Journal of Science Education and Technology, Vol. 14, No. 1, March 2005 (© 2005)
DOI: 10.1007/s10956-005-2738-x

An Investigation of Factors Affecting the Degree of Naïve Impetus Theory Application

Xiufeng Liu[1,3] and Dan MacIsaac[2]

This study investigates factors affecting the degree of novice physics students' application of the naïve impetus theory. Six hundred and fourteen first-year university engineering physics students answered the Force Concept Inventory as a pre-test for their calculus-based course. We examined the degree to which students consistently applied the naïve impetus theory across different items. We used a 2-way repeated measures ANOVA and linear regression to analyze data coded from *incorrect* student responses. It was found that there were statistically significant main effects for item familiarity and item requirement for explanation vs. prediction on the measured degree of impetus theory application. Student course grades had no significant effect on impetus theory application. When faced with items that were unfamiliar and predictive, students appeared to rely on non-theoretical, knowledge-in-pieces reasoning. Reasoning characteristic of naïve theories was more frequently applied when students were completing familiar problem tasks that required explanation. When considering all the above factors simultaneously, we found that the degree of naïve impetus theory application by students is attributable to variables in the following order: familiarity, prediction, and explanation.

KEY WORDS: mental models; physics; student conceptions.

INTRODUCTION

Promoting student conceptual understanding in science has been a focus in science education reform over the past two decades. The theoretical foundation for such a reform is mainly related to students' alternative conceptions and conceptual change (see comprehensive reviews by Duit and Treagust, 2003; Keil, 1998; Tyson *et al.*, 1997; Wandersee *et al.*, 1994). Conceptual change theories claim that students have various alternative or naïve conceptions about a science concept they are going to learn, and the key to successful science teaching is to identify students' preconceptions and design appropriate learning sequences for students to critically evaluate their preconceptions and construct new meanings compatible with current scientific theories.

Although conceptual change science teaching is a pedagogical issue, effective conceptual change science teaching requires appropriate understanding about the nature and development of students' alternative conceptions, which is a cognitive issue. It is commonly accepted in science education that student cognition is domain specific (Driver, 1983; Driver and Easley, 1978; Erickson, 1994). Children are universal novices; one key difference between novices and experts is the amount of knowledge they have in a specific domain and how such knowledge is organized (Bedard and Chi, 1992; Chi and Koesk, 1983; Chi, 1978; Schneider *et al.*, 1993). Thus, conceptual change science teaching is to facilitate children's knowledge reconstruction through building more coherent knowledge networks from less coherent knowledge networks instead of replacing old knowledge systems

[1]Department of Learning and Instruction, Graduate School of Education, SUNY, University at Buffalo, Buffalo, New York.
[2]Department of Physics, Buffalo State College, SUNY, New York.
[3]To whom correspondence should be addressed Department of Learning and Instruction, Graduate School of Education, SUNY, University at Buffalo, Buffalo, New York 14260-1000; e-mail: xliu5@buffalo.edu

by new ones. However, student cognition is also context specific (Lave, 1978; Lave and Wenger, 1991). Thus, conceptual change science teaching needs also to help students reason and apply appropriate knowledge in various contexts.

The above two views on cognition, i.e. the domain specific or structural and the context dependent, have resulted in two different views about the characteristics of students' alternative conceptions. In general, those different views are of two types – those emphasizing the inconsistency among various alternative conceptions within and among students and those emphasizing the consistency among various alternative conceptions within and among students. Building on previous research to resolve the difference between the above two views, this study is another effort to clarify this cognitive dispute. The pedagogical importance of resolving this cognitive dispute is to inform how conceptual change science teaching should be designed and implemented. That is, advocates believing in the consistency of students' alternative conceptions may design and implement conceptual change science teaching around developing and modifying students' cognitive structures, while those believing in the inconsistency of students' alternative conceptions may design and implement conceptual change science teaching around developing appropriate learning contexts conducive to developing scientific understanding.

Different Views on the Consistency of Student Alternative Conceptions

Views emphasizing the inconsistency of student conceptions are represented by diSessa (1988, 1993) and diSessa and Sherin (1998). diSessa and Sherin (1998) claimed that there are two components in a student's concepts associated with structure and performance. The structural component includes a set of readout strategies for getting information and a causal net of general beliefs. The readout strategies are specific to problem tasks, while the causal net is the general class of knowledge and reasoning strategies that determine when and how some observations are related. diSessa and Sherin further claim that students' causal nets in physics are usually an uncoordinated and naive "senses of mechanism" called "p-prims" (phenomenological primitives). P-prims are tiny ideas associated with perceived features of particular problem situations. This view of the uncoordinated nature of students' alternative conceptions, particularly as applied to introductory physics, appeared in earlier publications by diSessa (1988, 1993).

Specifically, diSessa argued that novice learners' knowledge of physical phenomena is not a logically organized structure, but more like knowledge in pieces. Students' naïve conceptions are phenomenological because they are responses to experienced and observed phenomena and are specifically linked to these phenomena rather than being general or abstract. Students' naïve conceptions are primitive because they are self-evident to themselves and therefore require no further explanation. Because p-prims are fragmented, uncoordinated, and context dependent, novices' conceptions have no commitment or systematicity. Other researchers also subscribe to the notion of knowledge-in-pieces. For example, Minstrell (2001) analyzed introductory physics learners reasoning at length. He identified and catalogued a large collection of various small units of knowledge he termed "facets" and related them to a particular learning outcome to form a hierarchical facet cluster. Unlike diSessa's p-prims, Minstrell's facet clusters bridge to scientific conceptions or learning standards.

Views emphasizing the consistency of students' alternative conceptions can be summarized into three types: theory theory, mental model theory, and ontological attribution theory. Theory theory suggests that students' alternative conceptions are organized into coherent, causal-explanatory systems within specific domains (Brewer and Samarapungavan, 1991; Carey, 1985; Gopnik and Wellman, 1994; Gopnik and Meltzoff, 1997; Schwitzgebel 1999; Wellman, 1990). "The hypothesis of theory theory is that there are deep similarities between the underlying cognitive mechanisms involved in the epistemological endeavors of childhood and of science" (Gopnik and Wellman, p. 259). Of course, students' theories may not be scientific; they are often everyday folk theories–coherent systems that organize and structure everyday thinking. Gopnik and Meltzoff described three classes of features characteristic of theories; they are structural features (i.e. abstractness, coherence, causality, and counterfactuals), functional features (i.e. prediction, interpretation, and explanation), and dynamic features (i.e. denial, *ad hoc* auxiliary hypothesis, alternative models, and intense experimentation and observation). Schwitzgebel proposed that: (a) a theory is a set of propositions; (b) any set of propositions can potentially be regarded as a theory, and to regard a set of propositions in this way is to be committed to evaluating that set of propositions in terms of its capacity to generate good explanations in a domain; and (c) to subscribe to a theory is to accept the propositions composing it and to employ them, or be disposed to employ them, in explaining phenomena within the theory's domain.

The mental model theory has been around for quite some time (Gentner and Stevens, 1983; Johnson-Laird, 1983), but it has only recently gained prominence in science education for explaining student alternative conceptions (Chiu *et al.*, 2002; Coll and Treagust, 2003a,b; Gilbert and Boulter, 2000; Taber, 2003; Taylor *et al.*, 2003; Vosniadou and Brewer, 1992). Mental models are mental representations for interpreting experiences and making sense of the physical world (Coll and Treagust). Mental models are functional evolving systems; they are incomplete and may not be scientific. Johnson-Laird asserts that the essence of mental models is to build a working model of the phenomenon in the mind in order to make predictions. Importantly, mental models are applied consistently across different contexts (Gentner and Stevens, 1983; McCloskey, 1983; Rogers *et al.*, 1992; Vosniadou, 1992, 1994; Vosniadou and Brewer, 1992).

The third view emphasizing the consistency of students' alternative conceptions is the ontological category theory (Chi, 1992; Chi and Roscoe, 2002; Chi *et al.*, 1994; Keil, 1989). For example, Chi and Roscoe claim that (a) concepts are loosely structured in something like a hierarchy tree according to different ontologies, such as material substance (e.g. natural kind and artifact), processes (e.g. procedure, event, and constraint-based interaction), and mental states (e.g. emotional and intentional); (b) trees are fundamentally distinct from one another in ontology; (c) people's conceptual structures correspond to different trees and branches. Based on the above assumptions, they argued that misconceptions are mis-categorizations of concepts across ontological categories. When examining the variety of student alternative conceptions from the above ontological framework, they found much consistency in student alternative conceptions.

The above three views on the consistency of student alternative conceptions are distinct, but also related. Vosniadou (1992, 1994) claimed that specific theories are built from everyday experiences or instruction to explain a limited range of phenomena. She considers that a theory is based on a few abstract and stable core presuppositions or beliefs, but that a mental model is a transient and analogical construction, elaborated on the spot for the purpose of solving a given problem. Giere (1988) claims that scientific theories are families of models. Similarly, Brewer (1999) claims that scientific models are theories, and it is the underlying model that gives scientific theories power to explain phenomena. In fact, proponents of mental models consider scientific theories as consensus models (Gilbert and Boulter, 2000). A relationship between the ontological attribution theory and mental model/theory theory also exists. For example, in a recently study, misattribution of ontology to a concept by a novice was considered to be a naïve mental model (Mazens and Lautrey, 2003).

Resolving the Difference

Attempts have been made to resolve the above differences. For example, Anderson *et al.* (1992) conducted an experimental study investigating the consistency of novices' conceptions of motion. They used 16 problems related to independent linear, dependent linear, circular, and pendular motion. They found that the subjects were likely to give a more accurate response if mass/velocity variables were either both small in value or both large in value. On the other hand, if one variable was high in value and simultaneously the other was low, the subjects were likely to give a less accurate response. In particular, the naïve impetus theory was more commonly applied when objects had high velocity/low mass than when objects had low velocity/high mass. The above findings suggest that mental models existed in novices only to a limited degree, not universally as implied by the mental model theory. However, when considering item requirement for prediction or explanation, they found that students' prediction of paths and directions was inconsistent (as indicated by low mean inter-item correlation, 0.12), but that students' explanation of motion was highly self-consistent (as indicated by high mean inter-item correlation, 0.73). They concluded that both mental models and knowledge-in-pieces played a role in novices' conceptions of motion, and which is predominant depends on item contextual characteristics (such as velocity and mass) and item requirement for prediction or explanation. It is not known whether there are other contextual factors correlated with the degree of student application of a naïve mental model.

In a more recent study, Mazens and Lautrey (2003) studied 6–10 year old children's conceptions of sound. They identified four mental models in students' naive conceptions from associating sound with material objects (the naïve idea) to considering sound as vibration (the scientific). Models identified by Mazens and Lautrey included: (a) sound cannot pass through other objects unless there are holes to pass through, (b) sound can pass through solids if sound is harder than the objects, (c) sound is immaterial, and (d) sound is a vibratory process. Specifically in the case of the mental model attributing all properties of matter to sound, Mazens and Lautrey did not find two distinct groups of children– those children who attributed all properties of matter to sound and those children who did not attribute all properties of matter to sound. Instead, different properties of matter seemed to be attributed and abandoned in a hierarchical rather than synchronic fashion from 6 to 10 years old. Although the various forms of inconsistency of student responses seemed to support the knowledge-in-pieces theory where different p-prims were activated according to the surface cues of the situation, examination of the arguments given by children also suggested relatively stable beliefs and presuppositions underlying the apparent inconsistency. Thus, Mazens and Lautrey concluded that a lack of consistency among the properties in children's responses could not be totally interpreted as corresponding to an absence of structure in naïve knowledge. They suggested that a continuum between knowledge-in-pieces and mental models might exist in students' naïve conceptions, and that students could experience a transition from naïve conceptions to more scientific conceptions.

The difference between student use of mental models and knowledge-in-pieces may also be related to students' academic background. For example, the Force Concept Inventory (Hestenes *et al.*, 1992/1995) was designed for assessing students' Newtonian and non-Newtonian conceptions of force through six closely related categories of questions. Through factor analysis of correct and incorrect responses, eliminating specific distracter data, based on both high school and university samples, Huffman and Heller (1995) found that only fewer than 6 FCI items converged on any single factor, and that the two samples suggested different factors. Huffman and Heller interpreted their results to suggest that students' reasoning of the force concept was uncoordinated and context dependent, supporting the knowledge-in-pieces theory. However, Hestenes and Halloun (1995) counter-argued that a single factor solution of factor analysis consistent with the Newtonian force concept could be expected if a non-novices (i.e. those who scored 60% – 80% on the inventory) or a physicist data sample were used. Hestenes and Halloun suggested that for novices, factor solutions from factor analysis would be more in agreement with the clusters of naïve conceptions identified in Hestenes *et al.* (1992/1995). In fact, one factor identified by Huffman and Heller contained some of those items identified as indication of the impetus naïve theory by Hestenes *et al.* If Hestenes and Halloun are correct, then the application of mental models may also depend on students' academic levels.

Experts are found to apply mental models more consistently than novices (Chi *et al.*, 1981; Chi *et al.*, 1988; Chi *et al.*, 1982), while novices commonly apply the impetus theory (Hallouin and Hestenes, 1985a,b). Experts are found to (a) possess extensive and highly integrated bodies of domain knowledge; (b) be effective at recognizing the underlying structure of domain problems; (c) select and apply appropriate problem-solving procedures for the problem at hand; and (d) retrieve relevant domain knowledge and strategies with minimal cognitive effort (Alexander, 2003). The above novice-expert distinction may suggest that the application of the impetus theory is a higher level of cognitive performance than knowledge-in-pieces. From a genetic epistemological view, Piaget perceived that the impetus theory was a more advanced theory than the Aristotelian theory of inherent motion (Piaget and Garcia, 1989).

Thus, it appears that there is no dichotomy between mental models and knowledge-in-pieces, because the application of mental models seems to depend on various factors. The literature reviewed above suggests that these factors may include students' academic backgrounds, specific item requirements for explanation or prediction, and students' familiarity with the problem tasks or contexts. Thus, mental models and knowledge-in-pieces differ not in quality but in quantity – degrees of consistency in responses to problems. No research has been reported that explicitly and systematically investigates those factors on the degree of consistency, which is the purpose of the present study. Specifically, we investigate the degree of novices applying the impetus theory and the factors affecting their application of the impetus theory. We test the hypotheses that students' academic achievement, familiarity with item context, and item requirement for prediction or explanation affect novices' application of the impetus theory.

The focal research questions for this study are:

1. Is there a statistically significant difference in percentages of students applying the impetus theory between items that require prediction and items that require explanation?
2. Is there a statistically significant difference in percentages of students applying the impetus theory between items that are familiar to students and items that are unfamiliar to students?
3. Are there statistically significant interaction effects between student academic achievements and the above independent variables (i.e. item familiarity, and item requirement of explanation or prediction)?
4. Which of these factors when simultaneously considered significantly predict the overall percentage of students applying the naïve impetus theory?

This study will further illuminate the nature of student alternative conceptions so that new theories on student conceptions and conceptual change may be developed. It will bring clarity to the ongoing debate on mental models/theory theory and knowledge-in-pieces regarding students' alternative conceptions. The results also have practical significance by shedding new light on the currently promoted model-based science teaching and learning approach (e.g. De Jong *et al.*, 1999; Korfiatis *et al.*, 1999; Linn and Muilenburg, 1996; Monaghan and Clement, 1999; Stewart *et al.*, 1992; Treagust *et al.*, 2002; Windschitl, 2001).

METHOD

Data Source

The data used in this study came from a pre-instructional survey involving 614 university calculus-based physics course registrants. The instrument was the revised version of the Force Concept Inventory (FCI) (Hestenes *et al.*, 1992/1995). The multiple-choice based paper and pencil survey "requires a forced choice between Newtonian concepts and common-sense alternatives" (Hestenes *et al.*, 1992/1995, p. 142). The FCI is a popular instrument in physics education as it has been given to thousands of physics students of various levels of instruction across dozens of institutions (Hake, 1998). The FCI was developed through analysis of interviews of students from 9th grade to university undergraduate and graduate physics (Hallouin and Hestenes, 1985a,b; Hestenes *et al.*, 1992/1995) with distracters derived from student alternative conceptions. The FCI's quite invidious distracters are culled from student reasoning; hence it is common for new students with no previous physics instruction to score much lower on the FCI than random chance alone would predict.

Most items in the 1995 version are exactly the same as in the 1992 version; however, some items were reordered, or had choices re-ordered, a few items were new and some original items were deleted. Because our interest in this study is on the naïve impetus theory, we will only focus on the items and their distracters identified as involving the impetus theory by Hestenes *et al.* (1992/1995, p. 144). According to Hestenes *et al.*, impetus is perceived by novices as an inanimate "motive power" or "intrinsic force" that keeps objects moving, which contradicts Newton's First Law. Evidence that a student believes in some kind of impetus is therefore evidence that the First Law is not understood.

Hestenes *et al.* listed the distracters of the FCI items indicating the application by novices of the impetus theory. There were 13 such items with various distracters (see Appendix). Some FCI distracters are related to an impetus supplied by a "hit," some related to the loss or recovery of an original impetus, others related to a gradual or delayed impetus build-up, and still others related to a circular impetus. Table I columns 1–4, present those distracters in the original 1992 version of FCI and their corresponding item distracters in the (now prevalent) 1995 revised version of the FCI. It can be seen that choices involving the impetus theory were essentially unchanged in the 1995 version. Although students make a selection on a multiple-choice question based on many possible reasons, and applying the impetus naïve theory is only one of them, the choices identified by Hestenes *et al.* indicating the application of the naïve impetus theory were validated through extensive interviews (Hallouin and Hestenes, 1985a,b). Because we consider applying the naïve impetus theory to be a continuous variable, rather than yes or no, we base our analysis on students' responses to a group of items to establish the degree of consistency of student application of the naïve impetus theory.

Definition of Independent and Dependent Variables

According to our focal research questions, the dependent variable is the degree of students applying

Table I. Correspondence Between the 1992 and 1995 FCI Versions on Distracters Indicating the Impetus Theory

1995 Version		1992 Version			
Item	Choices	Item	Choices	Familiarity	Prediction/ Explanation
Q7	A, D	Q4	A, D	Familiar	Prediction
Q13	A, B, C	Q5	A, B, C	Familiar	Explanation
Q8	C, D, E	Q6	C, D, E	Familiar	Prediction
Q10	B, C, D	Q8	B, C, D	Familiar	Prediction
Q11	B, C	Q9	B, C	Familiar	Explanation
Q6	A	Q10	A	Unfamiliar	Prediction
Q12	C, E	Q16	C, D	Unfamiliar	Prediction
Q30	B, D, E	Q22	B, C, E	Familiar	Explanation
Q14	E	Q23	E	Unfamiliar	Prediction
Q21	A, D	Q24	A, D	Unfamiliar	Prediction
Q23	A, D, E	Q26	A, D, E	Unfamiliar	Prediction
Q24	C, E	Q27	C, E	Unfamiliar	Prediction
Q27	B, D, E	Q29	B, D, E	Familiar	Prediction

the impetus theory in answering a group of items. Because each question is scored as applying the impetus theory (coded as 1) or not applying the impetus theory (coded as 0) according to Hestenes *et al.* (1992/1995), the mean for each item over all the students is the percentage of students applying the impetus theory to the item. This percentage represents the consistency *among* students in applying the impetus theory. Similarly, the mean for each student over a group of items is the percentage of items to which a student has applied the naïve impetus theory. This percentage represents the consistency *within* a student in applying the impetus theory. Accordingly, the mean over both students and a group of items represents the degree of students applying the impetus theory in answering a group of items; this mean defines the dependent variable in our study. Different ways of forming a group of items by independent variables (to be described next) result in different dependent variables (e.g. the dependent variable for prediction questions).

The independent variables involved are item context familiarity, item requirement for prediction or explanation, and students' academic achievement. The criterion used to determine whether or not a question is familiar to students or not is whether or not the question involves an everyday context that most students, not necessarily every student, have likely experienced. For example, ice or street hockey should be reasonably physically familiar to many US students, but experiences with rocket movement would quite rare. The criterion used to determine if a question requires prediction or explanation is whether or not a question asks students to predict paths of the movement of an object (e.g. the path of a hockey puck after being hit) or to identify forces involved in a movement (e.g. identifying forces on a golf ball on its way up). The classification of questions by the above two independent variables, familiarity and prediction/explanation, are presented in columns 5 and 6 of Table I, as well as at the end of each item in the Appendix. Finally, the independent variable of student academic achievement was defined by students' final physics course grades (i.e. A, B, C, D, and F).

Data Analysis

Because the central issue of this study is about the consistency of students' naïve conceptions when they cannot answer the questions correctly, we had to exclude those correct responses by coding them as missing. *Henceforth, we only consider those incorrect responses*. To help understand the distribution among students who answered questions correctly, students who answered the questions incorrectly by choosing naïve impetus theory distracters, and students who answered the questions incorrectly by choosing other distracters, Table II presents the frequencies of students in each of the categories arranged by student course grade. From Table II, we see that overall students with better course grades are more likely to answer the questions correctly. However, for those students who did not answer the questions correctly, course grade does not seem to matter in likelihood of selecting the impetus theory choices. For Questions 11, 13, and 30, a large number of students answered them incorrectly by selecting the impetus theory choices, regardless of their course grades.

Table II. Frequency Distribution of Students Who Answered Questions Correctly, Incorrectly with Impetus Choices, and Incorrectly with Other Wrong Choices by Course Grade

Item	Course grade	# Correct (%)	# Incorrect: impetus (%)	# Incorrect: other choices (%)
Q6	A, B	223 (87.8)	23 (9.1)	8 (3.1)
	C	187 (81.7)	32 (13.9)	10 (4.4)
	D, F	66 (79.5)	14 (16.9)	3 (3.6)
Q7	A, B	207 (81.5)	11 (4.3)	36 (14.2)
	C	172 (75.1)	25 (10.9)	32 (14.0)
	D, F	56 (67.5)	10 (12.0)	17 (20.5)
Q8	A, B	199 (78.3)	37 (14.6)	18 (7.1)
	C	150 (65.5)	51 (22.3)	28 (12.2)
	D, F	53 (63.9)	20 (24.1)	10 (12.0)
Q10	A, B	214 (84.3)	30 (11.8)	10 (3.9)
	C	134 (58.5)	73 (31.9)	22 (9.6)
	D, F	59 (71.1)	21 (25.3)	3 (3.6)
Q11	A, B	88 (34.6)	120 (47.3)	46 (18.1)
	C	45 (19.7)	142 (62.0)	42 (18.3)
	D, F	18 (21.7)	50 (60.2)	15 (18.1)
Q12	A, B	216 (85.0)	37 (14.6)	1 (.4)
	C	161 (70.3)	67 (29.3)	1 (.4)
	D, F	63 (75.9)	20 (24.1)	0 (0)
Q13	A, B	120 (47.2)	134 (52.8)	0 (0)
	C	58 (25.3)	171 (74.7)	0 (0)
	D, F	20 (24.1)	63 (75.9)	0 (0)
Q14	A, B	182 (71.7)	0 (0)	72 (28.3)
	C	119 (52.0)	0 (0)	110 (48.0)
	D, F	46 (55.4)	0 (0)	37 (44.6)
Q21	A, B	141 (55.5)	38 (15.0)	75 (29.5)
	C	94 (41.0)	64 (27.9)	71 (31.1)
	D, F	43 (51.8)	15 (18.1)	25 (30.1)
Q23	A, B	145 (57.1)	68 (26.8)	41 (16.1)
	C	98 (43.0)	82 (36.0)	48 (21.0)
	D, F	39 (47.0)	22 (26.5)	22 (26.5)
Q24	A,B	222 (87.4)	26 (10.2)	6 (2.4)
	C	175 (76.4)	43 (18.8)	11 (4.8)
	D, F	66 (79.5)	15 (18.1)	2 (2.4)
Q27	A, B	197 (77.6)	21 (8.2)	36 (14.2)
	C	65 (78.4)	31 (13.6)	44 (19.3)
	D, F	153 (67.1)	9 (10.8)	9 (10.8)
Q30	A, B	95 (37.4)	156 (63.6)	0 (0)
	C	62 (27.6)	161 (71.5)	2 (.9)
	D, F	17 (20.5)	64 (77.1)	2 (2.4)

We conducted two-way mixed designs Analysis of Variance (ANOVA) to test the effects of students' academic achievements (between-subject variable), item characteristics of familiarity, and prediction or explanation nature (within-subject variables), and the interaction between academic achievement and item characteristics. We conducted two separate 2-way ANOVAs, one for each of familiarity and prediction/explanation. Given the sample sizes to be adequate (see descriptive statistics below) and thus the statistical power to be sufficient, we used .05 as the alpha level for making statistical decisions. Because each 2-way ANOVA addressed a different research question, we did not need to adjust the inflation of type I error caused by multiple ANOVAs. The data sample used in this study was part of a larger data set of 1,313 students using both paper-and-pencil and web before and after instruction (Cole and MacIsaac, 2001, MacIsaac, Cole, and Cole, 2002); they can reasonably be considered to be representative of all large lecture calculus-based introductory physics course registrants at most large research universities. Table III presents the frequency statistics regarding the students' final course grades. It can be seen that although the sample is slightly skewed toward the more able students, overall the departure from normality is acceptable. Thus, we considered that the assumptions of representative-ness, normality, and homogeneity of variance were met. Further, because students completed the FCI paper and pencil survey independently, the assumption of observation independence for 2-way ANOVA was also met. The assumption for sphericity will be discussed in the Results section. Step-wise regression analysis was also conducted to simultaneously analyze effects of all the factors on the degree of students applying the naïve impetus theory.

Table III. Frequency Distribution by Students' Course Grades

Course grade	Frequency	Percentage	Valid percentage
A	72	11.7	12.7
B	182	29.6	32.2
C	229	37.3	40.5
D	40	6.5	7.1
F	43	7.0	7.6
Missing	48	7.8	
Total	614		

RESULTS

Descriptive Statistics

Table IV presents the descriptive statistics of students who selected distracters of the impetus theory. Because we coded selecting impetus theory choices or not as 1 or 0, the means in Table IV can be interpreted as percentages of students who selected the impetus theory distracters of particular questions. It is interesting to see that no students of any academic achievements who answered question 14 incorrectly selected distracters of the impetus theory, and all students of all academic achievements who answered the question 13 incorrectly selected

Table IV. Descriptive Statistics of Students Who Answered the Questions Incorrectly by Selecting Distracters of the Impetus Theory Arranged by Course Grade

Item	Course grade	Mean/% impetus	Std.	*n*
Q6	A, B	.74	.445	31
	C	.76	.431	42
	D, F	.82	.393	17
Q7	A, B	.23	.428	47
	C	.44	.501	57
	D, F	.37	.492	27
Q8	A, B	.67	.474	55
	C	.65	.481	79
	D, F	.67	.479	30
Q10	A, B	.75	.439	40
	C	.77	.424	95
	D, F	.87	.338	24
Q11	A, B	.72	.452	166
	C	.77	.424	184
	D, F	.77	.425	65
Q12	A, B	.97	.162	38
	C	.99	.121	68
	D, F	1.0	.000	20
Q13	A, B	1.0	.000	134
	C	1.0	.000	171
	D, F	1.0	.000	63
Q14	A, B	.0	.000	72
	C	.0	.000	110
	D, F	.0	.000	37
Q21	A, B	.34	.475	113
	C	.47	.501	135
	D, F	.38	.490	40
Q23	A, B	.62	.506	109
	C	.63	.484	130
	D, F	.59	.487	44
Q24	A,B	.81	.397	32
	C	.80	.407	54
	D, F	.88	.332	17
Q27	A, B	.37	.487	57
	C	.41	.496	75
	D, F	.50	.514	18
Q30	A, B	1.0	.000	156
	C	.99	.110	163
	D, F	.97	.173	66

distracters of the impetus theory. The most popular impetus theory distracters were chosen for items 12 and 30 by almost 100% of students of all academic achievements. The sample sizes of different academic achievement groups for all the items ranged from 17 (D and F students for items 6 and 24) to 184 (C students for item 11).

Table V presents the descriptive statistics of the dependent variables. We see that, among those students who answered the 13 items incorrectly, the degrees of consistency over different groups of items varied greatly from the highest (.87) for explanation questions to the lowest (.48) for prediction and unfamiliar questions. We also see that the consistency over all students and all the items, the overall consistency, is .64, indicating that students applied both the naïve impetus theory and knowledge-in-pieces, but slightly more students applied the naïve impetus theory than knowledge-in-pieces.

Table V. Descriptive Statistics of Dependent Variables

Item type	*n*	Mean	Std
Prediction	527	.48	.34
Explanation	544	.87	.28
Familiar	566	.76	.29
Unfamiliar	498	.48	.38
Overall	587	.64	.28

Effect of Item Familiarity

In order to test the effect of item familiarity on percentage of students selecting distracters from the impetus theory, we combined the items that were classified as "familiar" in Table I into one new dependent variable with the mean scores of individual students as its values. Similarly, we combined the items that were classified as "unfamiliar" in Table I into one new dependent variable with the means scores of individual students as its values. A 2-way mixed design ANOVA (3×2) was conducted with student course grade as the between-subject variable and "familiar" and "unfamiliar" as within-subject variables. The Mauchly's test for sphericity in the two-way ANOVA with repeated measures showed that the equal covariance assumption between the two variables of repeated measures was met. Thus, it was not necessary to adjust the degree of freedom in our interpretation of the 2-way ANOVA results. Table VI presents the ANOVA results. It can be seen that there was a significant effect of item

Table VI. 2-Way ANOVA with Repeated Measures for the Effects of Course Grade and Item Familiarity for Students not Selecting Correct FCI Responses

Source	df	MS	F	P	Partial eta squared
Between-subject					
Grade	2	0.27	2.18	.115	.010
Error	437	0.12			
Within-subject					
Familiarity	1	16.35	189.46	.000*	.302
Familiarity × grade	2	0.29	3.39	.035*	.015
Error	437	0.09			

*$p < .05$.

Table VII. Confidence Intervals of Means of Students Selecting Distracters of the Impetus Theory on Items of Different Familiarity

Course grade	Familiarity	Mean	Std. error	95% Confidence interval Lower bound	Upper bound
A or B	Familiar	.739	.020	.699	.778
	Unfamiliar	.456	.028	.401	.511
C	Familiar	.771	.019	.734	.808
	Unfamiliar	.531	.026	.479	.583
D or F	Familiar	.819	.032	.756	.882
	Unfamiliar	.427	.045	.339	.516

Table VIII. 2-Way ANOVA with Repeated Measures for the Effects of Course Grade and Item Requirement for Explanation vs. Prediction for Students not Answering the FCI Items Correctly

Source	df	MS	F	P	Partial eta squared
Between-subject					
Grade	2	0.11	1.1	.333	.005
Error	444	0.10			
Within-subject					
Item requirement	1	28.92	385.3	.000*	.465
Item requirement × grade	2	0.16	2.2	.117	.010
Error	444	0.08			

*$p < .05$.

familiarity [$F(1, 437) = 189.46$, $p < .05$, $\eta^2 = 0.30$]. There was also a statistically significant interaction effect between the course grade and item familiarity [$F(2, 437) = 3.39$ $p < .05$, $\eta^2 = 0.015$]. Because the effect size for this interaction is only 1.5% variance ($\eta^2 = 0.015$), we consider this interaction effect to be practically insignificant. The effect of student course grade was not statistically significant ($p > .05$).

Table VII presents the confidence intervals for the combinations of course grade and item familiarity. It can be seen that the confidence intervals for familiar items and unfamiliar items do not overlap for all course grades, indicating a main effect of item familiarity.

Effect of Item Explanation or Prediction

In order to test the effect of item requirement for prediction or explanation on percentage of students selecting distracters of the impetus theory, we combined the items that were classified as "prediction" in Table I into one new dependent variable with the mean scores of individual students as its values. Similarly, we combined the items that were classified as "explanation" in Table I into one new dependent variable with the mean scores of individual students as its values. A 2-way mixed design ANOVA (3 × 2) was conducted with student course grade as the between-subject variable and "prediction" and "explanation" as the within-subject variable. The Mauchly's test for sphericity in the 2-way ANOVA with repeated measures showed that the equal covariance assumption between the two variables of repeated measures was met. Thus, we did not need to adjust the degree of freedom in our interpretation of the 2-way ANOVA results presented in Table VIII. It can be seen that there was a statistically significant effect of item requirement for prediction or explanation [$F(1, 444) = 385.28$, $p < .05$, $\eta^2 = 0.465$]. There was no statistically significant main effect of course grade, nor was there a statistically significant interaction effect between course grade and item requirement for prediction or explanation ($p > .05$).

Table IX presents the confidence intervals for the items requiring prediction and explanation. Because the two confidence intervals do not overlap, thus there is a significant difference between the two means. Statistically there are more students selecting distracters of the impetus theory on items requiring explanation ($M = .89$) than items requiring prediction ($M = .49$).

Effects on Overall Consistency

Table X presents the correlation coefficients between various dependent variables. It can be seen that a significant correlation exists between all pairs of the dependent variables. A significant correlation between two dependent variables indicates that students who apply the naïve impetus

Table IX. Confidence Intervals of Means of Students Selecting Distracters of the Impetus Theory on Items Requiring Prediction or Explanation

Item requirement	Mean	Std. error	95% Confidence interval Lower bound	Upper bound
Prediction	.487	.017	.453	.522
Explanation	.886	.014	.860	.913

Table X. Correlation Coefficients Between Dependent Variables

Variable	Prediction	Explanation	Familiar	Unfamiliar
Explanation	.164** ($n = 484$)			
Familiar	.428** ($n = 506$)	.803** ($n = 544$)		
Unfamiliar	.885** ($n = 498$)	.187** ($n = 467$)	.188** ($n = 477$)	
Overall consistency	.803** ($n = 527$)	.678** ($n = 544$)	.827** ($n = 566$)	.722** ($n = 498$)

*$p < .05$.
**$p < .01$.

theory to one groups of questions (e.g. Familiar Questions) are also likely to apply the theory to the other group of questions (e.g. Prediction Questions). It can also be seen that the Overall Consistency, i.e. the mean of all items for each student, is significantly correlated with other dependent variables (e.g. Familiar Questions, Explanation Questions, etc.).

In order to examine the simultaneous effects of all the above factors (Familiar, Unfamiliar, Prediction, and Explanation) as well as the student course grade on the overall consistency, a step-wise linear regression was conducted. The results are presented in Table XI. From Table XI we see that all the variables entered into the regression equation with the exception of Course Grade, which means that those variables all significantly contribute to the overall consistency statistically. Those variables together explained 93.4% of total variance ($R^2 = .934$). However, those variables differ in their degrees of contribution to the prediction of consistency. Familiar items contribute most (with a beta of .5) to the consistency of applying the impetus model, with unfamiliar and prediction items contribute similarly (with a beta of .3), and explanation items contribute least (with a beta of .1). This indicates that the overall degree of consistency depends mostly on item familiarity (the more familiar the higher the degree), and item prediction (the more prediction involved, the higher the degree). Course grade does not have a significant effect in predicting the overall degree of consistency in student application of the impetus model.

Table XI. Regression on Degree of Overall Consistency by Various Factors

Variable	B	Std. error	Beta	*t*
Prediction	.211	.023	.310	9.171**
Familiar	.453	.023	.497	19.924**
Unfamiliar	.194	.019	.322	10.271**
Explanation	.098	.019	.117	5.254**
(Constant)	.028	.010		2.769**

**$p < .01$

DISCUSSION

Answering our research questions, we conclude the following:

1. *Is there a statistically significant difference in percentages of students applying the impetus theory between items that require prediction and items that require explanation?* The findings presented above show that there is a statistically significant difference in percentages of students applying the impetus theory between items requiring prediction and items requiring explanation. Significantly more students apply the impetus theory on items requiring explanation than on items requiring prediction.
2. *Is there a statistically significant difference in percentages of students applying the impetus theory between items that are familiar to them than items that are unfamiliar to them?* The findings presented above show that there is a statistically significant difference in percentages of students applying the impetus theory between familiar items and unfamiliar items. Significantly more students apply the impetus theory on familiar items than on unfamiliar items.
3. *Is there a statistically significant interaction effect between student academic achievements and the above independent variables (i.e. item familiarity, and item explanation/prediction nature)?* The findings presented above show that student academic achievements do not interact with any of the variables.
4. *Taking into consideration all the variables simultaneously, which variables significantly contribute to the percentage of overall consistency by students applying the naïve impetus theory?* The findings presented above show that both item familiarity and item prediction/explanation significantly contribute to

the overall consistency of students applying the naïve impetus theory. However, student course grades do not significantly contribute to this overall consistency.

How can we make sense of the above conclusions? It is clear from our findings that applying mental models/naïve theories or p-prims/knowledge-in-pieces is a matter of degree. Under no conditions do students exclusively apply a mental model or a p-prim. This is consistent with previous studies, such as those by Anderson *et al.* (1992), and Mazens and Lautrey (2003). Because the purpose of this study is to evaluate factors contributing to the degree of students applying a naïve theory, or a mental model, p-prim, and knowledge-in-pieces, it is necessary to consider all the factors at the same time.

Keep in mind that our analysis excluded the correct responses made by students that are consistent with Newtonian theories. It is reasonable to assume that at least some students who did not answer the FCI questions correctly hold the impetus naïve theory. Item familiarity affects significantly the percentages of students applying the naïve impetus theory as concluded above. This could be explained by the fact that generating mental models depends on students' personal experiences. If an item is familiar and easy, students are more likely to be able to generate specific mental models to answer the question. Table X also indicates that Explanation has a very high correlation with Familiar Questions, which suggests that familiar questions are more likely to require explanation, thus resulting in higher degrees of consistency. On the other hand, if the item is less familiar, students may be less likely to apply their personal experiences to generate mental models, or may be more likely to resort to random guessing, which decreases the overall percentage of consistency in applying the naïve impetus theory.

If a question requires explanation, because the impetus theory is a coherent system, the impetus theory gives those students ability to generate answers consistently due to the requirement of a theory to explain (Brewer, 1999; Schwitzgebel, 1999). This accounts for a significantly higher percentage of students applying the impetus theory to explanation items. If a question requires students to make predictions, then students need to generate a mental model from a theory to make a prediction (Johnson-Laird, 1983). Because mental models are generated on the spot and different students may use different naïve theories, mental models are more likely to vary from student to student and from problem task to problem task. Therefore, questions requiring students to predict have more potential to produce inconsistency or variation than questions requiring students to explain. This explains why prediction questions contribute most significantly to the overall percentage of students applying the impetus theory in the regression analysis. Explanation only contributes minimally (although statistically significantly) to the overall percentage in the regression because students only need to apply a theory such as the impetus theory to explain, which potentially results in more consistency and less variation or variance.

Student course grade did not contribute significantly to the percentages of students applying the naïve impetus theory. This also suggests that, if a student answered a FCI question incorrectly, no matter what the student's grade is, the student may go through a same mental process involving the impetus theory or knowledge-in-pieces.

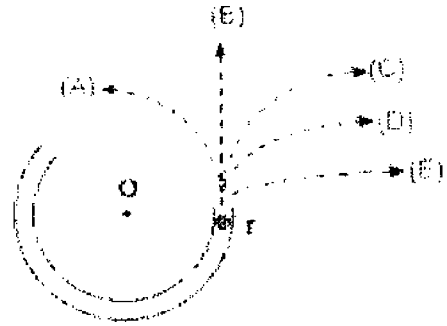
The above mental process assumes that students have a desire to explain using a theory and to predict using a mental model, which is consistent with views of mental models (Johnson-Laird, 1983; Gentner and Stevens, 1983; Vosniadou, 1992, 1994; Vosniadou and Brewer, 1992) and theory theory (Brewer and Samarapungavan, 1991; Carey, 1985; Gopnik and Meltzoff, 1997; Schwitzgebel 1999; Wellman, 1990). However, the desire to explain and predict may be moderated by other factors such as item familiarity, which provides a potential for students to apply p-prims or knowledge-in-pieces. Thus, naïve theories, mental models, p-prims, and knowledge-in-pieces are closely related in a student's mental process of answering a question.

The above findings suggest that we need to conceptualize naïve theories, mental models, p-prims, and knowledge-in-pieces into a coherent theoretical framework. From a pedagogical point view, we believe that as a goal of science education, we must promote the current consensus models or the scientific theories, which is promoted by some researchers (e.g. Duit and Treagust, 2003; Gilbert, 1999; Schwitzgebel, 1999). However, achieving that goal may require students to develop modeling abilities in applying their own theories – although sometimes these may be naïve theories or less complete theories than the current consensus models in science. We may conceptualize a hierarchy from p-prims to naïve mental models or theories to scientific consensus models or theories, and it is the responsibility of science teachers to help students make the progression from p-prims to appropriate scientific consensus models. This conceptualization is consistent with the constructivist view on knowledge in transition (Smith *et al.*, 1993). Smith *et al.* argue that knowledge-in-pieces and scientific theories share certain functionality and should be viewed in continuity and in a systemic framework. Attempting to "unlearn" student naïve conceptions may be neither plausible nor even desirable.

The above conceptualization is also consistent with the conceptual change views put forward in the literature. For example, Linder (1993) argues that conceptual change involves changes in understanding of relations between constructs and their contexts (dealing with p-prims); Ebenezer and Gaskell (1995), Marton and Booth (1997), as well as Liu (2004) promoted the notion of relational conceptual change that is similar to Linder's. Other conceptual change views (see reviews by Duit and Treagust, 2003; Tyson *et al.*, 1997) promote such theory change. Whatever conceptual change view is to be adopted, we must incorporate *both* p-prims and students' naïve theories. The facet clusters that Minstrell (1991, 2001) and colleagues have been developing over the past few years appear to be based on the hierarchical continuum from p-prims to scientific consensus models. Of course, the continuum we suggest here is much broader than what facet clusters imply. We perceive that each major science concept entails student progression in understanding from multiple p-prims to multiple naïve mental models with various degrees of power in making consistent predictions and explanation, and onwards to the scientific theory that has the most power in making most consistent predictions and explanation. Thus, we perceive a continuous trajectory of student conceptual development in science concepts starting with pieces of ideas and ending with scientific theories as consensus models. If we can map out such a trajectory for each of the major concepts learned in science curricula, we will be in a much stronger position to develop instruction and assessment that are much more appropriate and relevant to student learning.

APPENDIX: FCI QUESTIONS USED (ITEM CLASSIFICATIONS ARE INDICATED AT THE END OF EACH ITEM IN ITALIC)

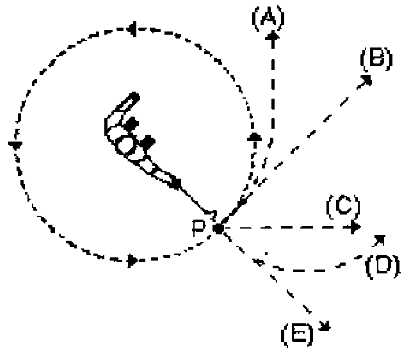
Q6. Which path in the figure at right would the ball most closely follow after it exits the channel at "r" and moves across the frictionless table top? [*Unfamiliar, Prediction*]



Q7. A steel ball is attached to a string and is swung in a circular path in a horizontal plane as illustrated in the accompanying figure.

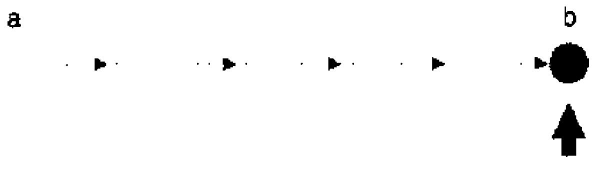
At the point P indicated in the figure, the string suddenly breaks near the ball.

If these events are observed from directly above as in the figure, which path would the ball most closely follow after the string breaks? [*Familiar, Prediction*]

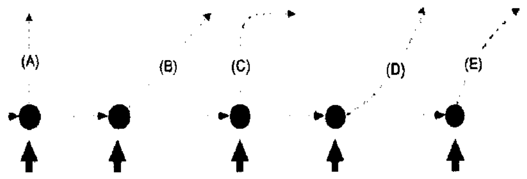


USE THE STATEMENT AND FIGURE BELOW TO ANSWER THE NEXT FOUR QUESTIONS (8 THROUGH 11)

The figure below depicts a hockey puck sliding with constant speed v_0 in a straight line from point "a" to point "b" on a frictionless horizontal surface. Forces exerted by the air are negligible. You are looking down on the puck. When the puck reaches point "b," it receives a swift horizontal kick in the direction of the heavy print arrow. Had the puck been at rest at point "b," then the kick would have set the puck in horizontal motion with a speed v_k in the direction of the kick.



Q8. Which of the paths below would the puck most closely follow after receiving the kick? [*Familiar, Prediction*]



Q10. Along the frictionless path you have chosen in question 8, the speed of the puck after receiving the kick:

(A) is constant.
(B) continuously increase.
(C) continuously decreases.
(D) increases for a while and decreases thereafter.
(E) is constant for a while and decreases thereafter.

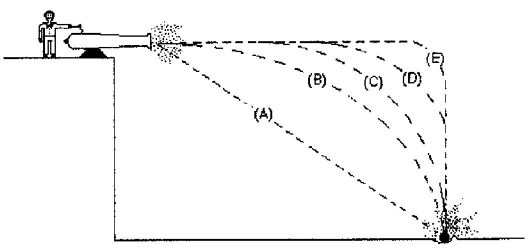
[*Familiar, Prediction*]

Q11. Along the frictionless path you have chosen in question 8, the main force(s) acting on the puck after receiving the kick is (are):

(A) a downward force of gravity.
(B) a downward force of gravity, and a horizontal force in the direction of motion.
(C) a downward force of gravity, an upward force exerted by the surface, and a horizontal force in the direction of motion.
(D) a downward force of gravity and an upward force exerted by the surface.
(E) none. (No forces act on the puck).

[*Familiar, Explanation*]

Q12. A ball is fired by a cannon from the top of a cliff as shown in the figure below. Which of the paths would the cannon ball most closely follow? [*Unfamiliar, Prediction*]



Q13. A boy throws a steel ball straight up. Consider the motion of the ball only after it has left the boy's hand but before it touches the ground, and assume that forces exerted by the air are negligible. For these conditions, the force(s) acting on the ball is (are):

(A) a downward force of gravity along with a steadily decreasing upward force.
(B) a steadily decreasing upward force from the moment it leaves the boy's hand until it reaches its highest point; on the way down there is a steadily increasing downward force of gravity as the object gets closer to the earth.
(C) an almost constant downward force of gravity along with an upward force that steadily decreases until the ball reaches its highest point; on the way down there is only a constant downward force of gravity.
(D) an almost constant downward force of gravity only.
(E) none of the above. The ball falls back to ground because of its natural tendency to rest on the surface of the earth.

[*Familiar, Explanation*]

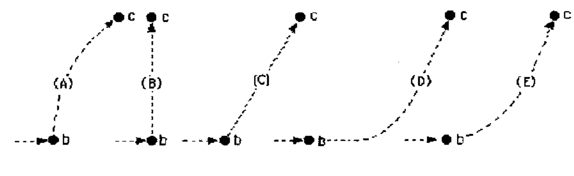
Q14. A bowling ball accidentally falls out of the cargo bay of an airliner as it flies along in a horizontal direction.

As observed by a person standing on the ground and viewing the plane as in the figure at right, which path would the bowling ball most closely follow after leaving the airplane? [*Unfamiliar, Prediction*]

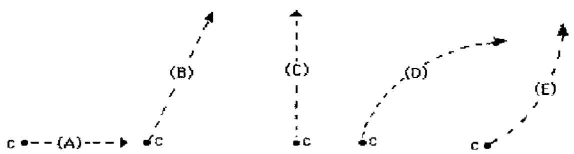
USE THE STATEMENT AND FIGURE BELOW TO ANSWER THE NEXT FOUR QUESTIONS (21 THROUGH 24).

A rocket drifts sideways in outer space from point "a" to point "b" as shown below. The rocket is subject to no outside forces. Starting at position "b," the rocket's engine is turned on and produces a constant thrust (force on the rocket) at right angles to the line "ab." The constant thrust is maintained until the rocket reaches a point "c" in space.

Q21. Which of the paths below best represents the path of the rocket between points "b" and "c"? [*Unfamiliar, Prediction*]



Q23. At point "c" the rocket's engine is turned off and the thrust immediately drops to zero. Which of the paths below will the rocket follow beyond point "c"? [*Unfamiliar, Prediction*]



Q24. Beyond position "c" the speed of the rocket is:

(A) constant.
(B) continuously increasing.
(C) continuously decreasing.
(D) increasing for a while and constant thereafter.
(E) constant for a while and decreasing thereafter.

[*Unfamiliar, Prediction*]

Q27. A woman exerts a constant horizontal force on a large box. As a result, the box moves across a horizontal floor at a constant speed "v_0". If the woman suddenly stops applying a horizontal force to the box, then the box will:

(A) immediately come to a stop;
(B) continue moving at a constant speed for a while and then slow to a stop.
(C) immediately start slowing to a stop.
(D) continue at a constant speed.
(E) increase its speed for a while and then start slowing to a stop.

[*Familiar, Prediction*]

Q30. Despite a very strong wind, a tennis player manages to hit a tennis ball with her racquet so that the ball passes over the net and lands in her opponent's court.

Consider the following forces:

1. A downward force of gravity.
2. A force by the "hit."
3. A force exerted by the air.

Which of the above forces is (are) acting on the tennis ball after it has left contact with the racquet and before it touches the ground?

(A) 1 only.
(B) 1 and 2.
(C) 1 and 3.
(D) 2 and 3.
(E) 1, 2, and 3.

[*Familiar, Explanation*]

ACKNOWLEDGMENTS

We acknowledge support from the US Department of Education through the Arizona Teacher's Excellence Coalition (AzTEC) project in data collection. We would also like to thank the Purdue University Physics Department for providing the opportunity to collect data.

REFERENCES

Alexander, P. A. (2003). Can we get there from here? *Educational Researcher* 32: 3–4.

Anderson, T., Tolmie, A., Howe, C., Mayes, T., and MacKenzie, M. (1992). Mental models of motion. In Rogers, Y., Rutherford, A., and Bibby P. A. (Eds.), *Models in Mind: Theory, Perspective, and Application* Academic Press, London, pp. 57–71.

Cole, R. P., and MacIsaac, D. L. (2001). *A comparison of paper-based and web-based assessments*. American Educational Research Association: Divison K (ERIC Document Reproduction Service No. TM 032 375).

MacIsaac, D. L., Cole, R. P., and Cole, D. M. (2002). Standardized testing in physics via the World Wide Web. *Electronic Journal of Science Education* 6(3).

Bedard, H., and Chi, M. T. H. (1992). Expertise. *Current Directions in Psychological Science* 1: 135–139.

Brewer, W. (1999). Scientific theories and naive theories as forms of mental representation: Psychologism revived. *Science & Education* 8: 489–505.

Brewer, W., and Samarapungavan, A. (1991). Children's theories vs. scientific theories: Differences in reasoning or differences in knowledge? In Hoffman, R., and Palermo, D. (Eds.), *Cognition and the Symbolic Processes,* Erlbaum, Hillsdale, NJ pp. 209–232.

Carey, S. (1985). *Conceptual Change in Childhood*. MIT Press, Cambridge, MA.

Chi, M. T. R. (1978). Knowledge structure and memory development. In Siegler, R. (Ed*.), Children's Thinking: What Develop*? Erlbaum, Hillsdale, NJ, pp. 73–96.

Chi, M. T. H. (1992). Conceptual change within an across ontological categories: Examples from learning and discovery in science. In Giere, R. N. (Ed.), *Cognitive Models of Science: Minnesota Studies in the Philosophy of Science,* Vol. 15, University of Minnesota Press, Minneapolis, pp. 129–186.

Chi, M. T. H., Feltovich, P. J., and Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science* 5: 121–152.

Chi, M. T. H., and Koesk, R. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology* 19: 29–39.

Chi, M. T. H., Glaser, R., and Farr, M. J. (1988). *The Nature of Expertise*. Erlbaum, Hillsdale, NJ.

Chi, M. T. H., and Roscoe, R. D. (2002). The processes and challenges of conceptual change. In Limon, M., and Mason, L. (Eds.), *Reconsidering Conceptual Change: Issues in Theory and Practice*. Kluwer Academic Publishers, Dordrecht, the Netherlands, pp. 3–27.

Chi, M. T. H., Slotta, J. D., and deLeeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction* 4: 27–43.

Chi, M. T. H., Glaser, R., and Rees, E. (1982). Expertise in problem solving. In Sternberg, R. (Ed.), *Advances in the Psychology of Human Intelligence,* Vol. 1, : Erlbaum, Hillsdale, NJ.

Chiu, M. H., Chou, C. C., and Liu, C. J. (2002). Dynamic processes of conceptual change: Analysis of constructing mental models of chemical equilibrium. *Journal of Research in Science Teaching* 39: 688–712.

Coll, R. K., and Treagust, D. F. (2003a). Investigation of secondary school, undergraduate, and graduate learners' mental models of ionic bonding. *Journal of Research in Science Teaching* 40: 464–486.

Coll, R. K., and Treagust, D. F. (2003b). Learners' mental models of metallic bonding: A cross-age study. *Science Education* 87: 685–707.

De Jong, T., Martin, E., Zamarro, J. M., Esquembre, F., Swaak, J., and van Joolingen, W. R. (1999). The integration of computer simulation and learning support: An example from the physics domain of collisions. *Journal of Research in Science Teaching* 36: 597–615.

Driver, R. (1983). *The Pupil as Scientist*. Open University Press, Milton Keynes.

Driver, R., and Easley, J. (1978). Pupils and paradigms: A review of literature related to concept development in adolescent science students. *Studies in Science Education* 5: 61–84.

diSessa, A. A. (1988). Knowledge in pieces. In Forman, G., and Pufall, P. B. (Eds.), *Constructivism in the Computer Age*. Lawrence Erlbaum, Hillsdale, NJ.

diSessa, A. (1993). Toward an epistemology of physics. *Cognition and Instruction* 10: 105–226.

diSessa, A. A., and Sherin, B. L. (1998). What changes in conceptual change? *International Journal of Science Education* 20: 1155–1191.

Duit, R., and Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education* 25: 671–688.

Ebenezer, J., and Gaskell, P. J. (1995). Relational conceptual change in solution chemistry. *Science Education* 79: 1–17.

Erickson, G. L. (1994). Pupils' understanding magnetism in a practical assessment context: The relationship between content, process and progression. In Fensham, P., Gunstone, R., and White, R. (Eds.), *The Content of Science: A Constructivist Approach to Its Teaching and Learning*. The Falmer Press, London.

Gentner, D., and Stevens, A. L. (Eds.) (1983). *Mental Models*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Giere, R. (1988). *Explaining Science*. University of Chicago Press, Chicago.

Gilbert, J. K. (1999). On the explanation of change in science and cognition. *Science and Education* 8: 543–557.

Gilbert, J. K., and Boulter, C. J. (Eds.) (2000). *Developing models in science education*. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Gopnik, A., and Meltzoff, A. (1997). *Words, Thoughts, and Theories*. MIT Press, Cambridge, MA.

Gopnik, A., and Wellman, H. M. (1994). The theory theory. In Hirschfeld, L. A., and Gelman, S. A. (Eds.), *Domain Specificity in Cognition and Culture*. Cambridge University Press, New York.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics* 66: 64–74.

Hallouin, I., and Hestenes, D. (1985a). The initial knowledge state of college physics students. *American Journal of Physics* 53: 1043–1055.

Hallouin, I., and Hestenes, D. (1985b). Common sense concepts about motion. *American Journal of Physics* 53: 1056–1064.

Hestenes, D., and Halloun, I. (1995). Interpreting the Force Concept Inventory: A response to Huffman and Heller. *The Physics Teacher* 33: 502–506.

Hestenes, D., Wells, M., and Swackhamer, G. (1992/1995). Force concept inventory. *The Physics Teacher* 30: 141–158.

Huffman, D., and Heller, P. (1995). What does the Force Concept Inventory actually measure*? The Physics Teacher* 33: 138–143.

Johnson-Laird, P. N. (1983). *Mental Models: Toward a Cognitive Science of Language, Inference, and Consciousness*. Harvard University Press, Cambridge, MA.

Keil, F. C. (1989). *Concepts, Kinds and Cognitive Development*. MIT Press, Cambridge, MA.

Keil, F. C. (1998). Cognitive science and the origins of thought and knowledge. In Lerner, R. M. (Ed.), *Theoretical Models of Human Development,* Wiley, New York, pp. 341–413.

Korfiatis, K., Papatheodorou, E., Stamou, G. P., and Paraskevopoulous, S. (1999). An investigation of the effectiveness of computer simulation programs as tutorial tools for teaching population ecology at university. *International Journal of Science Education* 21: 1269–1280.

Lave, J. (1978). *Cognition in Practice: Mind, Mathematics, and Culture in Everyday Life*. Cambridge University Press, Cambridge, UK.

Lave, J., and Wenger, E. (1991). *Situated Learning: ligitimate Peripheral Participation*: Cambridge University Press, Cambridge, UK.

Linder, C. J. (1993). A challenge to conceptual change. *Science Education* 77: 293–300.

Linn, M. C., and Muilenburg, L. (1996). Creating lifelong science learners: What models form a firm foundation? *Educational Researcher* 25: 18–24.

Liu, X. (2004). Using computerized concept mapping for promoting and assessing student relational conceptual change. *Science Education* 88: 373–396.

Mazens, K., and Lautrey, J. (2003). Conceptual change in physics: children's naive representations of sound. *Cognitive Development* 18: 159–176.

Marton, F., and Booth, S. (1997). *Learning and Awareness*.: Erlbaum, Mahwah, NJ.

McCloskey, M. (1983). Naikve theories of motion. In Gentner, D., and Stevens, A. L. (Eds.), *Mental Models*, Laurence Erlbaum, Hillsdale, NJ.
Minstrell, J. (1991). Facets of students' knowledge and relevant instruction. In Duit, R., Goldberg, F., and Niedderer, H. (Eds.), *Research in Physics Learning: Theoretical Issues and Empirical Studies*, University of Bremen, Germany, pp. 110–128.
Minstrell, J. (2001). Facets of students' thinking: Designing to cross the gap from research to standard-based practice. In Crowley, K., Schunn, C. D., and Okada, T. (Eds.), *Designing for Science: Implications from Everyday, Classroom, and Professional Settings,* Lawrence Erlbaum Associates, Mahwah, NJ, pp. 415–443.
Monaghan, J. M., and Clement, J. (1999). Use of a computer simulation to develop mental simulation for understanding relative motion concepts. *International Journal of Science Education* 21: 921–944.
Piaget, J., and Garcia, R. (1989). *Psychogenesis and the History of Science*. Columbia University Press, New York.
Rogers, Y., Rutherford, A., and Bibby, P. A. (Eds.) (1992). *Models in the Mind: Theory, Perspective and Application,* Academic Press, London.
Schneider, W., Gruber, H., Gold, A., and Opwis, K. (1993). Chess expertise and memory for chess position in children and adults. *Journal of Experimental Child Psychology* 56: 328–349.
Schwitzgebel, E. (1999). Children's theories and the drive to explain. *Science and Education* 8: 457–488.
Smith, J. P., diSessa, A. A., and Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences* 3: 115–163.
Stewart, J., Hafner, R., Johnson, S., and Finkel, E. (1992). Science as model building: Computers and high-school genetics. *Educational Psychologist* 27: 317–336.
Taber, K. S. (2003). Mediating mental models of metals: Acknowledging the priority of the learner's prior learning. *Science Education* 87: 732–758.
Taylor, I., Barker, M., and Jones, A. (2003). Promoting mental model building in astronomy education. *International Journal of Science Education* 25: 1205–1225
Treagust, D. F., Chittleborough, G., and Mamiala, T. L. (2002). Students' understanding of the role of scientific models in learning science. *International Journal of Science Education* 24: 357–368.
Tyson, L. M., Venville, G. J., Harrison, A. G., and Treagust, D. F. (1997). A multidimensional framework for interpreting conceptual change in the classroom. *Science Education* 81: 387–229.
Vosniadou, S. (1992). Knowledge acquisition and conceptual change. *Applied Psychology: An International Review* 41: 347–357.
Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction* 4: 45–69.
Vosniadou, S., and Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Pyschology* 24: 535–585.
Wandersee, J. H., Mintzes, J., and Novak, J. (1994). Research on alternative conceptions in science. In Gabel, D., (Ed.), *Handbook of Research on Science Teaching and Learning*. MacMillan Pulishing Company, New York.
Wellman, H. M. (1990). *The Child's Theory of Mind*, MIT Press, Cambridge.
Windschitl, M. (2001). Using simulations in middle school: Does assertiveness of dyad partners influence conceptual change? *International Journal of Science Education* 23: 17–32.