion". In the first case, sound uses the motion that exists all the time with or without the propagation and in the second one the motion is created when the source creates the

sound. This distinction between these two models is clear if a student is self-consistent in all 6 questions but may not be recognizable from each of the triplets. In this case, the sub-model differences can not be distinguished and the triplet model is assigned to a "parental" sub-model or a model that encompasses both of the indistinguishable modalities. In the mentioned case this parental model is "Generic Dependent Entity".

With respect to the model analysis, an important difference between the air and the wall context pertains to the fact that in order to determine the model, in the wall context one more question is needed than in the air context (given that the number of the choices stays the same). In the wall context, the answer choices that describe the sound propagation as traveling of the air particles from one side of the wall to another do not determine what the sound is. A model pertaining to the propagating air mechanism in the wall context has to be determined from the additional question which is in that case question 6.

Programs for the data analysis can be found in folder No. 3 on the CD that accompanies this dissertation or online (see Appendix W).

4.5.4 Display of results in terms of mental models

The program for data analysis displays results in the five different graphs that show following information:

1. Percentages of times that a particular model is used,
2. Percentages of students using a particular model at least once,
3. Movements of particles of the medium,
4. Students' model states, and
5. Correctness of the answers

Models that each of the students in the sample uses are probed 4 times in each of the tests. The graph that shows percentages of times that a particular model is used displays them so that contributions from the consistent and inconsistent usage are displayed as parts of the same column and separated by color. Figure 4.2. shows this graph. Transversal and Circular Wave Models are displayed together but separately from the Longitudinal Wave Model. Two columns with Wave models are spaced out from the

rest of the models because this makes it easier for the instructor to quickly estimate the size of the "good side".
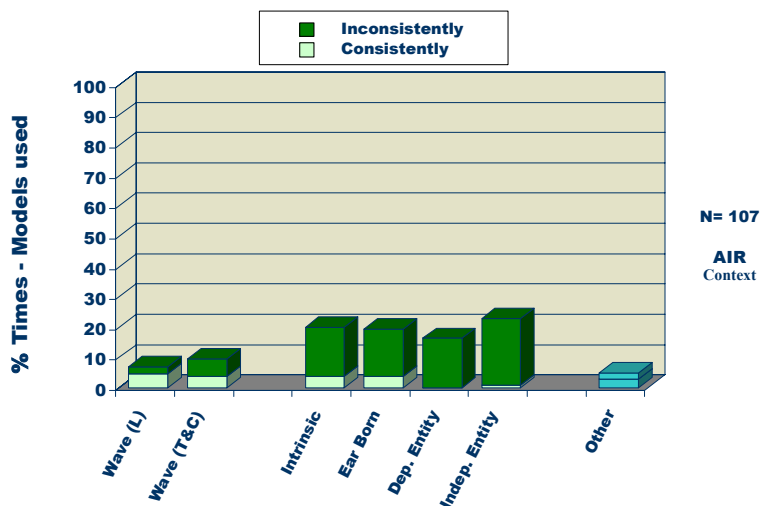

Figure 4.2. Percentages of times that a particular model is used

The Figure 4.2. as shown here appears in each of the power Point presentations with the results. These presentations are given as a part of the electronically presented results on the CD that accompanies the dissertation as well as on the mentioned web sites. The graph in the analysis program (also given on the CD) is different from the one shown as Figure 4.1. in that it does not have the context and the number of students shown at the right side.

Second graph embedded into the analysis program displays the percentage of the students that use each of the models at least once. Results pertaining to each of the models in this graph are also separated by contributions from the consistent and inconsistent usage. This graph and the graphs that follow will be presented in a way they appear in the analysis program (not power point presentations of the results). Results presented in these sample graphs are obtained from an actual algebra-based introductory physics class at KSU.
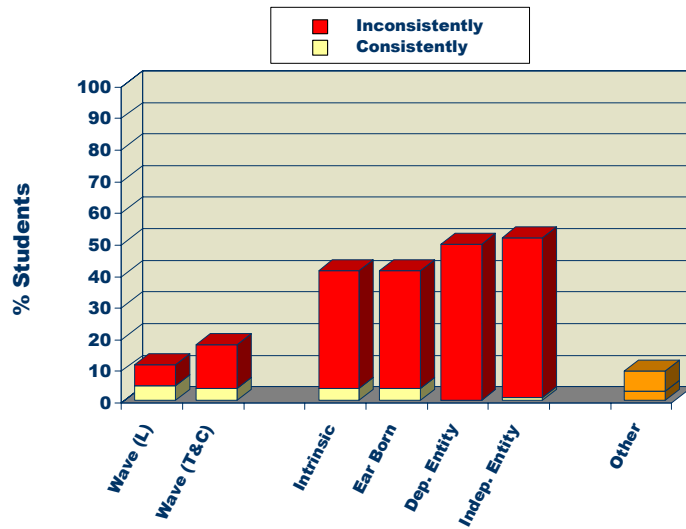
Figure 4.3. Percentages of students using a particular model at least once

Movements of particles of the medium are displayed in the third kind of the bar chart. Horizontal axes display different vibrations and translational motion is added on top of each of them according to the combinations that students picked.

Figure 4.4. Movements of particles of the medium.

The next diagram (Figure 4.5.) shows the results of students' answers according to the number of students in different model states. The left column shows the number of

students that consistently use a particular model. In the bottom part are students that consistently use the correct model, while in the upper part are those that consisten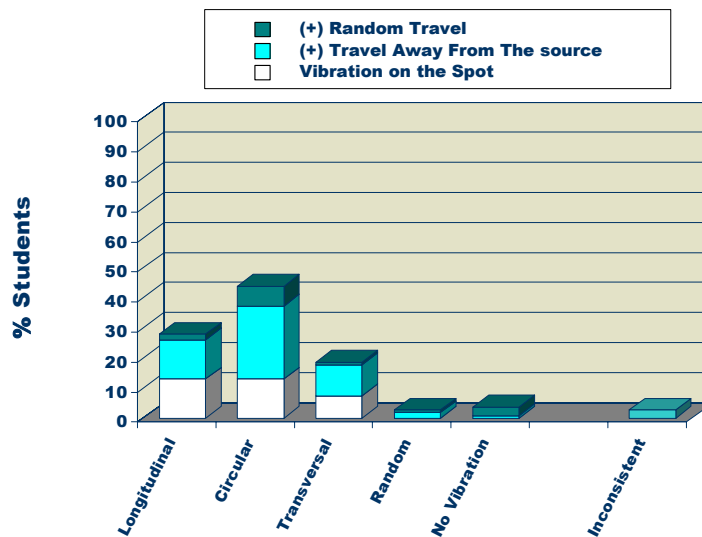tly use any other model. Students that consistently use a model are in a pure model state. The column on the right side of the graph shows students (in absolute numbers) that do not use consistently any of the models. In this column two particular mixtures of the models are separated out. First (displayed at the bottom) is the mixture of the Wave and Ear-born models. This one is sorted out because one might argue that there is nothing wrong with this combination if put together. A second distinguished mixture of models is the combination of Independent entity and Dependent Entity Models. These two models are not separated by a clear cut borderline but lay along the continuum of "dependency". Because of this continuity, Dependent – Independent Model mixtures are frequent.
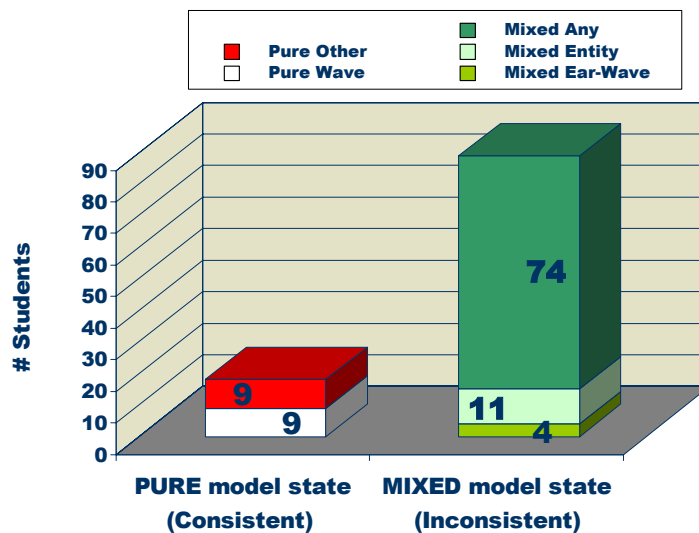

Figure 4.5. Students' model states

The final representation of the results of the survey that the analysis program provides is simple percentage of the students that get each of the questions right. This graph is shown in the Figure 4.6.
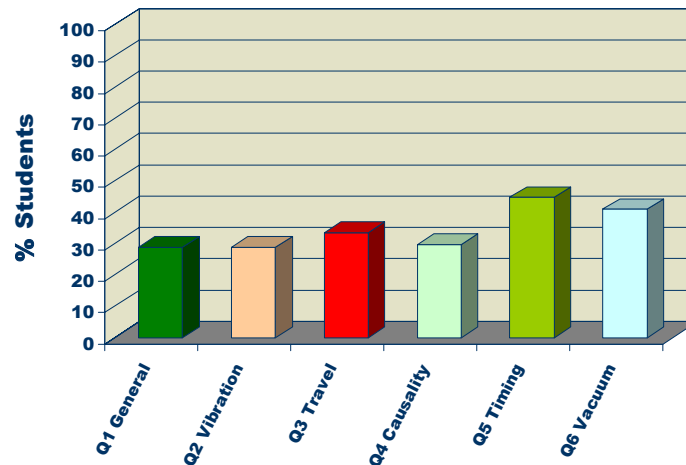
Figure 4.6. Correctness of the answers

In addition to these graphs, the analysis program contains also a sheet with detailed representation of each of the sub models in a way that each of them is presented separately.

4.5.5 Survey testing

The test was administered as a survey to a large and diverse sample to find quantitative values relevant to determining reliability and validity of the test. The survey was done during the Spring and Fall semesters of 2003. Table 4.1 shows all the samples that were tested and number of students in each sample. The test was administered at different institutions in US and at several institutions in Croatia (International code HR). The accuracy of the author's translation of the test to Croatian language was verified by a court interpreter for English language in Croatia.

To distinguish different versions of the test that were administered at this stage of the research from modified versions that were made in the post-survey phase (based on the results in the survey phase of the research), we will refer to the test that was used for the survey and interviews during 2003 as "survey test" or "survey version of the test". The survey version of the test and corresponding analysis programs were assigned a numerical code 8.9 and both designations will be sometimes used.

Table 4.1.

Institutions and classes that participated in the survey

Institution Level	Semester	State / Country	Course Math Level	Data Taken	Students Enrolled	Students Participated	Participation at least (%)	Incentive	All Pre	Air Pre	Wall Pre	All Post	Air Post	Wall Post
University	Spring 03	NY	Calculus	Online	113	100	88.50	Homework	100	100		95	95	
University	Spring 03	PA	Algebra	In Class	14	12	85.71	Class activity	12	6	6	10	6	4
University	Spring 03	KS	Calculus	In Lab	160	127	78.75	Request				127	69	58
University	Spring 03	KS	Algebra	In Lab	320	207	64.69	Request				207	107	100
University	Spring 03	KS	Concepts	In Class	156	38	24.36	Request				38	33	5
University	Fall 03	NC	Calculus	Online	83	57	68.67	Extra Credit for Correctness	57	57		19	19	
University	Fall 03	HR	Calculus	In Class	100	60	60.00	-//-				60	29	31
University	Fall 03	KS	Algebra	In Class	268	177	66.04	-//-	175	99	76	177	98	79
University	Fall 03	KS	Concepts	Online	156	96	61.54	-//-				96	96	
University	Fall 03	KS	Concepts	In Class	162	105	64.81	Request				105	105	
University	Fall 03	KS	Concepts	In Class	128	78	60.94	Request				78	78	
University	Fall 03	IL	Concepts	In Class	68	53	77.94	Request				53	20	33
University	Fall 03	LA	Concepts	In Class	47	37	78.72	Request				37	19	18
C. College	Spring 03	KS	Concepts	In Class	20	19	95.00	Request				19	19	19
SUM Univ.					**1795**	**1166**	**64.96**		**344**	**262**	**82**	**1121**	**793**	**347**
High S.	Spring 03	KS	Concepts	In Class	118	102	86.44	Request				102	82	20
High S.	Spring 03	MN	Concepts	In Class	48-50	47	94.00	Request				47	23	24
High S. (1,2)	Spring 03	HR	Algebra	In Class	51-57	51	89.47	Request				51	51	51
High S. (1)	Fall 03	HR	Algebra	In Class	55	51	92.73	Request	49	28	21	51	28	23
High S. (2)	Fall 03	HR	Algebra	In Class	54	51	94.44	Request				51	24	27
SUM H.S.					**326-334**	**302**	**90.42**		**49**	**28**	**21**	**302**	**208**	**145**
Middle S. (1)	Spring 03	HR	Algebra	In Class	42-50	42	84.00	Request				42	20	22
Middle S. (2)	Spring 03	HR	Algebra	In Class	95	90	94.74	Request				90	44	46
SUM M.S.					**137-145**	**132**	**91.03**		**0**	**0**	**0**	**132**	**64**	**68**

4.5.6 Survey results

Results of the testing in the survey phase are summarized in the Table 4.2 in terms of the model distribution and students' self consistency. The results of each of the samples are shown in the Appendix M. The results in the appendix M are given in terms of the percentages of times that a particular group of students used a particular model. In the Table 4.2 results are presented separately for pre and post instruction results obtained from students at different educational levels. Percentages in the Table 4.2 reflect simple averages of the percentages that each of the models was used in each of the samples pertaining to the specific category. A standard deviation was calculated with respect to these simple averages. In addition to this information, tables in the Appendix M show weighted averages for each of the categories.

In addition to the model distribution, Table 4.2 as well as tables in the Appendix M, show the percentage of students that used a particular model consistently. The percentage of those who used a Wave Model (either Longitudinal, Transversal or Circular) is also shown.

While calculating the averages we excluded samples that had fewer than 15 students. We also excluded samples that were tested after instruction during which a particular intervention was made in order to address student's understanding of the sound that would not have been made without the test or without instructors' familiarity with the test. These samples were excluded because they were not compatible with others and can be considered outliers by their characteristics although they were not necessarily outliers by results. The number of samples that were included and number of incompatible samples in each of the analyzed categories is reported in the Table 4.2.

Table 4.2.

Results of the surveys in terms of the model distribution and students' self consistency (in percentages)

Context	Pre / Post	Institution	Included Samples	Incompatible samples	N	Result	Consistent (Pure model state)	Consistent Wave	Wave (L)	Wave (T&C)	Intrinsic	Ear-born	Dependent Entity	Independent Entity	Other
Air	Pre	University	3	1	257	Average	14.05	4.26	3.99	5.41	22.93	16.64	18.28	27.84	4.24
						SD	8.16	4.11	4.84	2.75	2.97	4.92	1.67	6.64	1.46
Air	Pre	High School	1	0	28	Average	7.14	0.00	0.89	5.36	13.39	14.29	28.57	26.79	10.71
Air	Post	University, CC	11, 1	3	689	Average	13.60	5.66	7.28	4.35	21.13	13.78	19.67	29.00	4.80
						SD	11.36	6.89	8.24	3.76	6.91	5.84	7.50	7.49	3.89
Air	Post	High School	3	2	156	Average	14.67	1.71	4.71	2.08	15.97	16.29	20.14	32.64	8.16
						SD	6.65	2.01	3.65	2.22	4.14	3.88	2.19	2.01	1.43
Air	Post	Middle School	2	0	64	Average	11.14	0.00	0.00	0.00	13.75	36.99	22.27	20.23	6.76
						SD	12.54	0.00	0.00	0.00	1.77	11.33	1.45	9.16	2.49
Wall	Pre	University	1	1	76	Average	14.47	3.95	3.29	6.91	12.83	6.25	23.03	43.09	4.61
Wall	Pre	High School	1	1	21	Average	14.29	0.00	0.00	1.19	23.81	7.14	21.43	26.19	20.24
Wall	Post	University, CC	6, 1	3	338	Average	13.45	5.56	6.20	8.06	19.01	2.90	24.37	32.66	6.78
						SD	11.49	7.70	7.20	5.06	8.14	3.30	6.58	13.32	5.26
Wall	Post	High School	3	2	95	Average	15.05	5.75	7.90	6.48	14.76	8.19	22.77	30.02	9.87
						SD	2.57	2.30	4.32	2.41	2.38	1.88	2.69	0.86	7.61
Wall	Post	Middle School	2	0	68	Average	7.81	4.55	0.27	10.20	14.23	4.10	31.18	29.77	10.25
						SD	1.82	6.43	0.38	12.89	3.98	4.19	4.12	14.78	1.64
Both	Pre	University	1	1	175	Average	13.14	2.86	2.57	4.43	17.86	13.86	20.00	37.14	3.57
Both	Pre	High School	1	0	49	Average	10.20	0.00	0.51	3.57	17.86	11.22	25.51	26.53	14.80
Both	Post	University, CC	6, 1	2	559	Average	10.93	4.27	5.71	5.72	21.07	8.31	21.62	31.09	6.47
						SD	9.34	5.31	5.13	4.09	7.29	3.90	7.10	9.77	4.66
Both	Post	High School	3	2	251	Average	14.17	3.38	5.19	4.11	15.65	12.75	21.05	31.39	9.85
						SD	3.84	1.24	2.42	1.84	3.73	2.75	0.67	0.67	1.96
Both	Post	Middle School	2	0	132	Average	9.37	2.38	0.14	5.34	14.01	19.90	26.90	25.14	8.57
						SD	6.96	3.37	0.20	6.76	2.92	3.00	1.52	11.98	2.02

4.6 Results relevant for determining the reliability of the instrument

Table 4.2 shows that obtained results are stable in several different ways. Each of these aspects of stability is elaborated in sections below.

4.6.1 Stability of the results across the different educational levels

An important and expected difference in the results was between different educational levels. Expected differences were related to student's usage of the correct model and with respect to their level of self-consistency. Table 4.2 show that these differences are in the expected direction, namely colleges perform better than high school students and high school students perform better than middle school students. An exception however is in the case of the post instruction tests and the wall context. In surveyed samples high school students on average outperformed the college students. This is not the case with the air context nor is the case when two contexts (all students) are taken together.

If results from different educational levels are compared with respect to the students' self-consistency, results are again different in expected direction in all of the cases but (again) in the case of the post instruction tests and the wall context. In this case high school students were by a small margin more self-consistent than the college students.

It should be noted however that all of these differences between educational levels (in terms of students' self-consistency and in terms of the usage of the correct model) are embarrassingly small for higher levels with respect to the lower ones. Another meaningful pattern of differences that Table 4.2. shows is the that post-instruction results are regularly better than pre-instruction results. Figure 4.7. graphically compares post instruction results at three educational levels as obtained through the air context of the test. The figure shows stable increasing slope when Wave and Intrinsic Models are compared at different levels. Ear-born Model has the opposite trend which (much less pronounced) exists also in the case of the Dependent Entity Model. There is no real pattern of this kind in the case of the Independent Entity Model.
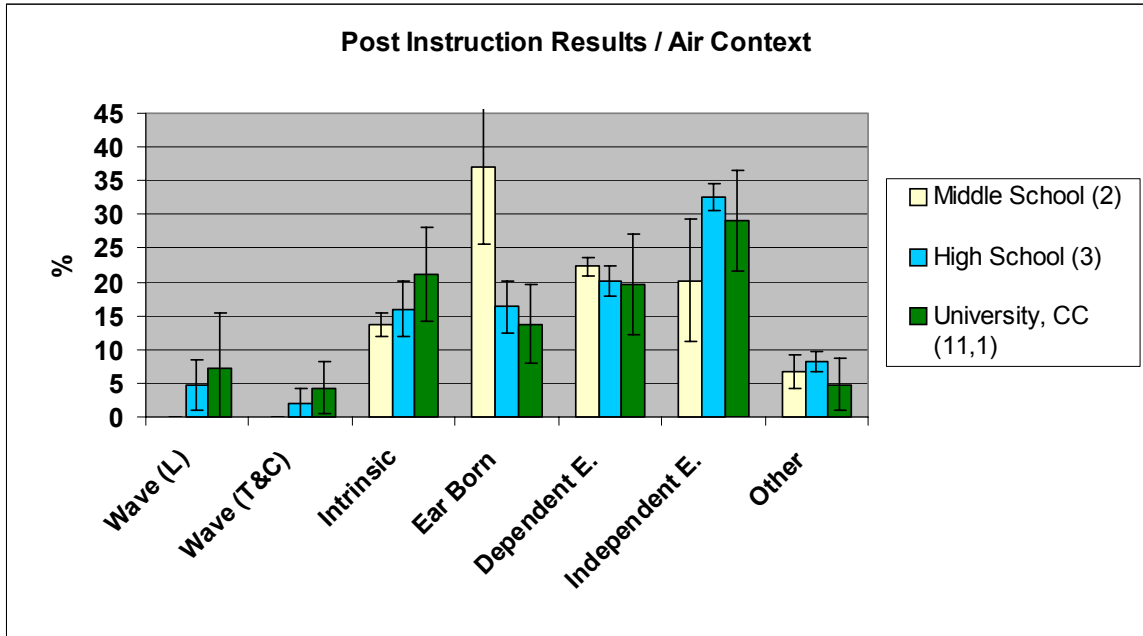
Figure 4.7. Comparison of post instruction results at primary, secondary and tertiary levels as obtained by the air context of the survey

Differences shown in the Figure 4.7. are easier to notice if models that are similar to some extent are grouped together. In this way we can group Wave Models and Intrinsic Model because same answer choices correspond to these models in questions 1, 4, 5, 6 and they are differentiated by dynamics of the particles of the medium in questions 2, 3. Dependent and Independent models have in common that according to both of them sound is a self-standing entity different from the medium through which it propagates. If these two groups of models are clustered together, Figure 4.7. looks as shown in the Figure 4.8.
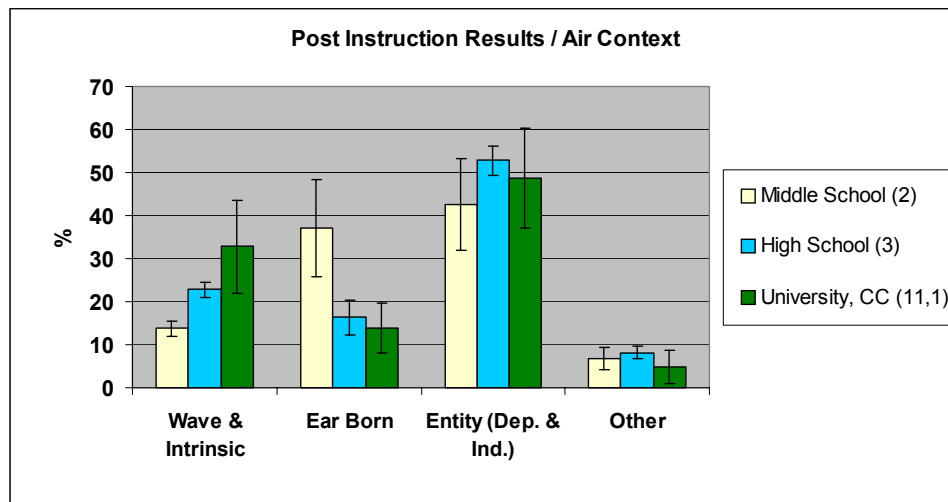
Figure 4.8. Comparison of post instruction results at primary, secondary and tertiary levels as obtained by the air context of the survey (grouped models)

When models are grouped this way, patterns described with respect to the Figure 4.7. become more pronounced. There is an upward slope when Wave and Intrinsic models are compared at different levels. This slope rises from the primary level toward the higher ones. A slope in the opposite direction is associated with Ear-born Model while no definite pattern is related to Entity models.

Figure 4.9. shows results related to wall context so that models are grouped there in the same way as in the Figure 4.8. Here we again have the upward slope associated with Wave and Intrinsic models but it is somewhat less pronounced than in the air context. There is no clear pattern of this kind related to Ear-born and Entity models. However absolute percentages of models in both of these groups are very similar at all three levels.
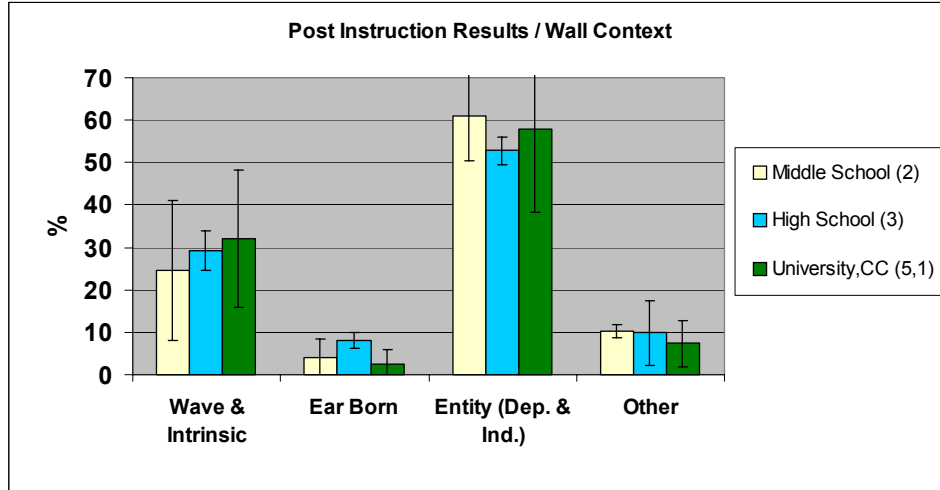
Figure 4.9. Comparison of post instruction results at primary, secondary and tertiary levels as obtained by the wall context of the survey (grouped models)

Figure 4.10 shows results of students in samples that took both air and wall context if all students are taken together and models grouped in similarity clusters. When results from the two contexts are combined, this is done in a way that weighted averages are found for each of the models in each of the contexts within the sample.

Upward slope is here again clear for the correct side of models as well as the downward slope for the Ear-born Model. The percentages of students who use Entity Models at these three levels are strikingly similar (52.04%, 52.44% and 52.71%).
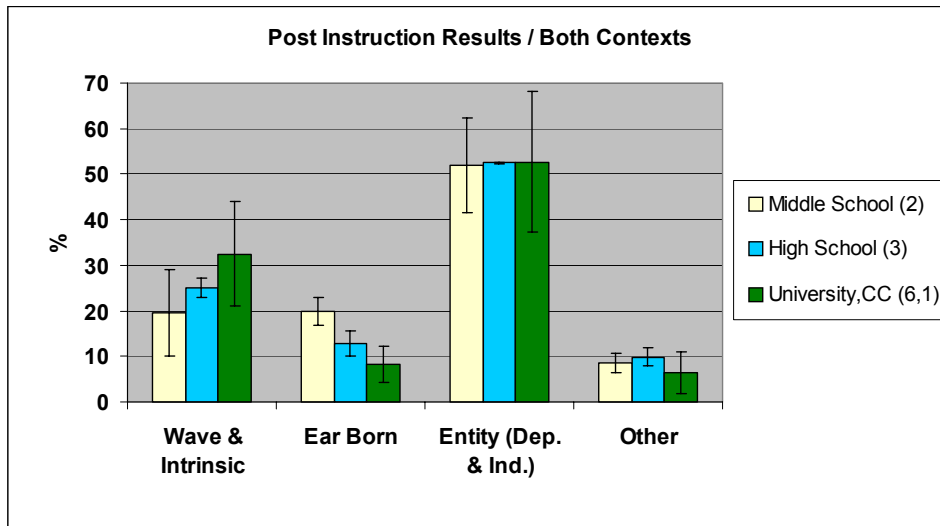

Figure 4.10. Comparison of post instruction results at primary, secondary and tertiary levels as obtained by both (air and wall) contexts of the survey (grouped models)

These results show that the test reliably measures students' progress in terms of their usage of correct models (and models that are close to correct). Ear-born notion of sound is less popular at higher than at lower levels and Generic Entity Model (Dependent and Independent) is very stable and on average does not change much with educational level.

4.6.2 Stability of results within the same institution

Another way to determine if results are distributed in a meaningful way is to look at the difference between results obtained from students at the same institution but enrolled in the courses at different levels. For this purpose students enrolled in concept-based, algebra-based and calculus-based introductory physics courses at Kansas State University were sampled. The expected result was that the students enrolled in calculus course will have the best results and students enrolled in the concept-based course least good results. The obtained results were in accordance with these expectations as can be seen when results of these groups are compared. In the case of the correct model there is a rising pattern that starts with the lowest level course and in the case of the most incorrect model (Independent entity) there is an opposite trend. Models in the middle of the scale are not consistently different.
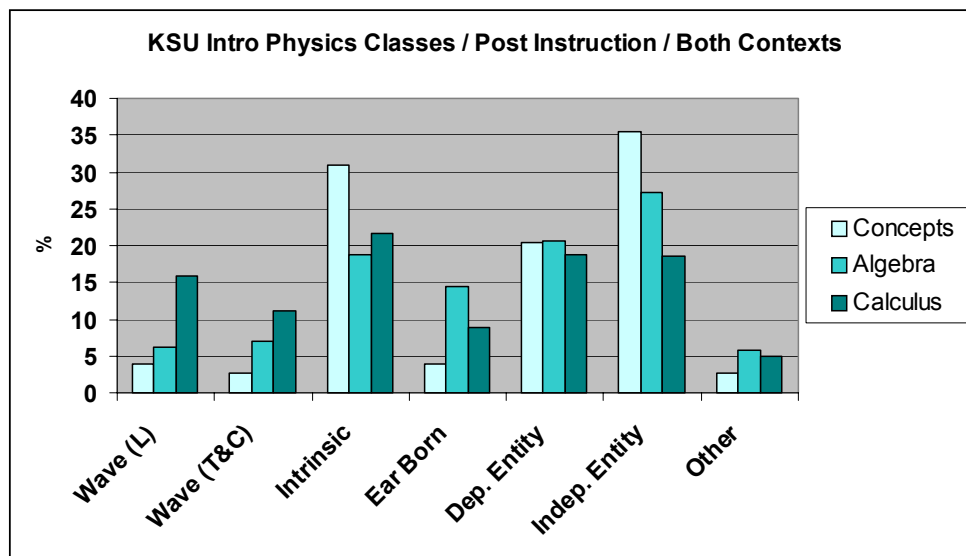

Figure 4.11. Comparison of post instruction results at Kansas State University in spring 2003 as obtained by both contexts

Table 4.3. shows results related to model distribution numerically and also results that pertain to self-consistency of students in these different classes. As shown in the table, differences in self consistency (with respect to Wave or all models) follow the same pattern as distribution of Wave models.

Table 4.3.

Comparison of post instruction results as obtained by both contexts at the same institution (KSU) in Spring 2003 and from classes at different levels (in percentages)

ourse Math Level	N	Consistent (Pure model state)	Consistent Wave	Wave (L)	Wave (T&C)	Intrinsic	Ear-born	Dependent Entity	Independent Entity	Other
Calculus	126	24.60	15.08	15.87	11.11	21.63	8.93	18.85	18.65	4.96
Algebra	207	15.46	5.80	6.28	7.00	18.72	14.37	20.65	27.29	5.68
Concepts	38	0.16	0.03	3.95	2.63	30.92	3.95	20.39	35.53	2.63

4.6.3 Stability of results across the different institutions at the same level

Table 4.2 and figures 4.3 and 4.12 show that for all models except for the correct one, standard deviations between the samples are relatively small when compared to averages. This shows that samples that were analyzed are not very different from each other. This is especially true when one takes into account the relatively small number of the samples and that some of the samples had less than 30 students.
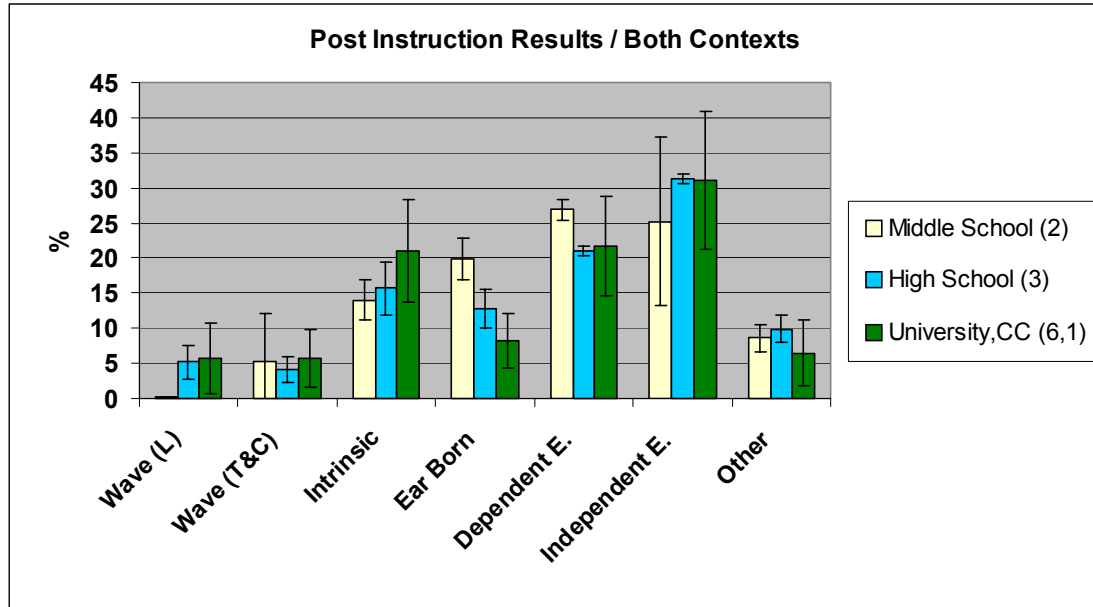
Figure 4.12. Comparison of post instruction results at primary, secondary and tertiary levels as obtained by the both context of the survey together

The most striking resemblance of the results from different institutions at the same level was obtained in the case of the high schools. In the spring semester of 2003 the test was administered after the instruction to students at high schools in Kansas, Minnesota and Croatia. None of the teachers knew about the test during the lessons on sound. Both contexts were administered in each of these samples. In the case of the Croatian sample all students took both contexts of the test and in other two samples each student took one context. The middle column of the three columns that represent different school levels in the Figure 4.12 shows these data graphically. SDs between Dependent and Independent Entity Models as obtained from these three schools are smallest of all models (0.67% each) and the greatest SD is related to Intrinsic Model (3.73%).

Small and relatively small standard deviations between samples at the same levels imply that on average, distribution of students' models is rather predictable i.e. based on these averages and standard deviations, the teacher at any of the levels can pretty accurately determine what he or she can expect in his or her classroom. On the other side, it is possible that the testing itself, in a proposed formative way, may have important instructional value that is worth the time investment in the classroom.

4.6.4 Difference between pre– and the post-instruction test results

In all of the cases when the test was administered both before and after the instruction, post instruction results were better than pre-instruction results. This pattern shows that the test is sensitive to the instructional changes. For the purpose of accurate measurement of the pre- and post-instruction differences each of the samples that were tested in these two instances is separately analyzed and results are shown in the Table 4.4. The difference is presented in terms of the gain (percentage increment of the correct answers) and the normalized gain. Normalized gain is the percentage gain achieved divided by the total possible percentage gain or: Normalized Gain = (posttest% - pretest%) / (100% - pretest%)

Hake (Hake, 1997), argues that a normalized gain is an accurate measure of the effectiveness (or the non-effectiveness) of a particular presentation style. This Hake's average normalized gain is usually referred to as the Hake Factor, h.

Table 4.4.

Results of pre- and post-testing

Institution	Course Math Level	Context	N Pre	N Post	Method	Pre Test Result (%)	Post Test Result (%)	Gain (%)*	Normalized gain (h)**
BOTH CONTEXST (WHOLE CLASS)									
University, NY	Calc.	Air	100	95	Research based	9.50	29.21	19.71	0.2178
University, PA	Algb.	Both	12	10	Lec./Demo/Lab	0.00	12.50	12.50	0.1250
High S., HR	Algb.	Both	49	51	Lecture / Demo	0.51	11.76	11.25	0.1131
University, NC	Calc.	Air	57	19	Research based	0.44	9.21	8.77	0.0881
University, KS	Algb.	Both	175	177	Lec./Demo/Lab	2.57	5.79	3.22	0.0330
AIR CONTEXT									
University, NY	Calc.	Air	100	95	Research based	9.50	29.21	19.71	0.2178
University, PA	Algb.	Air	6	6	Lec./Demo/Lab	0.00	20.83	20.83	0.2083
High S., HR	Algb.	Air	28	28	Lecture / Demo	0.89	16.07	15.18	0.1532
University, NC	Calc.	Air	57	19	Research based	0.44	9.21	8.77	0.0881
University, KS	Algb.	Air	99	98	Lec./Demo/Lab	2.02	4.59	2.57	0.0262
WALL CONTEXT									
University, PA	Algb.	Wall	6	4	Lec./Demo/Lab	0.00	0.00	0.00	0.0000
High S., HR	Algb.	Wall	21	23	Lecture / Demo	0.00	6.52	6.52	0.0652
University, KS	Algb.	Wall	76	79	Lec./Demo/Lab	3.29	7.28	3.99	0.0412

*Gain (G) = (post-test) – (pre-test)

**Normalized gain (h) = gain / (maximum possible gain) (Hake, 1997).

The model distribution of the sample that had the highest gain looked (before and after instruction) as shown in the Figure 4.13.
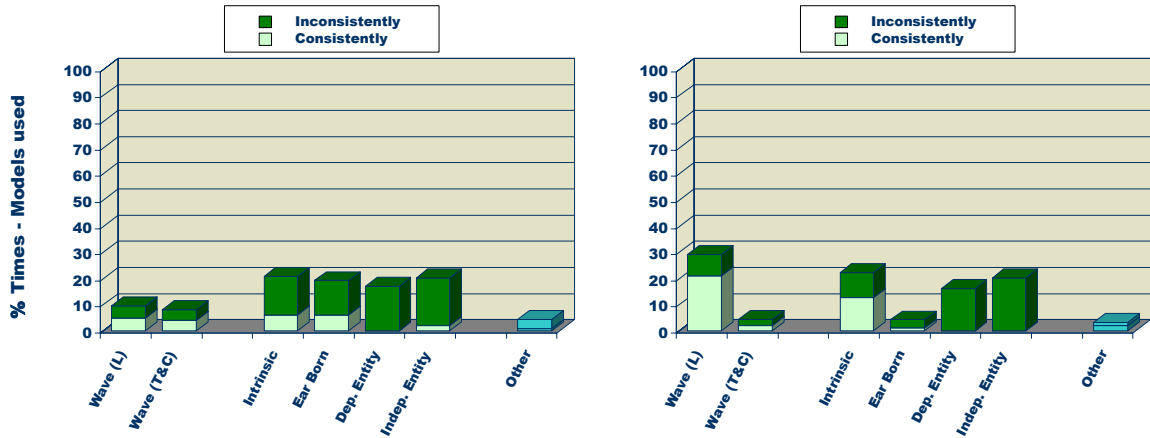

Figure 4.13. The model distribution of the sample that had the highest gain before (left figure N=100) and after (right figure N=95) the instruction. Air context was administered both before and after instruction.

The corresponding graphs that show the dynamics of the particles of the medium are shown in the figure 4.14.
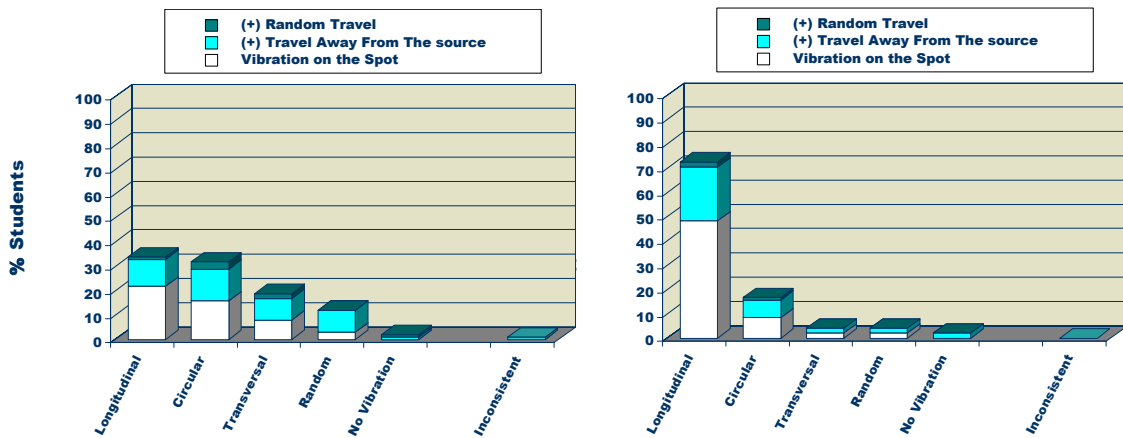

Figure 4.14. The movements of the particles of the medium expressed before (left figure N=100) and after (right figure N=95) the instruction by the sample that had a highest gain. Air context was administered both before and after instruction.

Although results after instruction are better than those before the instruction in each of these samples and also when all samples are compared together (see Table 4.2)

101

overall difference in model distribution among the students at the same level before and after the instruction is far from satisfactory. Figure 4.15, for example, shows model distribution as obtained before and after instruction from all the samples at tertiary level in the air context. Although in this case we compare different institutions (with only three of them common), the comparison is still informative. The percentage of correct models increases from 3.99 % (SD 4.8) before the instruction to 7.28 (SD 8.24) after the instruction. The usage of the other models stays practically the same.
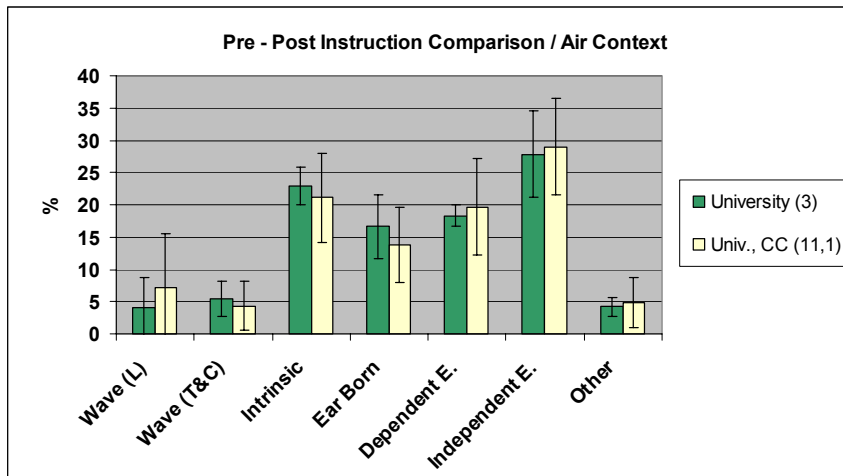

Figure 4.15. Model distribution as obtained from all the samples at tertiary level in the air context.

4.7 Results relevant for determining the validity of the instrument

In building the case for the test validity we use some aspects of the survey results and interviews with students and experts. In the final word on the test validity, we also add arguments built on the nature of the test itself as well as on the process of the test construction.

4.7.1 Validation through the interviews

This section reports on the findings of the interviews that were conducted during the Fall semester of 2003. In the same semester the survey was administered in order to combine

the findings of these two approaches (quantitative and qualitative) to validate and further improve the test.

The same survey version of the test (Appendix F and G) that was administered to a large sample was used in the interviews, too. We also used pictorial representations of the fundamental mechanisms of propagation in air and wall contexts as an additional check-up point of the model (see first graphs in appendices I-1 and I-2).

Several procedures were used in interviews, and in each different case they were combined differently. Students' models were determined through the open-ended questions either before or during the test taking (as a part of think aloud protocol) and through the discussion based on graphical representations of the models (appendices I-1 and I-2). In those protocols in which graphical representations were used, they were used as a last thing in the protocol. Students were taking the test in two different modes: either in a think aloud mode (denoted in the Table 4.5. as "aloud") or silently on their own ("silent"). In the second case discussion would have been conducted either before or after the test or both. The protocols employed are shown in the table 4.5 below:

Table 4.5.

Variations of the research protocols employed in the interviews with students

Student No.	Background			Protocol			
	Course level	Had lecture on sound or waves in general before the interview	Took the test before the Interview	Context	Interview	Test	Graphics
1.	Arithmetic	Waves	No	Air	No	Aloud	Yes
2.	Arithmetic	Waves	No	Air	No	Aloud	Yes
3.	Concept	No	No	Wall	Yes	Aloud	No
4.	Concept	No	No	Wall	Yes	Aloud	No
5.	Concept	No	No	Air	Yes	Aloud	No
6.	Calculus	Waves	No	Air	Yes	Aloud	Yes
7.	Arithmetic	Waves	No	Wall	Yes	Aloud	No
8.	Arithmetic	Waves	No	Wall	Yes	Silent	Yes
9.	Arithmetic	Waves	No	Wall	Yes	Silent	Yes
10.	Arithmetic	Waves	No	Wall	Yes	Silent	Yes
11.	Calculus	Waves	No	Air	Yes	Silent	Yes
12.	Calculus	Waves	No	Air	Yes	Silent	Yes
13.	Algebra	Sound	Yes	Air	Yes	Silent	Yes
14.	Algebra	Sound	Yes	Air	Yes	Silent	Yes
15.	Algebra	Sound	Yes	Wall	Yes	Silent	Yes
16.	Algebra	Sound	Yes	Air	No	Silent	Yes
17.	Algebra	Sound	Yes	Wall	No	Silent	Yes

The protocols were deliberately different one from another because the input often has a significant if not crucial influence on the students' reasoning. By combining the approaches we compensated or counterbalanced different kinds of influences that the protocol itself had on students' answers. Depending on the each particular protocol, students had a chance to verbally express their understanding (a) before the test (without seeing it yet), and/or (b) during the test when they already saw some of the answer choices), and/or (c) in a discussion after the test, and/or (d) with respect to graphical representations of the models that were presented to them. Participants also had a chance to change the multiple-choice items they picked earlier in the protocol if they wanted to.

4.7.1.1 Comparisons of students free answers with their results on the test

In this section, results pertaining to questions Q2 and Q3 (that deal with the dynamics of the particles of the medium) will be reported separately from the results related to all

other questions, purpose of which is to determine the relationship of that dynamics and sound propagation.

Results pertaining to the questions 2 and 3

Questions 2 and 3 together serve to elicit the dynamics of the medium particles during the sound propagation. One of the purposes of the interview was to determine if Q2 and Q3 perform this function in intended way. Students were asked to describe the dynamics of the medium particles during the interview and their interview responses were compared with the responses to Q2 and Q3 on the test. Of 13 students that were interviewed before the test, 7 were able to either partially or completely describe the dynamics of the medium particles. In each of these cases their selection of the answer choices in the test matched their initial statements. In addition to this, students were asked to interpret other choices in questions 2 and 3 (although special attention was devoted to those they picked as correct) so that validity of the choices was further probed and established this way.

We also wanted to determine if students correctly understood the introductory sentence preceding Q2 and Q3. Students' understanding of this introduction was verified so that students were specifically asked about the statement and its meaning to them. This item was validly understood in 16 instances (researcher omitted to ask one of the students about this).

Results pertaining to the questions 1, 4, 5 and 6

The main purpose of this particular interview protocol was to determine whether the mixed state projected by the test actually reflects the reasoning that involves different models or mixed model states, where identified, are possibly due to the invalidity of the test items (i.e. due to the students' interpretation of the choices in ways not intended by the authors). Validity of questions and answer choices that pertain to questions 1, 4, 5, 6 were crucial to answer this research question because these four questions in different ways deal with the relationship of the sound and the dynamics of the medium particles during the sound propagation.

Validating students answer choices was not a straightforward procedure that would involve a simple comparison of students' models of sound propagation before and during the test taking. For the following reasons, this simple comparison was not sufficient:

1. Students may not have a model to begin with (and therefore not the ready ideas to map on the test choices).
2. Students may like more than one model at the same time.
3. Students may change a model while taking the test.
4. Any or all of the above three may combine during the same interview.

These issues caused that validation of the test through the interviews was complex procedure and for this reason in this section we summarize only major findings and implications. An interested reader can find results related to validation of the test through the interviews described in a great detail in Appendices N, N-1, N-2 and N-3.

Findings

Invalid probes are those instances in which a student, for whatever reason, picked the choice that did not correspond to the model that he or she was expressing verbally.

The invalid probe of the model could have happened four times in each of the tests because there are four model-defining triplets in each of the six-question tests. The total number of probes of student's model in interviews was four times the number of interviewees (so 4x17=68).

In some instances a student did not interpret correctly a choice which was not related to his or her model. In these cases the misinterpretation would not have caused the invalid probe.

A total of six model probes (i.e. 8.8%) were deemed invalid based on the interview validity procedure described above. In these cases student's model would have been invalidly interpreted. Six invalid probes that occurred were made by six different students. In three of these instances the probe was invalid because of misinterpreted choice 5a. Additional three students misinterpreted the choice 5a, but that choice was not related to their model so this misinterpretation did not cause the invalid model probe associated with that choice. So choice 5a was misunderstood frequently and it was

always in the same way. Details are elaborated in Appendix N-1. This is the only answer choice related to which a pattern of misinterpretation was observed.

Of the remaining three invalid probes two occurred not because an item was misinterpreted but rather because the statement was misread. These two students in second reading noticed their "mistake" and corrected themselves. Final, 6th invalid probe occurred because of the misinterpretation of the choice 6a. Details related to this misinterpretation are also in the Appendix N-1.

4.7.2 Correlation analysis of answer choices

As a quantitative complement of the validity verification through the interviews, correlation coefficients between all of the answer choices were calculated using data collected through the survey administered to the large sample. The rationale for this procedure was that although we expected that many students may be in a mixed model state, if large sample is taken, the answers that are related to the same model should not have negative correlations. Appendix O shows overall results (in terms of correlations between answer choices) for all the data we collected at three educational levels. Data collected in two different contexts are analyzed separately. The tables in Appendix O have shadowed areas that pertain to the combinations of the choices that correspond to the same model so the reader can easier navigate through the table. Secondary choices for a model (choices that may correspond to more than one model and are rarely used by students), in instances when they exist are shadowed in a lighter gray than primary choices. Primary model choices do not necessarily have the highest of all correlations pertaining to the particular question, (due to mixed model states and due to the secondary choices) but nevertheless, choice combinations that pertain to the same model as primary choices should not be negatively correlated. Negative correlations of the primary choices related to the same model may indicate a possible problem in the way in which students interpret any of those choices. Another indicator of possible problems in interpretation can be a significant correlation between the answers that correspond to different models.

These main points (mentioned above) that we were primarily interested in with respect to the correlation factors in the Appendix O, are summarized in the Table 4.6. Table 4.6. shows correlation factors that pertain to the correct model separately. High

correlations indicate that those students who have the correct model "know what they do" and are not "mixed" a lot. Indicators in Table 4.6 add to the quantifiable results that help to determine possible issues in test validity, but they are also very useful in determining the applicability of the test at a specific level. All the data presented in the Table 4.6 were collected in 2003. For the purpose of determining these correlations results from the pre- and post- instruction tests were taken together but sorted out with respect to the context.

Table 4.6

Identifying possibly problematic answer choices through correlation analysis of the choices – survey results

School level		Tertiary		Secondary		Primary	
Context		Air	Wall	Air	Wall	Air	Wall
N		1132	429	185	115	64	68
Desirable correlations related to the correct model between relationship defining questions (Q1, Q4, Q5, Q6) (6 possible)	Correlation of correct choices is highest in respective question	**6**	**6**	**6**	**6**	**1**	**6**
	Sig. at 5%*	6	6	6	6	0	5
	Sig. at 1%*	6	6	6	6	0	4
Desirable correlations related to the correct model between all questions (15 possible)	Correlation of correct choices is highest in respective question	**15**	**15**	**14**	**15**	**2**	**9**
	Sig. at 5%*	15	14	13	13	0	6
	Sig. at 1%*	15	12	12	11	0	5
Problematic correlations between relationship defining questions (Q1, Q4, Q5, Q6) (180 possible)	Primary choices related to the same model with negative correlations**	1 (1c-5a)	1 (1c-5a)	1 (1c-5a)	1 (1c-6a)	13	8
	Significant positive correlations between different models (at 5% sig. *)	8 Dep. & Indep. models	0	3 Dep. & Indep. models	0	7 Various models	6 Various models

* Two tailed

** In counting these instances we ignored situations when primary choice (or their sum) was negative but secondary choice in the question was the one with highest correlation among those in the particular question.

Table 4.6 shows that at the university level, there is one instance in each of the test versions (air and wall contexts) in which two primary choices are negatively correlated. In both cases this is between choices 1c and 5a. That negative correlation show that students who have the Dependent Entity Model in question 1 will in principle avoid what we considered the corresponding choice in the question 5. This result perfectly corresponds to our findings in the interviews. In the interviews we identified the nature of the problem and through correlation analysis we identified that the problem is present at a large scale. Due to the insight obtained through the interviews and related to the nature of the problem, we were able to address the problem in the next version of the test.

With respect to the strength of the correlations pertaining to the correct choices, university students had a perfect score in both of the contexts. All combinations of choices pertaining to the correct model had highest correlations among the choices in respective questions. Also, all correlations between the relationship defining questions (Q1, Q4, Q5, Q6) were significant at 1% level two tailed. If dynamics defining questions are considered too (Q2 and Q3) all correlation coefficients between the correct choices were highest of all in a particular question and most of them, although not all, were also highly significant. An unexpected result that table 4.6 shows with respect to the university students is related to the number of significant positive correlations that pertain to different models. However in each of these cases mixing occurred only among independent and dependent choices. As said before, these two models do not have a clear boundary and from this perspective this result is not worrisome. In addition, we demonstrated that these two models may hybridize during the test taking into a model that is a combination of the two in which case a student may pick choices petering to either of them. Should this model be addressed on its own is a question that will be discussed in the Chapter V.

Finally, unlike in the case of the correct model which is used by students who likely know what they are doing, Dependent and Independent models are at the bottom of the correctness scale. Students at this end can not be expected to have as stable ideas as those who have the correct model. Combining these arguments with the results from the

interviews, gives a solid ground for the claim that Dependent-Independent mixtures indicated in the right-most column of the Table 4.10 reflect valid mixed states. The program for model analysis of the test results sorts out students that use mixture of dependent and independent entities exclusively. Surprisingly, these mixtures are not pronounced in the wall context of the test at all.

Analysis of the same data with respect to high school students reveals similar issues. All that was said related to correlations between the choices at university level applies here too.

Significant positive correlations between different models in air context were even less frequent at high school level than at university level. Where they existed they were also related exclusively to the mixture of choices pertaining to Dependent and Independent Models. An issue that showed up in the case of the high school students and the wall context but not elsewhere, is slightly negative correlation between choices 1c and 6a which pertain to the Dependent Model. Is this a validity issue? According to the results of the interviews, it is not. In interviews, the choice 6a was raised as an issue only once and in that instance the student interpreted it as a correct answer in a highly sophisticated way (see appendix N-1).

Possible reason for this negative correlation may be related to the attractiveness of the Ear-born Model which in the wall context appears as an option in question 6 for the first time. Because of this late appearance, after seeing that option in the last question, a student who does not have a firm model but likes the ear-born idea may at that point abandon the model he or she earlier went with. Exactly this situation happened in the interview during the case study No. 10 (details in the Appendix N-3). This explanation is plausible because it is the Ear-born choice in question 6 has highest correlation with choice 1a (see Appendix O). Finally, the overall survey results show that Ear-born Model is more popular at lower educational levels which may be the reason why this negative correlation (1c-6a) was not the problem with college students.

Another possible explanation is that negative correlation between 1c-6a happened simply because questions 1 and 6 pertain to different contexts and contextual differences caused the shift away from the original model. The author does not claim that any of these alternative possibilities are true, but he wanted to show that there are a number of

plausible possibilities other than misinterpretation of choice 6a of any kind have caused this negative correlation. But existing data cannot dismiss that possibility either, although none of the other indicators suggest that misinterpretation of any kind caused this negative correlation (of choices 1c and 6a in the wall context, high school level). User of the product (test) may wish to use wall context with caution at high school level until more research is done at this level and with respect to this context. Other indicators except this one suggest that students are even more self-consistent in the wall context than in the air context at high school level. This is demonstrated through higher correlations of correct answers in the wall context and absence of the significant positive correlations pertaining to different models.

When middle school students are considered, the results show evident need for further study on applicability of the test at this level. Some encouraging results were obtained in the spring semester of 2004 and these will be discussed in Section 5.4.2.

4.7.3 Post survey test modifications and validations

In the post survey phase, issues that were identified in the survey testing were addressed and validity of the new version was verified again. The test choices were improved based on the qualitative and quantitative results that were collected in the survey phase and based on the direct suggestions that students gave during interviews. The modifications were made primarily to address the problem with answer choice 5a but question 6 was also modified to avoid possible issues with understanding of the answer choice 6a. The changes that were made are elaborated in detail in Appendix P.

This new test version was additionally validated in three ways:

1. Through additional round of expert reviews in which we followed the same procedures as before.

2. Through verification of positive change of earlier problematic correlations.

3. Through role-playing validation in which experts in physics assumed the roles of students having models that test probes for and took tests that way

Based on experts' suggestions a few additional minor changes were made in the test before it was once again administered to students. Obtained results showed favorable

change in correlations of answers 1c-5a and 1c-6a that we were aiming at. The result of the role-playing validation was that all of the experts straightforwardly picked choices that were corresponding to the models they were "assigned to". Details of these procedures and related findings are in the Appendix P. Based on these results the final version of the test was made (see Appendices S-1 and S-2).

4.8 Reliability of the test addressed

Results presented in the Section 4.7.3. show that in the post survey versions of the test, weak points (choices 5a and 6a) of the survey test version were addressed, while other relevant parameters stayed the same as in the survey version and no new problematical issues aroused. This gives the ground to use results obtained with the survey version of the test (8.9) as a basis for conclusions about validity and reliability of the final version of the test (9.2). It further makes plausible the claim that, had the final version of the test been administered to a large sample as the survey version was, the results would have been the same or superior to those of the survey version of the test.

Reliability pertains to the degree to which a test consistently measures something. Results presented in previous sections show that the test results obtained from the large sample are stable in several different ways each of which contributes to the case that the test is a reliable measurement tool. This reliability of the test is reflected through:

1. Meaningful correlations between the answer choices (at secondary and tertiary levels but not at primary level) as demonstrated through:

   ➢ Positive and frequently significant correlations between the choices in different questions that pertain to the same model

   ➢ Absence of the significant positive correlations between the choices that pertain to different models. The only exception here is the combination of the Independent and Dependent Entity Models because the level of dependence of the entity on the medium is a continuum. Responses of the same student may vary under different conditions or different contexts. However even these variations were very few in the case of the test version 9.0 as well as in the wall context in general.

112

2. Stability of the results across the different institutions at the same level as reflected through the small standard deviations around the average percentages at each of the models is represented in each of the samples.

3. The expected direction of differences between results in terms of the usage of the correct models and in terms of the students' self-consistency. Correct models and self-consistency are more frequent among students:

   ➤ at higher educational levels than on lower,

   ➤ at more advanced introductory physics courses at the same instruction than on the lower ones, and

   ➤ after the instruction than before it.

Because of the lack of the meaningful correlations between the answer choices at primary level, at this point we use above mentioned results to demonstrate reliability and validity only at secondary and tertiary levels.

Meaningful correlations between the answer choices indicate that content sampling error is not an issue in this test. The content sampling error is further reduced by probing a single model multiple times in this test.

The second and the third reliability indicator listed above show that the test is resistive toward the occasion sampling error. Examiner error, the third of the four reliability threats, is not measurable and it was reduced through the standard introduction. Finally the scorer error was not the issue at all because of the computerized analysis of results. This closes the list of the threats to the test reliability. Because all four of the treats to the reliability of the test were well addressed in the study, this gives the ground for the claim that the test is a reliable instrument.

4.9 Validity of the test addressed

A conclusion that can be made based on the Section 4.8 and results that support it, is that the test reliably measures something. The concept of validity determines if this something is what it is supposed to be. Here we summarize results that support the claim that the test validly measures mental models of sound propagation. The test validity is

demonstrated below with respect to the three aspects of validity (content, criterion and construct) as well as through the steps taken in the test construction process.

4.9.1 Primarily content-related validity verifications

A case that the test addresses the content-related aspect of validity will be made based on the table of (content) specifications, experts' reviews of the test and the argument of the face validity. Instructional sensitivity is a feature of the test that adds to this case too.

Table of content specifications

 A table of specifications is a two-dimensional chart. It lists in the vertical dimension the content areas to be addressed by the test and it lists in its horizontal dimension the categories of performance the test measures (Oosterhof, 2001). Categories of performance associated with it are: (1) concepts and (2) information so horizontal dimension of the content-related table of specifications consists of these two categories. In vertical dimensions are "content areas to be addressed by the test". Content-wise, questions of this test serve to probe the understanding of the sound, which is a longitudinal mechanical wave. The test covers only one content area and probes its different aspects. Therefore it is those aspects that are listed in the vertical dimension of the table of specifications.

Table 4.7.

Table of content specifications

	Concepts	Information	Totals
Determining the **mechanism of sound propagation** in the air/wall?	1		1
Determining how do particles of the medium **vibrate**, if at all, while the sound propagates?		1	1
Determining how do particles of the medium **travel**, if at all, while the sound propagates?		1	1
Determining what does this motion have to do with sound propagation – **cause and effect relationship**?	1		1
Determining what does this motion have to do with sound propagation – **time (temporal) relationship**?	1		1
Determining what happens with sound propagation **in the vacuum**?	1		1

The table 4.7 shows that test questions comprehensively cover aspects of the phenomena called sound propagation. The kind of knowledge probed through questions two and three was classified there as information and this should be understood conditionally. It is information more than a concept but the dynamics of the particles of the medium is aspect of sound propagation closely tied to the mechanism (concept) of propagation.

Experts' reviews of the test

Eight experts reviewed the test and they verified that there is only one correct answer in each of the questions. This procedure verified that test does address the correct model and also validated the choices that correspond to it. Results of these expert validation procedures are described in sections 4.5.2 and 4.7.3. All other models were also verified in a similar manner by other experts who did the role-playing validation (Section 4.7.3). Role-playing procedure showed that test validly addresses all models and not just the correct one.

Face validity of the test

Face validity is not actually validity but rather its artificial aspect. Face validity exists if it *appears* that test measures what it supposedly does. If face validity is not a feature of the test, this may cause issues with the way in which participants take and give answers to the test. The argument that this test has the face validity is based on experts' reviews of the test (sections 4.5.2. and 4.7.3), procedure of role playing (section 4.7.3) and on interviews with students (section 4.7.1). All participants in these protocols took the test and gave their feedback and in none of those instances any issue with face validity appeared.

Instructional sensitivity

Instructional sensitivity of the test that was demonstrated in the Section 4.6.4. These results show that what the test measures is related to what students learned in a meanwhile. Because the topic of interest in corresponding lectures was sound

propagation (sometimes exclusively, sometimes as a part of the broader topic of mechanical waves), this also builds up the case for the test's content –related validity.

4.9.2 Primarily criterion-related validity verifications

To demonstrate criterion validity we employed think aloud interview protocols with students and role-playing validation with experts. Results of these two procedures were 91.25 of valid probes in the students' case (see Section 4.7.1) and 100% in the case of the expert's role playing (see Section 4.7.3. and Appendix P). Favorable correlations between answer choices demonstrated in sections 4.7.2. and 4.7.3. also contribute significantly to the case for criterion- related validity.

4.9.3 Primarily construct-related validity verifications

"Construct validation consists of building a strong logical case based on circumstantial evidence that a test measures the construct it is intended to measure (Hanna, 1993 p.402). Previous research and the table of (construct) specifications will be used to demonstrate construct-related validity.

Construct built on the previous research

We show here that the psychological construct that we are addressing in this test is – a mental model. One way to do this is based on the previous research. This test is eliciting constructs that were identified in an earlier study (Hrepic, 2002; Hrepic et al., 2002). The authors of this earlier study thoroughly elaborated why conceptualizations of sound propagation that were identified and described as mental models are – mental models. Namely they fulfill general criteria for mental models that were listed in the Section 2.3.2. of this dissertation and in addition they satisfy specific requirements that define a mental model outlined by diSessa (2002a, 2002b): They

1. Involve the strong "base descriptive vocabulary" e.g., spatial configuration of identifiable kinds of things.

2. Involve only a small, well defined class of causal inferences i.e., just a few principles (e.g., "gears work by conveying motion via contact" or "resistors work by Ohm's law".

3. Allow explicit hypothetical reasoning e.g., "if this gear moves that way then the connected gears move ...".”

Table of construct specifications

Case for construct-related validity can be further strengthened through the table of specifications related to a mental model as a psychological construct. Analogously to the table of content specifications, this table will list in its horizontal dimension specific requirements that define a mental model mentioned above. Vertical dimension lists the same items that the test probes in each of the questions (the same ones listed in the table of the content specification).

Table 4.8.

Table of construct specifications

	Determining the things involved	Determining how system works	Probing hypothetical reasoning	Totals
Determining the **mechanism of sound propagation** in the air/wall?	1			1
Determining how do particles of the medium **vibrate**, if at all, while the sound propagates?		1		1
Determining how do particles of the medium **travel**, if at all, while the sound propagates?		1		1
Determining what does this motion have to do with sound propagation – **cause and effect relationship**?		1		1
Determining what does this motion have to do with sound propagation – **time (temporal) relationship**?		1		1
Determining what happens with sound propagation **in the vacuum**?			1	1

The table 4.8. shows that each of the three requirements are satisfied by this test at least once.

4.9.4   Other validity-strengthening procedures based on item development

Haladyna (1999) describes validation procedures primarily from logistics viewpoints: validity from item development and validity from analysis of item responses. A set of practical suggestions about item development suggested by this author were followed in construction of the test. The procedure of the test construction described in chapters three and four was another validity strengthening component. Because choices in the test were based on students ideas identified in earlier studies. In the pilot testing phase the open-ended questionnaire with small number of questions was combined with comprehensive semi-open-ended questionnaire to determine if anything was missed in terms of students' models or in terms of optimal contextual situations for their eliciting. After choices were written, they were first probed with additional "None of the above" and "More than one of the above" options. Finally the test was validated quantitatively and qualitatively.

In the Section 4.9. a full range of arguments were presented to show that this test is an assessment tool that validly addresses students mental models of sound propagation at high school and college level. This is especially true because of the proposed formative use of the test. Test validity is not an attribute of the test, but "of the interaction of a test with a situation in which the test is used to make decisions" (Hanna, 1993, p. 382). And based on this test a decision about optimal instructional approach related to sound propagation can be confidently made.

CHAPTER V
CONCLUSION

Based on the results in the previous chapter, conclusions will be drawn related to applicability of the test as a formative assessment instrument. After that, instructional implications of these findings will be discussed and applications of the test suggested.

5.1 Using the test

The test is a formative real-time assessment tool. Each of two contextual versions (Air and Wall) of the test consists of 6 questions. For a quick formative assessment all six are not necessary. The minimal number of probes needed to determine students' models and models states is two. This translates into four questions as minimal number of questions needed. For the purpose of obtaining all desired information (models and model states) relatively quickly, air context of the test has been divided into two shorter tests. Both of them have questions 2 and 3 in common because these two questions determine the dynamics of the particles of the medium. Out of the remaining four questions in the full test version two are assigned to each of the shorter test versions so that questions 1 and 4 are paired together and questions 5 and 6 are paired together. Therefore one of the two shorter test versions consists of questions 1, 2, 3, 4 and the other one of questions 2, 3, 5, 6. They will be labeled accordingly as Q1234 and Q2356 test versions. That name perhaps is not the simplest possible but is informative and leaves no doubt about which version is which. The optimal way of using Q1234 and Q2356 test versions is as a real-time, in class assessment. In order to perform a real-time assessment, in addition to the test, the teacher needs a class response system, computer and software for analysis of results. The accompanying CD contains programs for analysis of each of the test versions (Air, Wall, Q1234 and Q2356). Instructions related to usage of these programs are in the Appendix U. The programs were adapted for usage in real time with class response system called Personal Response System or PRS and can be easily adapted to

other systems if needed. Instructions related to usage of the programs in real time with PRS are given in the Appendix U-1.

5.1.1 Final package

Although the test is the central product of this dissertation, several other products facilitate the usage of the test. These include templates for data entry and scoring, programs for model analysis of results, programs for statistical analysis of results and programs for presentation of findings. All tests and programs for their analysis can be found on the CD that accompanied the dissertation. The final version of the test was labeled 9.2 and consists of the full, or 6-question, tests in air and wall contexts as well as of two shorter versions pertaining to the air context (Q1234 and Q2356). These tests are on the accompanying CD in the folder no 1.

In addition to tests, the CD contains folders with following files:

- Templates for data entry and scoring
- Programs for model analysis of results
- Programs for statistical analysis of results
- Templates for presentation of findings

Downloads and updates related to these files are also available at KSU PERG web page and at the author's web page (addresses of these web sites are in the appendix W). These web pages will contain current files related to tests, programs for data analysis, programs for representation of results and, if needed, updated information related to the test usage. Test and related programs are free for use by teachers in their own educational setting.

5.1.2 Applicability of the test at different levels

Table 4.6 in section 4.7.2 shows a number of undesirable correlations related to the survey data collected at primary level (which is not the case with the secondary and tertiary level), especially in the air context. This might be an indicator that the test might be demanding for this age and therefore not applicable. Situation was not much better

when the test was administered to middle school level students as a pre-instruction test during the post-survey phase. However when middle school students took the test after the instruction (which aimed at eliciting alternative mental models), the improvement with respect to usage of the correct model and students' self-consistency was surprisingly large. Percentage of students who consistently used a model increased from 1.33% (before the instruction) to 9.9% after the instruction. More importantly, Longitudinal Wave Model became the most frequently used model of all, with 5.5% students using it consistently (out of 9.9% total). Figure 5.1. shows the model change as obtained from this sample before (N=75) and after the instruction (N=99).
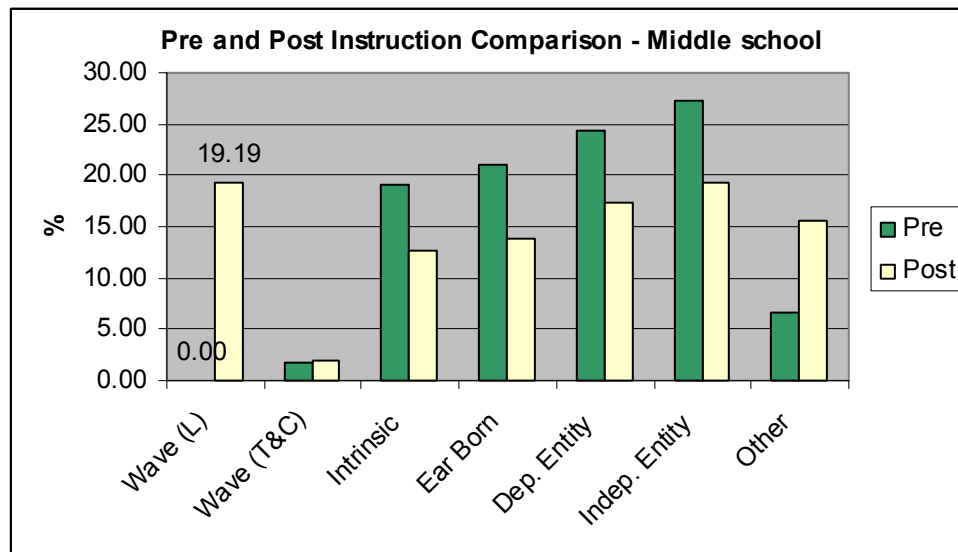

Figure 5.1. Model change at middle school level as obtained after model targeted instruction

The learning gain obtained from this middle school sample (19.19% unmatched and 18.57% matched) was one of the highest observed in this study (see sections 4.6.4 and 5.3). These results show that the test might be applicable also at the middle school level in some form but more research is needed related to this. Another reason for not abandoning the middle school level too soon is the fact that correlations at this level were based on a far (roughly 10 times) smaller sample than for the college students.

Results obtained at the high school and college level that pertain to reliability of the test are similar and both are very good. Unlike at the high school level, at the college

level validation of the test was done also through the interviews with students. The implication is that the test can be safely used as a valid instrument at the college level and reasonably safely at the high school level (had we done the same validity procedure through the interviews also at the high school level, that would enable us to give the recommendation with the same level of confidence for the both of these levels). The test shows promising results also at the middle school level but results collected so far are inconclusive and it is not clear whether students at this level validly interpret the test items.

5.1.3 Test limitations due to the multiple-choice nature of the test

Several limitations of the test became evident throughout the test validation process and these are to a large extent unavoidable with any multiple-choice instrument.

➢ The test does affect students understanding in a way that test items (questions and answer choices) play significant role in the model dynamics and change.

➢ The test taking strategies (that are otherwise meaningful and effective) may obscure test results.

➢ The test may project no model state as a mixed model state and possibly even as a pure model state if a student picks and sticks to a model.

➢ Students may change their opinion without being aware of this change.

Although these limitations are typical for the multiple-choice tests, they deserve to be mentioned because the user should have them in mind when interpreting the test results. Details related to above mentioned limitations observed in this test can be found in Appendices N, N-1 and N-3.

5.2 Interpreting test results

Four issues should be mentioned concerning test usage. The first is related to usage of the test as a summative assessment and second one to observed issues in model mixing. Further, issues related to usage and interpretation of the data obtained through different

contexts need addressing here and finally the difference between meaningful data and the random distribution of answers will be shown and discussed.

5.2.1 Issues in using the test as a summative assessment

A word of caution is in place with respect to usage of the test as a summative assessment. Based on the test results, points could be assigned in two ways. Either for each correctly answered question or for each correct model probe (a question triplet). Grading the test such that each answer is assigned a point may overestimate the student knowledge because a student who answered any of the questions correctly may not be associated with a correct model. Another possibility is to grade triplets. In this case we grade actual number of correct probes. However if done in this way, grading may underestimate student's knowledge because if a student makes a mistake in questions Q2 or Q3, he or she will get no points no matter what the answers on other questions are. This approach is therefore questionable also from the ethical perspective.

Therefore our suggestion is that if a teacher wants to use the test as a summative assessment, he or she should do it in a way that each question is graded on its own. This approach may cause an overestimate of students understanding and whether possible overestimate of the results is acceptable for the intended purpose or not is something a teacher needs to decide for him/herself.

5.2.2 Issues in model mixing

5.2.2.1 Different kinds of mixed model states

Due to variations in mixed model states that were observed during the validation interviews (see Section 4.7.1 and appendices N and N-3) it is possible to classify mixed model states into several categories. There are:

- ➢ Simultaneous mixed model state
  - o Nonselective
  - o Selective
- ➢ Consecutive mixed model state

In the case of a simultaneous mixed model state (MMS), a student finds more than one model attractive while answering the same question. This situation can have two variations. Nonselective simultaneous MMS (which is in fact no model state) is a situation when a student finds all of the choices plausible. For example, student No. 11 (see Appendix N-3) said at one point of the interview:

S: …since I don't know anything about it, any one of them could be right. I could be totally off. For what I don't know. So I just picked the one that sounds the most right to me.

However even in these cases we did not find a student that found all of the choices *equally* plausible. The same student mentioned above (No. 11) at the beginning of the discussion about his answer choices was asked how did he go about answering the test and he answered:

S: Oh, I kind of went off with what I was trying to say before. It's a written form this time. I kind of found myself trying to answer, trying to circle two of them in few of the questions.

Answers that student 11 picked in different questions corresponded to different models so the test projected him into mixed model state. However a student who likes several models may possibly prefer one and stick with it. An example for this situation is student No. 9, who said at one point:

S: And, actually like almost all of these answers I agree with, like every single one of them, but I just picked one that I kind of agreed more with. Because they all make sense.

This student misread one word in one of the choices and in the second reading corrected herself so her final answers corresponded to a single model i.e. a pure model state. The example of student No. 9 illustrates a disadvantage of the test which is that it may project the mixed model state (or even no model state) into a pure model state. On the other hand, it is reasonable to claim that if a student decides to stick with a single model four times in a row, although he or she may not be sure about that answer, s/he has a very strong preference toward that model. Therefore an argument could be made that

projecting the student with this attitude into a pure model state (that corresponds to his or her chosen model) is probably the most valid analysis of the result. The student No. 9, to continue with the same example, expressed the same model that she picked in the test choices also verbally before the test although she was not sure about it during all this time.

Selective simultaneous MMS is a situation when the student favors some of the choices and clearly rejects others: For example: "So it's either d) or b)". A good example for this is student No. 1 (see Appendix N-3). Throughout the test, to a different degree, this student liked three different models. He was well aware of the difference between them and in various instances different factors determined the model he settled on. This example also shows the need of probing the models with multiple questions.

Consecutive mix is the situation when a student subscribes to only one model at the time, but this model changes at some points in the test. An example is student No. 3 She at one point during the test consciously decided to change the model with which she agreed.

5.2.2.2 Mixture and Dependent Entity and Independent Entity Models

In section 4.5.4 we discussed the nature of the sound entity in students' responses as laying on a continuum of "dependency". An example of the answer in the middle of this continuum would be that the air "helps" the sound to propagate. However, because the researchers' definitions of these two models are polarized, so are the corresponding answer choices. For this reason some students that lay somewhere in the middle of this conceptual continuum could have possibly mixed the choices that correspond to Independent and Dependent Model. In the data analysis graph related to model states (figure 4.5.) we separated out students who mix exclusively Dependent and Independent Models. This serves to inform the instructor about the possible number of students that might have been displayed as being in the mixed model state because of being in the middle of dependency continuum related to Entity Models. Another reason for this separation is a Dependent-Independent Hybrid Model that appeared in several instances *during* the test taking (appendices N and N-3).

5.2.3 Differences between contexts

Students at 6 universities and one community college took both the air and wall contexts. Figure 5.2 shows results from these 7 samples in a way that results for each context are put one next to another. All results were collected after the instruction.
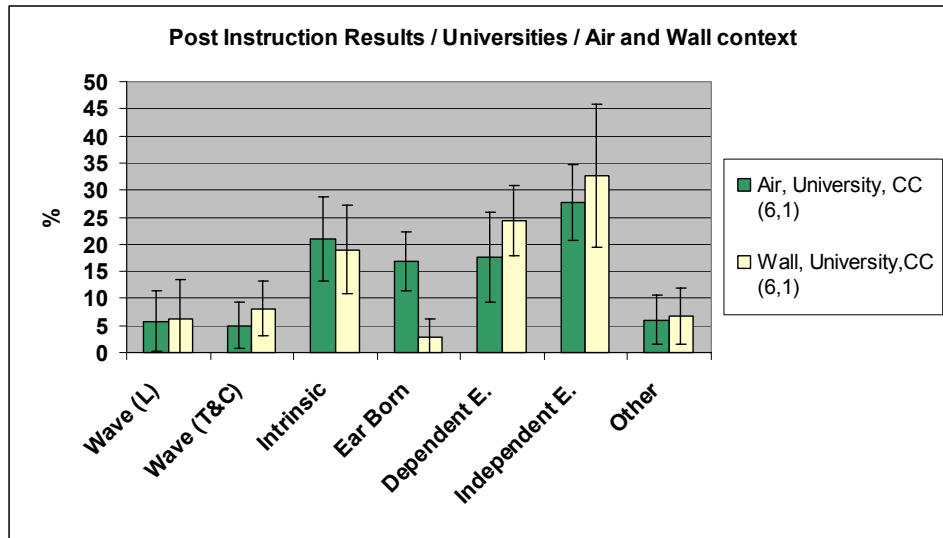

Figure 5.2. Comparison of the results obtained through different contextual versions of the test at university level

If Wave and Intrinsic Models are analyzed, Figure 5.2. shows certain variations in the height of each of the columns corresponding to these models. However, the sum of the Wave models and Intrinsic Model (models that share the same answer in questions 1, 4, 5, 6) is 31.63% and in the wall context 33.28% -- almost the same. It is reasonable to assume that from the pool of students who had any of these (Wave and Intrinsic) models, nearly an equal number took each of the contextual versions of the test because there was no pattern in the distribution of the tests. Then variations in each of the model columns shows that among the student with generically intrinsic models (Wave and Intrinsic) contextual features will cause the differences with respect to the dynamics of the medium particles. Within the air context, traveling of the particles in the direction of the sound propagation is twice as popular as vibration at the place. Within the wall context the difference is much smaller because vibrations are here more popular than in the air and traveling of the particles is less popular.

Because the percentages of unclassified models (other) are very close in these samples, the percentages of any of the remaining models are also almost the same (62.29% in air context and 59.94% in wall context). Of these models Ear-born is much more popular in the air context than in the wall context. The difference is "compensated" in a way that both Dependent and Independent Entity Models are more popular in the wall context. This result is related to the nature of the test. As explained in section 4.5.1.2, unlike in the air context in which the Ear-born choice is offered in each of the model-defining questions (1, 4, 5, 6), in the wall context, the Ear-born choice is offered only in the question 6. In questions 1, 4 and 5 the Ear-born choice is replaced with a propagating air choice. In analysis of models (see appendix L), if a student is not self consistent and picks propagating air choice in any of the questions 1, 4, or 5, the program "consults" question 6 to determine how is the motion of air related to nature of sound. For this reason, a random chance of getting the Ear-born Model in air context is more than twice as big as in the wall context (15.2% versus 6.62%). The inability of students who like the Ear-born option to choose it throughout the test in the wall context *on average* shifted them toward Entity Models so both of these models "gained". A test user should be aware of these differences between contexts.

In two of our samples the same students took the air and wall tests back to back. The results obtained from the community college students show all the patterns described above in exactly the same way only with bigger differences. All students took the wall context before the air context. These results are shown in Figure 5.3. In this case, the sum of the Wave and Intrinsic Models in the air context is 13.16% and in the wall context 11.84% -- again very close. The sum of the remaining models is 80.26% (air context) and 81.58% (wall context) -- almost identical. Percentages of unclassified models are the same (6.58%) in both cases.
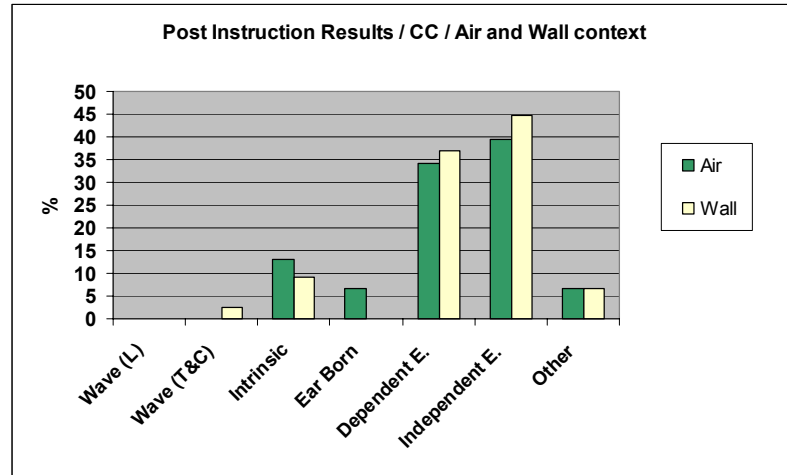
Figure 5.3. Comparison of the results obtained through different contextual versions of the test from the same students at community college level

A variation of these trends can be seen in the results obtained when high school students took two versions back to back. In this case half of them first took the air context and another half first took the wall contest. In this case both Wave Models as well as Intrinsic Model were more popular in the case of wall context. And in the case of the remaining two context independent entity was (unlike with college students more popular in the air context).
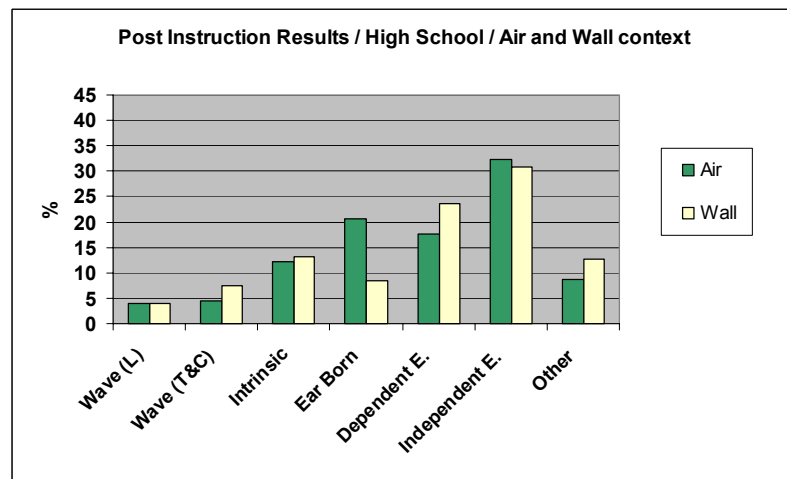

Figure 5.4. Comparison of the results obtained through different contextual versions of the test from the same students at high school level

If dynamics of the medium particles is analyzed per se (regardless of the model), in each of the above comparisons vibration of the spot is much more popular in the wall context than in the air context. The implication for the user is that:

➢ Due to contextual features, vibrational dynamics of the medium particles tends to be more popular that translational dynamics in the wall context (than in the air context).

➢ Due to the nature of the test, the wall context tends to underestimate the Ear-born Model and to the same extent overestimate Independent and Dependent Entity Models.

5.2.4 The issue of random distribution

Model distributions as obtained by students reflect to some extent a random distribution of models i.e. the distribution of models that would be obtained as a result of all possible answer combinations (see appendix L-1). This opens two questions:

1. How do we know that the obtained results are not random?
2. How can a user know that what he or she obtained from students is not some random distribution i.e. how to know if students are not serious?

To this pool of evidence given in Chapter IV we can add the z-test of the difference in results of each of the answer choices if random "sample" is compared to the actual one. Table 5.1 shows these results in the case when the sample consists of all college students that took the air context of the survey version of the test. Out of 30 answer choices all but two are significantly different at 5% level of significance two tailed.

Table 5.1.

Z-test of the difference between the random "sample" and college student sample that took the air context of the test

Z test	Q1	Q2	Q3	Q4	Q5	Q6
a	8.9766*	27.9068*	39.1408*	24.8025*	22.8600*	-5.6090*
b	-8.2510*	8.4007*	-8.1381*	-4.7828*	3.1652*	-17.0224*
c	7.4140*	-7.1129*	-1.4345	-3.2315*	-7.4281*	10.3067*
d	-0.7486	7.3080*	-8.7566*	3.5024*	-14.6621*	12.2137*
e	-2.9481*	-15.2319*	6.9940*	-6.7953*	16.1776*	14.9181*

* $p < 0.05$ %; N1=15625; N2=1132

In the case of the wall context taken also by college students, all but 7 answer choices are significant. These results are shown in the table 5.2.

Table 5.2.

Z-test of the difference between the random "sample" and college student sample that took the wall context of the test

Z test	Q1	Q2	Q3	Q4	Q5	Q6
a	2.4087*	18.2799*	19.0782*	11.6587*	14.9209*	-3.9759*
b	0.2895	3.3683*	-14.3863*	-3.4437*	-1.9212	-8.6480*
c	3.9668*	-3.1477*	4.8722*	-1.9212	-5.8013*	5.2188*
d	-1.7554	7.9099*	0.6845	-1.1669	-3.4437*	6.6595*
e	-3.5878*	-11.0489*	5.6369*	-0.0238	6.0728*	8.1049*

* $p < 0.05$ %; N1=15625; N2=429

If all students who took the survey version of the test are taken together (college, high and middle school) results are even more dramatic. In the air context, all answer choices are statistically significantly different from the random "sample". And in the wall context all but 5 choices are statistically different.

Above mentioned evidence show that students in our samples were not "gambling" although at the first sight there are some similarities of the obtained model distribution and the random distribution. However, none of the above mentioned evidence is suitable to determine quickly, in real time if students are "messing around" or are seriously doing the task. Three different indicators are suitable for this purpose so we

will suggest them as the answer on the second major question of this section: How to know whether students are seriously taking the task or not?

These three indicators are the differences between the obtained results and the random results with respect to:

➢ Frequency of correct model - should be significantly *larger* in real data

➢ Frequency of unclassified models ("Other") - should be significantly *smaller* in real data

➢ Frequency of self-consistent students - should be significantly *larger* in real data

Examples (and also evidence) of the differences in these three frequencies in the survey data are here again shown on the sample of college students that took air context during 2003 (survey test version). The sample is chosen as the example because it is the largest reasonably homogeneous group of data we have. Table 5.3 shows random distribution of models and the distribution of these models obtained from this group. In addition to percentages of models as found in this sample, factors that compare those percentages with corresponding random percentages are given in the table.

Table 5.3

Comparison of random model distribution and model distribution obtained from college students in air context

AIR RANDOM	N=15625	Wave (L)	Wave (T&C)	Intrinsic	Ear Born	Dep. Entity	Indep. Entity	Other
	Consistently	0.01	0.01	0.10	0.12	0.29	0.61	20.93
%	Inconsistently	0.79	1.59	11.90	15.08	14.97	24.92	8.67
	Total	0.80	1.60	12.00	15.20	15.26	25.54	29.60

AIR COLLEGES	N=1132	Wave (L)	Wave (T&C)	Intrinsic	Ear-born	Dep. Entity	Indep. Entity	Other
	Consistently	5.65	1.86	5.39	0.00	0.35	1.77	1.06
%	Inconsistently	3.36	3.14	16.74	12.68	19.63	25.51	2.87
	Total	9.01	4.99	22.13	12.68	19.99	27.27	3.93
Factor with	Consistently	441.70	144.93	56.13	0.00	1.20	2.88	0.05
respect to	Inconsistently	4.26	1.98	1.41	0.84	1.31	1.02	0.33
random	Total	11.24	3.12	1.84	0.83	1.31	1.07	0.13

Table 5.3 shows that percentage of students that express the correct model correctly is 11.24 times larger than the random probability for this. And percentage of

those who use the correct model consistently is about 440 times larger than the random chance. Chance that this happened randomly is close to none. On the other hand random chance that students' models will be classified as "other" is 7.53 times bigger in random distribution than in the real data.

The next Table (5.4) illustrates the difference between students' self-consistency and the random probability for self-consistent answers. The table shows that students are on average 13 times more self consistent than expected from random distribution. If Wave Models (Longitudinal, Transversal and Circular) are combined, this factor is 293.3, and for all other models it is 6.67. Differences this large are unlikely to happen by change. Another point worth mentioning here is related to difference between consistent students who have correct (or nearly correct) models and those who have incorrect ones. Data shown in the table 5.3 show one more time that students who are further away from the correct model will more likely find options related to different models attractive.

Table 5.4.

Comparison of random probability for self-consistency with results obtained from college students in air context

AIR RANDOM	Consistent (Pure Model State)	Consistent Wave (L+T+C)	Consistent Other Models	N
%	1.15	0.03	1.13	15625

AIR COLLEGES	Consistent (Pure Model State)	Consistent Wave (L+T+C)	Consistent Other Models	N
%	15.02	7.51	7.51	1132
Factor with respect to random	13.04	293.31	6.67	

Results related to high school and middle school students as well as results related to wall context are in some aspects less dramatic than these but in all of the cases the same conclusions apply. Results from these other groups as well as from the other context can be found in the appendix M-1.

5.3 Suggestions for instruction: Addressing alternative models of sound propagation

The assessment results obtained through this test can in principle be combined with any type of instructional approach -- traditional or progressive. A model is a knowledge structure at a higher level than misconception and this is the basis for the claim that addressing a model may at the same time target several misconceptions. For example, a misconceptions related to Entity Model may be that louder sound propagates faster, that sound propagates slower in solids than in gasses, that speed of the sound depends on the movement of the sound source and so on. By addressing the Entity Model these misconceptions are in principle automatically challenged. Of 6 questions in this test, question 1 and 6 can be used also on their own, in a peer instruction mode, but on their own they don't fully elicit students' models.

Although we did not investigate what instructional strategies work best in addressing students models of sound propagation, we will make several suggestions related to possibly optimal instructional approach. We begin by analyzing results obtained from the samples when the test was administered before and after the instruction. In addition to samples shown in the Section 4.5.7.4, Table 5.5. also includes a sample tested in Spring 2004 with test version 9.0 (all others were tested with version 8.9).

Table 5.5

Results of pre- and post-instruction testing and employed instructional methods

Institution	Course Math Level	Context	N Pre	N Post	Method	Pre Test Result (%)	Post Test Result (%)	Gain (%)	Normalized gain
BOTH CONTEXST (WHOLE CLASS)									
University, NY	Calc.	Air	100	95	Research based	9.50	29.21	19.71	0.2178
Middle S., HR	Algb.	Air	75	99	Research based	0.00	19.19	19.19	0.1919
University, PA	Algb.	Both	12	10	Lec./Demo/Lab	0.00	12.50	12.50	0.1250
High S. (1), HR	Algb.	Both	49	51	Lecture / Demo	0.51	11.76	11.25	0.1131
University, NC	Calc.	Air	57	19	Research based	0.44	9.21	8.77	0.0881
University, KS	Algb.	Both	175	177	Lec./Demo/Lab	2.57	5.79	3.22	0.0330

Gain (G) = post-test - pre-test. Normalised gain (g) = gain / (maximum possible gain)

Gains are shown in terms of percentage and normalized gains and samples are ordered in a way that those with higher gains are higher in the table. The order is the same according to both indicators. In two of these 6 cases the instruction strategy different from the regular one was utilized as a consequence of the instructor's familiarity with the test and with students' mental models of sound propagation. This happened in the case of the middle school (ranked second) and in the case of university that was ranked 6th. The middle school teacher adopted learning cycle approach to address students' models and in the case of the university sample, the instructor tried out the strategy of emphasizing the correct model while making sure not to mention the incorrect ones.

Table 5.5 indicates that research-based methods "do better" than traditional ones. A research-based instructional method that was utilized in two instances was more successful in the case when students "quality" was better at the beginning. Traditional or mostly traditional methods worked better when the class was smaller. Impressive gain (when compared to others in the table) was achieved by the instructional approach called Modeling Instruction (Hestenes, 1996; Physics Education Group at Arizona State University, 2000) which in an adaptation of the learning cycle (Abraham, 1998). This method was employed in the case of the middle school sample. The approach was suggested by researchers to this particular teacher and at this point we will present it as our suggestion for instruction (see appendix V). This approach (to teaching of sound as presented in appendix V) however needs further research in more controlled conditions. The basis for our optimism related to this suggestion for instruction is that the achieved gain in this case was close to the highest one and the highest one was achieved by students that were far better in the beginning. Also here we compare middle school students with university students. In carrying out the lecture, the middle school teacher did not have at the time all the equipment necessary for the desired demonstrations and it is reasonable to assume that otherwise, results would have been even better.

The instructional approach that we suggest is based on combination of several of the research-based instructional methods all of which to a good extent fit the description "Guided discovery". These are: The Learning Cycle (Abraham, 1998), Modeling Instruction Program (Hestenes, 1996; Physics Education Group at Arizona State

University, 2000), Socratic Dialog-Inducing (SDI) Labs (Hake) (Hake, 1992) and several others.

5.4 Suggestions for further studies

Research studies typically open more questions that they solve and this one was not unique in that respect. Some of the questions that the study opened are big themes while others are specific issues. Below are some of the broad questions that this study opened:

- Can the testing approach (Linked Item Model Analysis or LIMA) that was developed in this study for eliciting of students mental models of sound propagation be applied to other concepts in physics? Is the usability of this approach limited to areas where hybrid models play an important role or the approach is "hybrid model-independent"? How about applicability of the approach to assessment of student's knowledge in domains of other natural sciences?

- What is the instructional utility of this type of testing? Is this approach to addressing of the underlying models in real time likely to help students learn and move toward desired conceptual change?

- What is instructional value of the test's real-time aspect? How effectively teachers can implement the real-time aspects of this testing approach and real-time aspects of the associated instructional strategies?

- Is it possible to create broad teaching strategies or simple instructional techniques that individually address student's mental models in real time?

- Is this testing approach applicable in psychological personality tests? Would it provide information that current tests in that field do not?

Some more specific questions also well worth investigation are:

- Is it useful to administer this test as an online assignment before the instruction on sound so the activity is not needed during the class time? Are there, one the other side, any advantages of administering the test in the classroom which are more important than time saved through online assessment?

- How applicable is this test at the middle school level?

➢ How would a branched version of the test look and would it have any advantages with respect to this one? It is possible that branched version (in which the next question would be based upon the student's answer on the previous one) would significantly simplify some of the questions and answer choices of this test (especially in questions Q4 and Q5). This might shorten the time of the test taking and additional simplicity might add to the test validity.

Hybrid mental models have been described as transitional cognitive elements in conceptual change related to several important physics domains like electrostatics (Otero, 2003), waves (Hrepic, 2002; Hrepic et al., 2002), Newtonian mechanics (Itza-Ortiz, Rebello, & Zollman, 2004) as well as in other domains of physical sciences (Vosniadou, 1994). For this reason it is likely that even if it turns out that LIMA approach is useful only in domains where hybrid models play a significant role, due to the number and importance of these domains, LIMA might play an important role in eliciting students' mental models and their state of understanding in future.

Uniqueness of this approach to testing and analysis of the test items opens a variety of questions and possibilities related to possible usages in eliciting of psychological constructs for diagnostic purposes not only in education but also in psychology. Mental model (state) as described in this dissertation is a psychological construct. So it is possible that this testing approach (which proved useful for eliciting mental model states associated with sound propagation) could be applicable also to other psychological constructs, which are not necessarily cognitive. The proposed tests would elicit the specific psychological state "utilized" by a subject (which could be pure or mixed state, which is what our test does too). It is possible that personality tests in which answers on different questions are combined into full sentences might provide insights into examinee's psychological states that are missed in inventories with self-standing questions. With respect to Likert scale tests, the test type that we utilized might enable significant reduction in number of questions needed to determine the desired construct.

5.5 Conclusion

Earlier studies (Hrepic, 2002; Hrepic et al., 2002) investigated students' mental models of sound propagation and found patterns in their structure and dynamics. In this study we developed a model inventory for sound propagation that elicits these models in real time i.e. the inventory can be utilized during the instruction as a formative assessment tool. The full name of the test is "Formative Assessment of Mental Models of Sound Propagation" or for short "FAMM-Sound". Results of the test are analyzed so that analysis program first determines if the student is self-consistent i.e. if he or she uses a single model throughout the test. If student is consistent, s/he is in a pure model state and otherwise in a mixed model state. For students in mixed model state, the analysis program identifies different models that each student uses in a way that it combines answers from different test questions. In the air context, a minimum of three questions are needed to determine the model once. In the more complex wall context, because of the larger number of the items involved in sound propagation, four and sometimes all 6 test questions are "consulted" and compared in order to determine a model associated with a single answer choice.

To the best of the author's knowledge, this approach to eliciting students' knowledge (in the form of mental models) has not been utilized earlier. And the search after similar approaches throughout different fields was not successful either.

The difference between the analytical method of analysis of students model states developed in this study and that suggested earlier (Bao & Redish, 2001), is that in this approach there is no one-on-one match between answer options and mental models.

Another major difference is that our approach does not treat students' model states probabilistically. Based on results obtained through interviews with students in this study we believe that mental models that students use are deterministically defined. A big number of various factors affect students' usage of models. These factors are previous knowledge, contextual clues (i.e. questions themselves), question sequence, student's epistemological state, answer choices (when a multiple-choice test is given) and so on. The multitude of these factors, their dynamic inter-relatedness and our limited understanding of each of them makes the students' model not predictable. However we believe that the model which a student utilizes is deterministically rather than

probabilistically determined. We are optimistic with respect to our ability to develop effective instructional approaches that address students' alternative models and we base this optimism on these deterministic aspects of the paradigm of conceptual change.

Literature:

Abimbola, I. O., & Baba, S. (1996). Misconceptions and alternative conceptions in science textbooks: The role of teachers as filters. *American Biology Teacher, 58*(1), 14-19.

Abraham, M. R. (1998). The learning cycle approach as a strategy for instruction in science. In B. J. Fraser, Tobin, K. G. (Ed.), *International handbook of science education, Part 1* (Vol. 1, pp. 349-362). Dordrecht, Netherlands: Kluwer Academic Press.

APA Publication Manual (Ed.). (2001). *Publication Manual of the American Psychological Association*. Washington, DC: American Psychological Association.

Ashworth, P., & Lucas, U. (2000). Achieving empathy and engagement: A practical approach to the design, conduct and reporting of phenomenographic research. *Studies in Higher Education, 25*(3), 295-308.

Asynchronous Learning Networks. (1998). *Constructivism* [www]. ALN Magazine 1(1). Retrieved February 23, 2002, from the World Wide Web: http://www.aln.org/alnweb/magazine/issue1/sener/constrct.htm

Bao, L. (1999). *Dynamics of student modeling: A theory, algorithms, and application to quantum mechanics*. Unpublished Ph.D. Disertation, University of Maryland, College Park, MD.

Bao, L., & Redish, E. F. (2000). Model Analysis: Modeling and Assessing the Dynamics of Student Learning". *Journal of Physics Education Research, submitted*.

Bao, L., & Redish, E. F. (2001). Model analysis: Assessing the dynamics of student learning. *Submitted to Cognition & Instruction*.

Bao, L., Zollman, D., Hogg, K., & Redish, E. F. (2000). Model analysis of fine structures of student models: An example with Newton's third law. *Journal of Physics Education Research, submitted*.

Barman, C. R., Barman, N. S., & Miller, J. A. (1996). Two teaching methods and students' understanding of sound. *School Science and Mathematics, 2*, 63-67.

Beaty, W. J. (2000). *Children's Misconcepts about Science - A list compiled by the AIP Operation Physics Project* [www]. William J. Beaty. Retrieved March, 14, 2002, from the World Wide Web: http://www.amasci.com/miscon/opphys.html

Bodner, G., Klobuchar, M., & Geelan, D. (2001). The many forms of constructivism. *Journal of Chemical Education, 78*(8), 1107.

Boudourides, M. A. (1998). *Constructivism and education: A shopper's guide.* Paper presented at the International Conference on the Teaching of Mathematics, July 3-6, Samos, Greece.

Brandt, D. S. (2002, February 25). *Constructivist approaches to teaching internet-related topics*. Purdue University, Purdue University Libraries. Retrieved March 02, 2002, from the World Wide Web: http://thorplus.lib.purdue.edu/~techman/constr.html

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

Cañas, J. J., & Antolí, A. (1998). The role of working memory in measuring mental models. In T. R. G. Green & L. Bannon & C. P. Warren & J. Buckley (Eds.), *Proceedings of the Ninth European Conference on Cognitive Ergonomics - Cognition and Cooperation*. Rocquencourt: EACE. INRIA.

Chen, I. (2001). *Constructivism* [www]. University of Huston, College of Education. Retrieved March 1, 2002, from the World Wide Web: http://pdts.uh.edu/~ichen/ebook/ET-IT/constr.htm

CIREM. (2002). *List of definitions*. CIREM. Retrieved 10. Feb. 2004, 2004, from the World Wide Web: www.cirem.org.uk/definitions.html

Clerk, D., & Rutherford, M. (2000). Language as a confounding variable in the diagnosis of misconceptions. *International Journal of Science Education, 22*(7), 703-717.

Clough, E. E., Driver, R. (1986). A study of consistency in the use of students' conceptual frameworks across different task contexts. *Science Education, 70*(4), 473-496.

Craik, K. (1943). *The Nature of Explanation*. Cambridge, UK: Cambridge University Press.

Creswell, J. W. (1994). *Research design: Qualitative and quantitative approaches*. Thousand Oaks, CA: SAGE Publications, Inc.

diSessa, A. A. (1993). Towards an epistemology of physics. *Cognition and Instruction, 10*(2-3), 105-225.

diSessa, A. A. (1996). What do "just plain folk" know about physics? In D. R. Olson & N. Torrance (Eds.), *Handbook of Education and Human Development: New Models of Learning, Teaching, and Schooling*. Oxford: Blackwell Publishers, Ltd.

diSessa, A. A. (2002). Personal communication.

diSessa, A. A. (2002). Why "conceptual ecology" is a good idea. In M. Limon & L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice* (pp. 29-60). Dordrecht, Netherlands: Kluwer Academic Publishers.

diSessa, A. A., & Sherin, B. L. (1998). What changes in conceptual change? *International Journal of Science Education, 20*(10), 1155-1192.

Donovan, S. M., Bransford, J. D., & Pellegrino, J. W. (Eds.). (1999). *How people learn: Bridging research and practice*. Washington, DC: National Academy Press.

Erickson, F. (1998). Qualitative research methods for science education. In B. J. Fraser, Tobin, K. G. (Ed.), *International handbook of science education, Part 1* (Vol. 2, pp. 349-362). Dordrecht, Netherlands: Kluwer Academic Press.

Glasersfeld, E. v. (1990). Environment and education. In L. P. Steffe & T. Wood (Eds.), *Transforming Children's Mathematics Education: International Perspectives* (pp. 200-215). Hillsdale, NJ: Lawrence Erlbaum.

Greca, I. M., & Moreira, M. A. (2002). Mental, physical, and mathematical models in the teaching and learning of physics. *Science Education, 86*(1), 106-121.

Hake, R. R. (1992). Socratic Pedagogy in the Introductory Physics Lab. *The Physics Teacher, 30*, 546 - 552.

Hake, R. R. (1997). Interactive-engagement vs traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *Am. J. Phys, 66*, 64-74.

Haladyna, T. M. (1999). *Developing and validating multiple-choice test items*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Hammer, D. (1996). More than misconceptions: Multiple perspectives on student knowledge and reasoning, and an appropriate role for education research. *American Journal of Physics, 64*(10), 1316-1325.

Hammer, D. (2000). Student resources for learning introductory physics. *American Journal of Physics, Physics Education Research Supplement,, 68*(S1), S52-S59.

Hammer, D. (2002). Personal Communication.

Hanna, G. S. (1993). *Better teaching through better measurement*. Orlando, Florida: Harcourt Brace Jovanovic, Inc.

Hestenes, D. (1996). *Modeling Methodology for Physics Teachers.* Paper presented at the 1996 International Conference on Undergraduate Physics Education, College Park, MD.

Hewitt, P. G. (2002). *Conceptual Physics* (9th ed.): Addison Wesley.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. R. (1989). *Induction: processes of inference, learning, and discovery*. Cambridge, Mass.: The MIT Press.

Hrepic, Z. (1998). *Ucenicke koncepcije u razumijevanju zvuka (Students' conceptions in understanding of sound).* Unpublished Bachelor's thesis, University of Split, Split, Croatia.

Hrepic, Z. (1999). *Ucenicke koncepcije u razumijevanju zvuka (Students' conceptions in understanding of sound).* Paper presented at the Fourth Croatian Symposium on Physics Education: Role of Experiment in Physics Education, Novi Vinodolski, Croatia.

Hrepic, Z. (2002). *Identifying students' mental models of sound propagation.* Unpublished Master's thesis, Kansas State University, Manhattan.

Hrepic, Z., Zollman, D., & Rebello, S. (2002). *Identifying students' models of sound propagation.* Paper presented at the 2002 Physics Education Research Conference, Boise ID.

Hrepic, Z., Zollman, D., & Rebello, S. (2003). *Students' mental models of sound propagation: Implications for a theory of conceptual change.* Manuscript submitted for publication.

Itza-Ortiz, S. F., Rebello, S., & Zollman, D. A. (2004). Students' models of Newton's second law in mechanics and electromagnetism. *European Journal of Physics, 25*, 81–89.

Johnson-Laird, P., & Byrne, R. (2002, Dec 18, 2000). *Mental Models Website: A gentle introduction*. Department of Psychology, University of Dublin, Trinity College, Dublin, Ireland. Retrieved Jun, 10. 2002, http://www.tcd.ie/Psychology/Ruth_Byrne/mental_models/theory.html, from the World Wide Web: http://www.tcd.ie/Psychology/Ruth_Byrne/mental_models/theory.html

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, UK: Cambridge University Press.

Kelly, G. A. (1955). *The psychology of personal constructs: A theory of personality, .* New York: W. Norton & Company, Inc.

Krathwohl, D. R. (1998). *Methods of educational and social science research: an integrated approach*. New York: Longman.

Linder, C. J. (1987). Tertiary physics: a case study in students' conceptions of sound. In J. Novak (Ed.), *Proceedings of the Second International Seminar: "Misconceptions and Educational Strategies in Science and Mathematics"* (Vol. 3, pp. 322-334). Cornell University, Ithaca, NY, USA.

Linder, C. J. (1992). Understanding sound: so what is the problem? *Physics Education, 27*(5), 258-264.

Linder, C. J. (1993). University physics students' conceptualizations of factors affecting the speed of sound propagation. *International Journal of Science Education, 15*(6), 655-662.

Linder, C. J., & Erickson, G. L. (1989). A study of tertiary physics students' conceptualizations of sound. *International Journal of Science Education, 11*(special issue), 491-501.

Maloney, D., & Siegler, R. S. (1993). Conceptual competition in physics learning. *International Journal of Science Education., 15*, 283-295.

Marton, F. (1978). *Describing conceptions of the world around us* (66). Molndal, Sweden (ERIC Document Reproduction Service No. ED 169 074): University of Goteborg, Institute of Education.

Marton, F. (1986). Phenomenography: A research approach to investigating different understandings of reality. *Journal of Thought, 21*(3), 28-49.

Marton, F., & Saljo, R. (1984). Approaches to learning. In F. Marton & D. Hounsell & N. Entwistle (Eds.), *The experience of learning* (pp. 36-55). Edinburgh: Scottish Academic Press.

Matthews, M. R. (1993). Constructivism and science education: Some epistemological problems. *Journal of Science Education and Technology, 2*(1), 359-370.

Maurines, L. (1992). Spontaneous reasoning on the propagation of visible mechanical signals. *International Journal of Science Education, 14*(3), 279-293.

Maurines, L. (1993). Spontaneous reasoning on the propagation of sound. In J. Novak (Ed.), *Proceedings of the Third International Seminar on Misconceptions and Educational Strategies in Science and Mathematics*. Ithaca, New York: Cornell University (distributed electronically).

McDermott, L. (1998). Students' conceptions and problem solving in mechanics. In A. Tiberghien, Jossem, E. , Barojas, J. (Ed.), *Connecting research in physics education* (pp. 1-11). Ohio: ICPE Books.

Merino, M. J. (1998a). Complexity of pitch and timbre concepts. *Physics Education, 33*(2), 105-109.

Merino, M. J. (1998b). Some difficulties in teaching the properties of sounds. *Physics Education, 33*(2), 101-104.

Millar, R. (1989). Constructive criticisms. *International Journal of Science Education, 11*, 587-596.

Minsky, M. L. (1986). *The Society of Mind*. New York: Simon and Schuster Inc.

Minstrell, J. (1992). Facets of Students Knowledge and Relevant Instruction. In R. Duit & F. Goldberg & H. Niedderer (Eds.), *Research in physics learning: Theoretical issues and empirical studies, Proceedings of an International Workshop held at the University of Bremen, March 4-8, 1991* (pp. 110-128). Kiel: Institut für die Pädagogik der Naturwissenschaften an der Universität Kiel.

Norman, D. A. (1983). Some observations on mental models. In D. A. Gentner & A. L. Stevens (Eds.), *Mental models*. Hillsdale, NJ: Lawrence Erlbaum.

Oosterhof, A. (2001). *Classroom applications of educational measurement*. Upper Saddle River, New Jersey: Prentice Hall, Inc.

Osborne, J. (1996). Beyond constructivism. *Science Education, 80*(1), 53-82.

Otero, V. K. (2003). *A cognitive processes and the learning of physics part I: The evolution of knowledge from a Vygotskian perspective*, To appear in: Proceedings of the International School of Physics "Enrico Fermi", edited by M. Vicentini and E.F. Redish (IOS Press, Amsterdam), July, 2003, Varenna, Italy.

Phillips, D. C. (1995). The good, the bad and the ugly: The many faces of constructivism. *Educational Researcher, 24*(7), 5-12.

Physics Education Group at Arizona State University. (2000). *Modeling Instruction Program* [www]. Arizona State University. Retrieved 24. Aug. 2003, 2003, from the World Wide Web: http://modeling.la.asu.edu/

Piaget, J. (1972). *Psychology and epistemology: Towards a theory of knowledge*. Harmondsworth: Penguin.

Polkinghorne, D. E. (Ed.). (1992). *Psychology and postmodernism*. Thousand Oaks, CA: Sage Publications.

Redish, E. F. (1994). The implications of cognitive studies for teaching physics. *American Journal of Physics, 62*(6), 796-803.

Runes, D. D. (1962). *Dictionary of philosophy* (15th ed.). Paterson, NJ: Littlefield, Adams & Co.

Schecker, H., & Gerdes, J. (1999). Messung von Konzeptualisierungsfähigkeit in der Mechanik. Zur Aussagekraft des Force Concept Inventory. *Zeitschrift für Didaktik der Naturwissenschaften, 5*(1), 75-89.

Solomon, J. (1994). The rise and fall of constructivism. *Studies in Science Education, 23*, 1-19.

Taber, K. S. (2000). Multiple frameworks?: Evidence of manifold conceptions in individual cognitive structure. *International Journal of Science Education, 22*(4), 399-417.

Tam, M. (2000). Constructivism, Instructional Design, and Technology: Implications for Transforming Distance Learning. *Educational Technology and Society, 3*(2).

Tobin, K., & Tippins, D. (1993). Constructivism as a referent for teaching and learning. In K. Tobin (Ed.), *The Practice of Constructivism in Science Education* (pp. 3-21). Hillsdale, NJ: Lawrence-Erlbaum.

University of Maryland. (1999). *Wave Diagnostic Test*. University of Maryland. Retrieved 1. Feb. 1004, 2004, from the World Wide Web: http://www2.physics.umd.edu/~wittmann/research/umd_wavetests.pdf

Van der Veer, G. (2000, October 15, 2001). *Mental models of incidental human-machine interaction* [www]. Faculty of Sciences, Vrije Universiteit, Amsterdam, The Netherlands. Retrieved Jun, 10, 2002, from the World Wide Web: http://www.cs.vu.nl/~gerrit/mmi9910-report1.doc

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning & Instruction, 4*, 45-69.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harward University Press.

Wandersee, J. H., Mintzes, J. J., & Novak, J. D. (1994). Research on alternative concepts in science. In D. L. Gabel (Ed.), *Handbook of Research on Science Teaching and Learning* (pp. 177-210): Macmillan Publishing Company, NY.

Wittmann, M. C. (1998). *Making sense of how students come to an understanding of physics: An example from mechanical waves.* Unpublished Ph.D. Dissertation, University of Maryland, College Park, MD.

Wittmann, M. C. (2001). The object coordination class applied to wavepulses: Analysing student reasoning in wave physcs. *International Journal of Science Education, 24*(1), 97-118.

Wittmann, M. C., Steinberg, R. N., & Redish, E. F. (1999). Making sense of how students make sense of mechanical waves. *The Physics Teacher, 37*(1), 15-21.

Wittmann, M. C., Steinberg, R. N., & Redish, E. F. (2002). Understanding and Addressing Student Reasoning about Sound. Manuscript submitted for publication.

Wolcott, H. F. (1994). *Transforming qualitative data: Description, analysis and interpretation*. Thousand Oaks, CA: SAGE Publications, Inc.

Young, R. M. (1983). Surrogates and mappings: Two kinds of conceptual models for interactive devices. In D. A. Gentner & A. L. Stevens (Eds.), *Mental models*. Hillsdale, NJ: Lawrence Erlbaum.

Zollman, D. (1999). NSF grant proposal: *"Technology & Model-Based Conceptual Assessment: Research in Students' Applications of Models in Physics & Mathematics"*.