JOURNAL OF RESEARCH IN SCIENCE TEACHING VOL. 46, NO. 5, PP. 570–596 (2009)

# Knowledge Structure Coherence in Turkish Students' Understanding of Force

Gokhan Ozdemir,[1] Douglas Clark[2]

*[1]College of Education, Department of Science Education, Nigde University, Derbent Campus, Nigde 51100, Turkey*
*[2]College of Education, Division of Curriculum & Instruction, Arizona State University, Tempe, Arizona*

*Received 22 November 2007; Accepted 6 November 2008*

Abstract: This study investigates Turkish students' knowledge structure coherence in physics. In particular, this study investigates the conflicting findings reported in Ioannides and Vosniadou's [Ioannides and Vosniadou [2002] Cognitive Science Quarterly, 2, 5–61] and diSessa, Gillespie, and Esterly's [diSessa et al. [2004] Cognitive Science, 28, 843–900] studies about students' understandings of force. Ioannides and Vosniadou's study of four different age levels of students in Greece demonstrated broad consistency in students' understandings of force. diSessa and colleagues' quasi-replication in the United States demonstrated conflicting results supporting a more fragmented elemental perspective on students' knowledge structure coherence. The current study investigates these conflicting findings by studying students in a third country using the analytic methods from both studies to clarify the debate over knowledge structure coherence. The levels of consistency demonstrated by students in the current study are somewhat higher than the levels reported by diSessa, Gillespie, and Esterly according to both coding schemes, but are closer overall to the levels reported by diSessa, Gillespie, and Esterly than to the levels reported by Ioannides and Vosniadou. In addition, closer inspection of students' explanations suggests that students' explanations may code as consistent according to the coding schemes for a particular force meaning category but not actually represent a coherent understanding of that force meaning. These results therefore more closely support fragmented elemental perspectives on knowledge structure coherence. The results, however, demonstrate important systematicities in students' thinking and support the possibility that differences between the student populations in the countries of the original studies contributed to the differences in findings of the original studies.

**Keywords:** conceptual change; misconceptions; cognitive science; science education

## Introduction

Is a student's knowledge most accurately represented as a coherent unified scheme of theory-like character (e.g., Carey, 1999; Chi, 2005; Ioannides & Vosniadou, 2002; Wellman & Gelman, 1992)? Or is a student's knowledge more aptly considered as an ecology of quasi-independent elements (e.g., diSessa, Gillespie, & Esterly, 2004; Harrison, Grayson, & Treagust, 1999; Linn, Eylon, & Davis, 2004)? These questions carry important implications for curricular design and cognitive research. From a constructivist perspective, students learn by restructuring and building upon their prior knowledge. Thus, different structures of students' prior knowledge implicate different instructional strategies to help students reorganize those structures. Similarly, from a research perspective, a great deal of research focuses on how students' ideas change over time. Essentially, to better understand the processes of conceptual change, we must better understand the structure of the knowledge that is undergoing those changes.

Differences in subject matters, age groups, and methods often complicate comparisons between studies conducted by proponents of theory-like perspectives and elemental perspectives. Recently, however, diSessa,

*Correspondence to*: G. Ozdemir; E-mail: gokhanozdemir@nigde.edu.tr
DOI 10.1002/tea.20290
Published online 20 March 2009 in Wiley InterScience (www.interscience.wiley.com).

Gillespie, and Esterly (2004) conducted a quasi-replication in the United States of a study originally conducted by Ioannides and Vosniadou (2002) in Greece focusing on students' understanding of force in physics. Although involving similar methods, the studies resulted in dramatically different findings. The current study addresses this debate by studying students in a third country using both sets of analytic methods to clarify the debate over knowledge structure coherence.

*Overview of Theory-Like Perspectives*

Theory-like perspectives hypothesize that naïve knowledge involves unified coherent structures based on persistent beliefs consisting of ontological and epistemological presuppositions (e.g., Carey, 1999; Chi, 2005; Ioannides & Vosniadou, 2002; Wellman & Gelman, 1992). Proponents of theory-like perspectives often characterize conceptual change as a wholesale shift from one concept to another, either in terms of Piagetian assimilation and accommodation or Kuhn's (1962) concepts of normal science and scientific revolution (e.g., Carey, 1985, 1999; Wiser & Carey, 1983). While some advocates argue that scientists' theories are similar to naïve knowledge structures, proponents of these perspectives also generally acknowledge differences with respect to lack of meta-conceptual awareness and hypothesis testing (Vosniadou, 2002). In summary, proponents of theory-like perspectives assert that naïve knowledge structures are highly coherent in the form of theories and that these theories allow learners to make consistent predictions and explanations across significant science domains.

*Overview of Elemental Perspectives*

Elemental perspectives hypothesize that naïve knowledge involves collections of multiple quasi-independent elements (e.g., Brown, 1995; Clark, 2006; diSessa, 1988, 1993; Hunt & Minstrell, 1994; Linn et al., 2004) such as p-prims (phenomenological primitives), intuitive knowledge pieces, facts, facets, narratives, and concepts at varying levels of development and specification. Because these simple knowledge pieces are not organized and constrained by overarching theories or schemes, naïve knowledge is often fragmented in its application across contexts. Elemental perspectives should not be incorrectly caricatured, however, as the random interaction of independent elements. Rather, elements interact with each other in an emergent manner where the combinatorial complexity of the system constrains students' interpretations of phenomenon. During conceptual change, these simple knowledge pieces undergo a process of restructuring and reorganization. This is a piecemeal evolutionary change rather than a revolutionary change.

*Significant Findings From Ioannides and Vosniadou (2002)*

Ioannides and Vosniadou (henceforth referred to as "I&V") investigated the meaning of force and its development among 105 students across four grade levels in Greece (essentially pre-k, elementary school, middle school, and high school). In their standardized sets of questions, I&V presented students with pictures of simple stick models and asked the students questions about the forces on the objects in the pictures (Figure 1). I&V also asked comparison questions to explore students' interpretations of force in greater detail (e.g., falling stones vs. stationary stones, a man trying to move a stone vs. the same man trying to move a smaller stone, unstable stones on the top of a hill vs. stable similar stones on the top of a hill). I&V developed a qualitative coding scheme that first distilled students' answers down to basic explanations and then coded the basic explanations into force meanings using a rubric.

I&V found that 88.6% of the students' responses fell into one of seven categories of internally consistent interpretations of force. The seven categories include (a) internal force, (b) internal force affected by movement, (c) internal and acquired force, (d) acquired force, (e) acquired force and force of push-pull, (f) force of push-pull, and (g) gravitational and other forces. A full explanation of each force meaning is

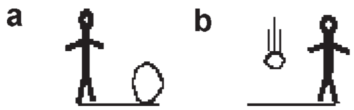



*Figure 1.* Samples question from I&V's questionnaire: (a) standing big stone, and (b) falling small stone.

provided in the Methods Section. For example, a student who has an *internal force* meaning believes that force is always related to an objects' size or weight. This student's predictions and explanations will therefore be consistent with this *internal force* meaning across all question sets regardless of context. I&V interpreted their findings of high consistency as supporting their claims regarding the coherent nature of students' knowledge.

### *Significant Findings From diSessa, Gillespie, and Esterly (2004)*

diSessa, Gillespie, and Esterly (henceforth referred to as ''DG&E'') conducted a study with thirty U.S. students that involved a condensed quasi-replication version of I&V's (2002) study along with additional extension questions. DG&E distilled and reorganized I&V's questions down to ten question sets for the quasi-replication component of the study. DG&E also extrapolated I&V's qualitative coding scheme into a coarse-quantitative coding scheme because they did not feel that they could reliably apply the qualitative scheme to their student interviews. In transforming the analytic methodology, DG&E attempted to create a coding scheme that would categorize students as being ''consistent'' more liberally than I&V's original coding scheme.

DG&E found that the U.S. students did not exhibit the same consistency across question sets that I&V had observed with their Greek students. Only 17% of the 30 U.S. students were completely consistent for a force meaning. Even when DG&E lowered the criterion for consistency to only require matching eight of ten question sets, only 43% of the students were consistent and the vast majority of these students were categorized in the *gravity and other* meaning, which DG&E considered suspect. DG&E interpreted their data as supporting a fragmented perspective on knowledge structure coherence in which naïve knowledge consists of unstructured small pieces that are unconsciously activated according to context (diSessa, 1993).

### *Purpose of the Current Study*

One hypothesis explaining the differences in findings between I&V's and DG&E's original studies focuses on the methodological differences between I&V's and DG&E's coding schemes. While DG&E attempted to create a coding scheme that coded students as ''consistent'' more liberally than the original I&V qualitative scheme, it is entirely possible that the differences in the schemes contributed to the substantial differences in findings. The current study therefore first compares the two coding schemes across a common data set. If the current study were to discover that the two coding schemes coded students substantially differently in terms of consistency, the current study might conjecture that coding scheme differences contributed significantly to the differences in findings of the original studies. If the two coding schemes code students similarly, however, comparing the levels of consistency demonstrated by the Turkish students with the levels of consistency demonstrated by the Greek and U.S. students in the original studies will help determine whether the findings for the Turkish students align more closely with DG&E's or I&V's findings in terms of consistency (and thus align more closely with the theoretical explanations proposed in those original studies). Finally, the current study also investigates the degree to which Turkish speakers express similar or different explanations about force than the Greek and U.S. students to examine other possible distinctions between the student populations in each country. Through these three analyses, the current study provides a foundation for our group's ongoing program of research investigating the debate over students' knowledge structure coherence and conceptual change processes across cultures (e.g., Clark, D'Angelo, & Schleigh, in preparation; Clark, Menekse, Ozdemir, & D'Angelo, in preparation; Schleigh, 2008).

## Methods

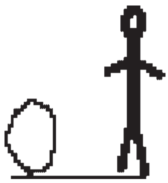
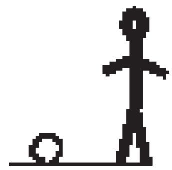
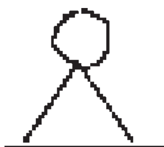
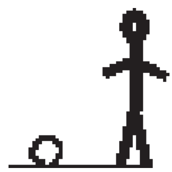
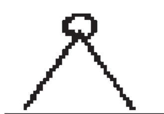
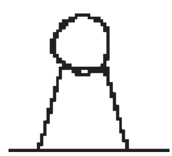
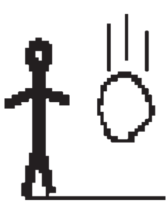
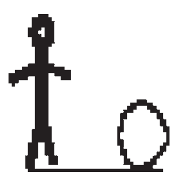
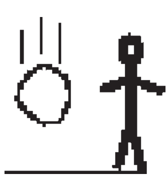
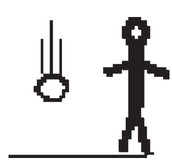
As outlined above, this study replicates DG&E's (2004) and I&V's (2002) work by applying the analytic methods from those studies to a third cultural/linguistic context.

### *Instrument*

Students were asked the ten sets of replication questions that DG&E (2004) condensed from I&V's (2002) questions. Table 1 presents these ten question sets.

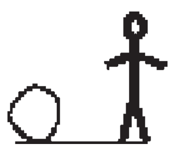
Each question set consists of two simple questions and one comparison question. In each set, students were first asked if there is a force on the stone in the first picture. Students were then asked if there is a force in

Table 1
*The ten sets*

| Set | Drawing A | Question A | Drawing B | Question B | Comparison question |
|---|---|---|---|---|---|
| (1) Big vs. small stones standing on the ground |  | “This stone is standing on the ground. Is there a force on this stone? Why?” |  | “This stone is standing on the ground. Is there a force on this stone? Why?” | “Is the force on this stone the same or different than the force on this stone? Why” |
| (2) Unstable vs. stable similar stones standing on a hill |  | “This stone is standing on a hill. It is unstable. That means it could easily fall down. Is there a force on the stone? Why?” |  | “This stone is standing on a hill. It is stable. That means it will not easily fall down. Is there a force on the stone? Why?” | “Is the force on this stone the same or different than the force on this stone? Why” |
| (3) Unstable small vs. unstable big stones standing on a hill |  | “This stone is standing on a hill. It is unstable. That means it could easily fall down. Is there a force on the stone? Why?” |  | “This stone is standing on a hill. It is unstable. That means it could easily fall down. Is there a force on the stone? Why?” | “Is the force on this stone the same or different than the force on this stone? Why” |
| (4) Falling big vs. standing big stones |  | “This stone is falling. “Is there a force on the stone? Why?” |  | “This stone is standing on the ground. Is there a force on this stone? Why?” | “Is the force on this stone the same or different than the force on this stone? Why” |
| (5) Falling big vs. falling small stones |  | “This stone is falling. Is there a force on the stone? Why?” |  | “This stone is falling. “Is there a force on the stone? Why?” | “Is the force on this stone the same or different than the force on this stone? Why” |
| (6) A man trying to move a big stone vs. small stone |  | “This man is trying to move this stone. Is there a force on the stone? Why?” |  | “This man is trying to move this stone. Is there a force on the stone? Why?” | “Is the force on this stone the same or different than the force on this stone? Why” |

*(Continued)*

Table 1
*(Continued)*

| Set | Drawing A | Question A | Drawing B | Question B | Comparison question |
|---|---|---|---|---|---|
| (7) A man trying to move a big stone vs. small stone but he cannot move either |  | "This man is trying to move this stone and it will not move. Is there a force on the stone? Why?" |  | "This man is trying to move this stone and it will not move. Is there a force on the stone? Why?" | "Is the force on this stone the same or different than the force on this stone? Why" |
| (8) A man trying to move a big stone vs. a child trying to move a big stone but they both fail |  | "This man is trying to move this stone and it will not move. Is there a force on the stone? Why?" |  | "This child is trying to move this stone and it will not move. Is there a force on the stone? Why?" | "Is the force on this stone the same or different than the force on this stone? Why" |
| (9) A man throwing a stone vs. a similar stone standing on the ground |  | "This man has thrown this stone. Is there a force on the stone? Why?" |  | "This stone is standing on the ground. Is there a force on this stone? Why?" | "Is the force on this stone the same or different than the force on this stone? Why" |
| (10) A man throwing a small stone vs. throwing a big stone |  | "This man has thrown this stone. Is there a force on the stone? Why?" |  | "This man has thrown this stone. Is there a force on the stone? Why?" | "Is the force on this stone the same or different than the force on this stone? Why" |

the second picture. Finally, if the student identified forces in both pictures, the student was asked to compare the forces in the two pictures. Students were asked to explain their answers throughout. The comparison question, when applicable, provided more information related to the student's understanding of force in terms of strength and contextual-related differences.

### *Subjects and Procedures*

The study was conducted with 32 Turkish students in order to parallel DG&E's sample of thirty students. Interviews were conducted at a typical K-8 school and a typical high school in Ankara, the capital of Turkey. As with I&V's and DG&E's studies, this study involved four different age groups of students: eight preschool students, eight elementary school students, eight middle school students, and eight high school students. The approximate mean ages were 5, 10, 13, and 16 years, respectively. All of the students spoke Turkish as their first language. Equal numbers of male and female students from middle class families were selected to participate. This was a random selection in terms of students' academic achievements. Students were selected from the same schools and district so that they presumably had relatively similar schooling backgrounds and out-of-school experiences.

The lead author was the interviewer. He collaborated closely with the teachers in selecting the students. The teachers and the interviewer informed the students about the research and asked the students if they would be willing to participate in the interviews. After identifying volunteers, the teachers of the classrooms selected the most talkative students so that the interviewer could collect enough information about the students' understandings of force. All students were interviewed individually for ~20–25 minutes. Students were asked all questions in one session. All interviews were videotaped.

### *Description of Coding Schemes and Data Analysis*

Each student's responses were analyzed for each question set. To address the possibility that the differences in the findings between I&V (2002) and DG&E (2004) resulted from differences in their schemes, this study separately applied both schemes to each student.

*Ioannides and Vosniadou's Coding Scheme.* The first scheme, which was developed by I&V (2002), consists of scoring at the question set level in terms of response categories that are then mapped to overall force meanings.

*Scoring for response categories at the question set level*. Students' responses for each question set were scored using a scoring key containing categories of responses for that question set. I&V referred to this as scoring at the question set level. It should be noted that DG&E reorganized the question sets such that simple and comparison questions were included in the same sets, which was not the case in I&V's study. Response categories for each question set were determined using rubrics converted from I&V's study to fit DG&E's ten-question-set format. The response categories for Question Set 1 are presented in Table 2.

*Scoring at the overall level*. Following the scoring at the question set level, the scoring at the overall level was based on a second scoring key that outlined the pattern of expected response categories for each of seven force meanings identified through I&V's research of the literature and their initial analyses of the data. Table 3 outlines I&V's criteria for each force meaning in detail. Table 4 shows the mapping of the question set response categories to overall level force meanings for Set 1.

*diSessa, Gillespie, and Esterly's Coding Scheme.* The second scheme was adapted by DG&E from I&V's study as exemplified for Set 1 in Table 5. DG&E's scheme is more holistic than I&V's in that students' explanations were not coded for each question and integrated into an overall code. Instead, students' answers for each question set were mapped directly onto potential force meaning matches. Furthermore, DG&E's scheme is less qualitative in that it only considers the relative forces in conjunction with exemptions for specific explanations. DG&E refer to this approach as being "coarse-quantitative." DG&E compared students' responses to expected patterns for I&V's seven meanings to identify any consistent matches across the question sets.

### *Transcript Examples*

This section presents two examples from the interviews to demonstrate how the coding schemes apply to the students' responses. The first example from the transcripts is for an elementary school student, Senay, whose answers for Set 1 match the *internal, internal/movement, and internal/acquired* force meanings. The second example from the transcripts is for a high-school student, Aysel, whose answers match for the *gravity and other* force meaning for Set 3.

*Senay's Transcript Segment: Set 1—(A) Big versus (B) Small Stones Standing on the Ground.*

Question A

I: This stone is standing on the ground. Is there a force on this stone? (Showing drawing A)

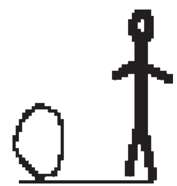
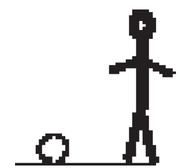
S: Yes.

I: Why?

S: Because, there is force on the stone and the weight of the stone is the reason for this force.

Question B

Table 2
*Response categories for question Set 1 (question set level scoring)*

| | | | | | |
|---|---|---|---|---|---|
| (1) Big vs. small stones standing on the ground |  | "This stone is standing on the ground. Is there a force on this stone? Why?" |  | "This stone is standing on the ground. Is there a force on this stone? Why?" | "Is the force on this stone the same or different than the force on this stone? Why" |

| Set 1 response categories | Response category description |
|---|---|
| (a) Force only on the big stone | Because the big stone is big and/or heavy and/or you cannot move it. No force on the small stone because it is small and/or light and/or you can move it easily |
| (b) Force on both stones but greater force on the big stone | Because both stones are heavy or they have weight but the first stone is bigger and/or heavier and/or you cannot move it |
| (c) Force of gravity on both stones | Same force on both stones. It is the force of gravity, the Earth's attraction |
| (d) Alternative interpretation of the force of gravity | Greater force of gravity or Earth's attraction on the big stone because it is heavier/has greater weight |
| (e) No force on any stones | Because they are not moving |
| (f) Force on the small stone, no force on big stone | Because the big stone is heavy and/or no one can move it easily. Because small stone is light and/or you can move it easily |
| (g) No force on any stones because no one pushes them | Because no one pushes them |
| (h) Force from the air on both stones | It is the force from the air above the stones. Same force on both stones because both stones are standing on the ground |

Table 3
*Criteria for each of the seven coherent force meanings identified by I&V*

(1) *Internal force*: Students were assigned to this meaning of force if they gave answers indicating that there is a force on all objects or only on big/heavy objects because they have weight or are big/heavy. Students do not refer to gravity, the object's motion, or another agent
(2) *Internal force affected by movement*: Students were assigned to this meaning of force if they gave answers indicating that the force is due to size/weight of object only, but also if moving objects and objects that are likely to fall have less *internal* force than do stationary objects
(3) *Internal and acquired*: Students were assigned to this meaning of force if they indicated that there is a force on stationary objects due to size/weight and that these objects acquired an additional force when they are set in motion. I&V included students in this category who were ambivalent about unstable objects and interpreted unstable objects as either lacking *internal* force or likely to acquire additional force
(4) *Acquired*: Students who indicated that force is a property of objects that explains motion and potentially acts on other objects were assigned to this meaning of force. These students answered that there is no force on stationary objects, and the force on moving objects disappears when the object stops moving. They also included students who thought that force was only acquired by heavy, moving objects and claim that this response indicates that these students relate the *acquired* force to both the weight and the motion of the object. Additionally, I&V included students who thought that unstable stones had more force because they could be set in motion more easily as well as those who thought all stones (stable and unstable) could be set in motion easily in this category
(5) *Acquired and force of push–pull*: Students were assigned to this meaning if they gave answers meeting the criteria described above for the *acquired* meaning of force, but also answered that there was a force on an object when acted on by an agent regardless of whether or not it moves
(6) *Force of push–pull*: Students were assigned to this meaning if they indicated that a force was exerted only on objects being pushed by an agent, whether or not the object was moving
(7) *Force of gravity and others*: Students were assigned to this meaning if they mentioned gravity and other forces. Students could be considered consistent with the "*gravity and other*" category for question Sets 6 (a man trying to move a big stone vs. small stone), 7 (a man trying to move a big stone vs. small stone but he cannot move either), and 8 (a man trying to move a big stone vs. a child trying to move a big stone but they both fail) even if they did not mention the word gravity in these sets (See response categories *b*, *c*, and *e* for the comparison questions of stationary objects which are pushed by a human agent in I&V study)

I: This stone is standing on the ground. Is there a force on this stone? (Showing drawing B)

S: Yes.

I: Why?

S: Because, the same reason; it can be either small or big; this does not matter.

I: What is the reason?

S: The reason is the weight of the stone.

Comparative Question

I: Is the force on this stone the same or different than the force on this stone?

S: Different.

I: Why:

S: Because, one is big and one is small.

I: And . . . Why is there a force?

S: If the stone is big, we cannot move it, but if it is small we can easily move it.

I: On which stone is the force greater?

S: The force on the big stone is greater than the force on small one.

*Application of I&V's coding scheme to Senay's responses for Set 1.* I&V's coding scheme for Set 1 codes a student's explanations for the *internal*, *internal/movement*, and *internal/acquired* force meanings

Table 4
*Conversion of set level response categories in overall level force meanings for Set 1 (big vs. small stones standing on the ground)*

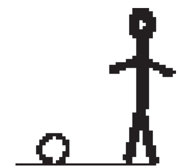
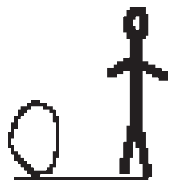
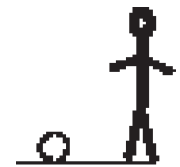
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| (1) Big vs. small stones standing on the ground |  | "This stone is standing on the ground. Is there a force on this stone? Why?" |  | "This stone is standing on the ground. Is there a force on this stone? Why?" | "Is the force on this s tone the same or different than the force on this stone Why" | | |
| Overall meaning of force | (1) Internal | (2) Internal/Move | (3) Internal/Acq | (4) Acq | (5) Acq/P–P | (6) Push–pull | (7) Gravity |
| | a, b | a, b | a, b | e, g | e, f, g, h | g | c, d |
| Set level response categories mapping to each overall force meaning (see full list in Table 2) | Force only or greater on the big stone because bigger and/or heavier and/or you cannot move it | Force only or greater on the big stone because bigger and/or heavier and/or you cannot move it | Force only or greater on the big stone because bigger and/or heavier and/or you cannot move it | No force on any stones because they are not moving | No force on any stones because they are not moving | No force on any stones because no one pushes them | Force of gravity on both stones |
| | | | | No force on any stones because no one pushes them | No force on any stones because no one pushes them | | Greater force of gravity on the big stone |
| | | | | | Force only on the small stone | | |
| | | | | | Force from the air | | |

Table 5
*DG&E's coding scheme for question Set 1 (big vs. small stones standing on the ground)*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| (1) Big vs. small stones standing on the ground |  | "This stone is standing on the ground. Is there a force on this stone? Why?" | |  | "This stone is standing on the ground. Is there a force on this stone? Why?" | | "Is the force on this stone the same or different than the force on this stone? Why?" |
| Meaning of force Set 1—big vs. small stones standing on the ground | (1) Internal<br>Force only on the big stone, but not due to air, gravity, or ground<br>Force on both stones but greater force on the big stone, but not due to air, gravity, or ground | (2) Internal/Move<br>Force only on the big stone, but not due to air, gravity, or ground<br>Force on both stones but greater force on the big stone, but not due to air, gravity, or ground | (3) Internal/Acq<br>Force only on the big stone, but not due to air, gravity, or ground<br>Force on both stones but greater force on the big stone, but not due to air, gravity, or ground | (4) Acq<br>No force on any stone | (5) Acq/P–P<br>No force on any stone<br>Force only on the small stone but not due to gravity | (6) Push–Pull<br>No force on any stone | (7) Gravity<br>Equal force on both stones<br>Force on both stones but greater force on the big stone |

categories if the explanations describe a force on the stones due to their own weight or size. Senay assigned forces to both stones due to ‘‘the weight of the stones.’’ She also indicated that ‘‘the force on the big stone is greater than the force on the small one.’’ This response exactly matches the question set response category that ‘‘there is force on both stones because both stones are heavy or they have weight but the first stone is bigger and/or heavier and/or you cannot move it.’’ Therefore, Senay’s response is coded as matching the *internal*, *internal/movement*, and *internal/acquired* force meanings categories for the overall scoring level for Set 1.

*Application of DG&E’s coding scheme to Senay’s responses for Set 1*. DG&E’s coding scheme also codes Senay’s explanations for Set 1 as *internal*, *internal/movement* and *internal/acquired* because explanations for this set are coded for these meanings if they describe a greater force for the big stone not due to air, gravity, or ground. As seen in the transcript segment above, Senay’s responses did not include air-related, gravity-related forces, or any force from the ground when she assigned a greater force to the big stone in Set 1.

*Aysel’s Transcript Segment: Set 3—(A) Unstable Small versus (B) Unstable Big Stones Standing on a Hill*

Question A

I: This stone is standing on a hill. It is unstable. That means it could easily fall down. Is there a force on the stone? (Showing drawing A)

A: Yes.

I: Why?

A: Because of gravitational force, it may fall down; gravity helps it to fall. . .

Question B

I: This stone is standing on a hill. It is unstable. That means it could easily fall down. Is there a force on the stone? (Showing drawing B)

A: Yes, but it is bigger than the first one.

I: Why?

A: Same again; due to gravitational force. . .

I: You said it is bigger, can you explain?

A: Gravity is proportional to the weight of the objects, this stone looks bigger.

Comparative Question

I: Is the force on this stone the same or different than the force on this stone?

A: Different.

I: Why?

A: Because gravitational forces depend on the masses of the stones; the force acting on the second stone is greater than the force on first stone because second stone is heavier than the first one.

*Application of I&V’s coding scheme to Aysel’s responses for Set 3*. Aysel clearly expressed gravitational force for both stones in her explanations. Aysel assigned a larger gravitational force to the big stone due to its larger size. At the question set level, this response is coded as ‘‘alternative interpretation of the force of gravity.’’ This response is then placed in *gravity and other* force meaning category at overall level scoring for Set 3.

*Application of DG&E’s coding scheme to Aysel’s responses for Set 3*. According to DG&E’s coding scheme for Set 3, a student’s response is assigned to the *gravity and other* force meaning category if the response describes equal gravitational forces on both stones or greater gravitational force on the big stone.

*The 20% Error Allowance.* After coding all of the questions, students' consistency across the 10 question sets was checked for both coding schemes against the seven force meanings identified by I&V. Additional possible consistent meanings were explored when they arose (e.g., force from being alive). A student who matched for a force meaning across all 10 question sets was categorized as "fully consistent" for that force meaning. A looser criterion was also applied allowing students to be considered consistent if they used the same force meaning on at least eight of the ten sets. This is in addition to the other specific exemptions outlined in Table 3 for the *gravity and other* meaning. This error allowance was included in DG&E's study. I&V did not need to employ an error allowance in their study because most of their students (88.6%) were coded as fully consistent.

*Inter-Rater Reliability.* All interviews were first coded by the first author in Turkish. The interviews were then translated to English and 25% of the students at each grade level were randomly selected for coding by a second trained coder from our research group to ascertain inter-rater reliability. Inter-rater reliability was 93.5% in terms of agreement between the two coders for the force meanings assigned for each question set for each student for each coding scheme.

## Results and Discussion

We first present and discuss the results for each age group in terms of consistency of force meanings for each coding scheme. We then discuss the implications of the results in terms of (1) whether differences in the coding schemes could have been responsible for the differences in the findings in the original studies, (2) which theoretical perspectives from the original studies are supported by the results of the current study, and (3) how the Turkish students compare to the U.S. and Greek students in terms of knowledge structure coherence and force meanings.

### *Pre-School Students*

Two of the eight pre-school students were strictly consistent according to both schemes (Table 6). This means that only two students consistently interpreted force across all 10-question sets. Both of these students applied the *internal force* meaning, consistently connecting force to the size/weight of the stones. If we apply an error allowance to the coding (which means allowing up to 2 out of the 10 question sets to not match a force meaning before deeming that student as not being consistent with that force meaning), five additional students' responses fell into one of two sub-categories of the *internal* force meaning (*internal/movement* or *internal/acquired*) or *force of living* category according to I&V's scheme (for a total of seven out of eight students). Only three additional students' force meanings were determined as consistent with the error allowance using DG&E's scheme (for a total of five out of eight students). These were three of the five students whose force meanings were identified as consistent with the error allowance for I&V's scheme (two with the *internal/acquired* force meaning and one with the *internal/movement* force meaning).

Table 6
*Pre-school participants' consistency table with and without error allowance*

| | Internal | Int/Move | Int/Acq | Acq | Acq/P–P | Push–pull | F to living | Fragmented |
|---|---|---|---|---|---|---|---|---|
| Sena | ID | | | | | | | |
| Nazli | | | | | | | | ID |
| Ali | | | I*D* | | | | | |
| Hasan | | I*D* | | | | | | |
| Ayse | | | I*D* | | | | | |
| Meral | | | | | | | I* | |
| Faruk | | I* | | | | | | |
| Can | ID | | | | | | | |

I, I&V's scheme; I*, I&V's scheme with error allowance.
D, DG&E's scheme; D*, DG&E's scheme with error allowance.

Appendix presents an expanded table showing the coding of students' consistent force meanings with and without error allowance when the two schemes are applied separately to the students' responses.

*Discussion of Pre-School Students.* The pre-school students demonstrated consistent predictions for the *internal force* meaning for the six question sets that focused on differences in the size of the stones. These six questions sets include Set 1 (big vs. small stones standing on the ground), Set 3 (unstable small vs. unstable big stones standing on a hill), Set 5 (falling big vs. falling small stones), Set 6 (a man trying to move a big stone vs. small stone), Set 7 (a man trying to move a big stone vs. small stone but he cannot move either), and Set 10 (a man throwing a small stone vs. throwing a big stone). In these sets, stones are on the ground, at the top of hills, falling, pushed, pushed, and thrown, respectively. In other words, both stones from each set were in the same state, but the stones were different in size. The students consistently gave answers that supported *internal force* meanings for these sets that compared stone size.

A potential issue with the error allowance becomes clear with the pre-school students' responses for Set 4 (falling big vs. standing big stones) and Set 9 (a man throwing a stone vs. a similar stone standing on the ground). These sets compare stones of the same size in different states. In Set 4 (falling big vs. standing big stones), the first stone is falling down while the second stone is standing on the ground. In Set 9 (a man throwing a stone vs. a similar stone standing on the ground), the first stone has been thrown while the second stone is standing on the ground. Pre-school students' causal responses varied when the objects were in different states in Sets 4 (falling vs. standing stones) and 9 (throwing vs. standing stones). Tables 7 and 8 show that only two students interpreted those two contexts with an *internal force* meaning. Four different force meanings other than the *internal* force meaning in Set 4 and five different force meanings other than the *internal* force meaning in Set 9 were observed for these two critical sets. Sets 4 and 9, however, could be ignored by the error allowance, masking the students' interpretation of a significantly different context from the six sets that compare stone sizes. Coding a student as consistent with the error allowance therefore does not necessarily demonstrate that the student maintains a coherent force meaning because important question contexts can be entirely eliminated.

*Odd reasoning when young students don't understand the meaning of the word "force."* It seems possible that the contexts led the students to give answers related to the size of the stone for the six sets that focus on stone size. The interview begins with a simple comparison of two stones of different sizes standing on the ground. It is possible that this led students to assign a greater force to the big stones when they do not really know what "force" means. Although the interviews suggest internal force meanings for the six question sets focusing on stone size, some confusion with the questions was apparent during the interviews. A few students insisted on assigning a greater force to one stone "because the stone is bigger" in question sets where the pictures of the two stones are exactly the same size. One student thought that there was no force on any stones because they were not alive. Another student repeatedly said that one stone is bigger than the other stone in one context even though the two stones are exactly the same size in the pictures. Moreover, students sometimes avoided giving any causal explanations at all.

*Summary: Pre-School Students.* In summary, only two of the eight students coded as consistent across the question sets without the error allowance. The error allowance brings this number up to seven out of eight for I&V's scheme and five out of eight for DG&E's scheme. This error allowance, however, may not validly represent coherence of understanding. Sets 4 and 9 resulted in different force meanings than the core *internal* force meaning, but the error allowance permitted these two sets to be dismissed. The six question sets focusing on differing stone sizes, however, resulted in generally consistent interpretations aligning with the *internal* force meaning.

Table 7
*Variation of pre-school students' force meanings in Set 4*

| Set 4 | Internal | Int/Move | Int/Acq | Both acq and acq/P–P | Push–pull | Gravity | F to living |
|---|---|---|---|---|---|---|---|
| Number of pre-school students | 2 | 2 | 2 | 2 | 0 | 0 | 0 |

Table 8
*Variation of pre-school students' force meanings in Set 9*

| Set 9 | Internal | Int/Move | Int/Acq | Both acq. and acq/P–P | Push–pull | Gravity | F to living |
|---|---|---|---|---|---|---|---|
| Number of pre-school students | 2 | 2 | 1 | 2 | 0 | 0 | 1 |

Two other potential issues were noticed. First, the term "force" was a barrier for pre-school students. The samples from the interviews above indicated that some pre-school students were confused about what the term meant in the given contexts. It seemed that they did not have a clear definition or meaning for the term "force." Second, the structure of the question sets supported *internal* force meanings because six of the question sets focus on differences in stone size. We are concerned that this may have led pre-school students to focus on stone size in attributing forces for these questions, thereby potentially over-representing the *internal* force meaning.

### *Elementary School Students*

Both I&V's and DG&E's schemes code only one elementary school student as demonstrating a consistent force meaning. That student was consistent in terms of the *gravity and other* force meaning for both I&V's and DG&E's schemes (Table 9). Applying the 20% error allowance causes three additional students to be coded as consistent for I&V's scheme and four for DG&E's scheme. Rana demonstrated consistent *internal* force meanings for both I&V's and DG&E's schemes using the error allowance. Utku demonstrated consistent *acquired* force meanings for both schemes and a consistent *acquired/push–pull* force meaning for I&V's scheme using the error allowance. Sezen demonstrated a consistent *internal/movement* force meaning only for I&V's scheme with the error allowance. Kenan (*internal/movement*) and Mert (*acquired*) demonstrated consistent force meanings only for DG&E's scheme. As Table 9 shows, elementary school students' force meanings varied considerably.

*Discussion of Elementary School Students.* Several elementary school students, including the students who demonstrated consistent meanings of force according to I&V's and DG&E's schemes, assigned more than one kind of force to each set in their explanations. Those multiple forces were not consistent in the students' causal explanations. One force would appear in one context but the same force would be absent in another context. For example, a few students talked about force from the hills on the stones standing at the top of the hills but the same students did not assign force from the ground for the standing stones. This indicated that those elementary students' responses were contextually sensitive.

As examples of this contextual sensitivity, Engin's interpretation of force includes all force meanings, except the core *internal* force meaning, according to I&V's and DG&E's schemes. However, none of his force

Table 9
*Elementary school participants' consistent table with and without error allowance*

| | Internal | Int/Move | Int/Acq | Acq | Acq/P–P | Push–pull | Gravity | Fragmented |
|---|---|---|---|---|---|---|---|---|
| Kenan | | D* | | | | | | |
| Mert | | | | D* | | | | |
| Rana | I*D* | | | | | | | |
| Sezen | | I* | | | | | | |
| Nur | | | | | | | ID | |
| Senay | | | | | | | | ID |
| Engin | | | | | | | | ID |
| Utku | | | | I*D* | I* | | | |

I, I&V's scheme; I*, I&V's scheme with error allowance.
D, DG&E's scheme; D*, DG&E's scheme with error allowance.

meanings are consistent across contexts. His interpretation of force changes from context to context. For example, in Set 1 (big vs. small stones standing on the ground), Engin did not assign a force to the standing stones due to "no external effect on the stones." When the interviewer probed his reasoning with the question, "In which cases can we say there is a force on this stone?" he expressed the verbs "push, pull, and lift." This response codes for the *acquired*, *acquired/push–pull*, and *push–pull* force meanings because his explanations clearly refer to the lack of push, pull, and/or motion in the contexts. However, for the same standing stone contexts in Sets 4 (falling big vs. standing big stones) and 9 (a man throwing a stone vs. a similar stone standing on the ground), he changed his reasoning. In Set 4, (falling big vs. standing big stones) he assigned a force to the standing stone because the "ground is carrying this stone." In Set 9 (a man throwing a stone vs. a similar stone standing on the ground), he assigned a force to the standing stone because the "ground is carrying it because of gravity." These explanations code as the *gravity and other* force meaning. How different contexts affect Engin's explanations are more apparent in Set 2 (unstable vs. stable similar stones standing on a hill) in terms of the *gravity and other* force meaning. While he assigned a gravitational force to the stable stone on the hill, he only attributed the *internal movement* and *internal acquired* force meanings to the unstable stone on the hill due to its instability. He thought that there was a force on the unstable stone because "the peak is sharp so the stone can fall down easily."

Similarly, Sezen codes as consistent for the *internal/movement* force meaning if an error allowance is applied with I&V's scheme. However, akin to Engin, Sezen demonstrated inconsistencies in her explanations. In Set 1 (big vs. small stones standing on the ground), for example, Sezen, did not assign a force to the small stone because "it is light and small," but she thought that there was a force on the big stone because the ground was pushing the big stone. This explanation codes as an *internal* force meaning because she related force to the size of the stones. However, her explanations were not simply based on the *internal* force meaning because she also stated a "force from the ground on the big stone only." Her explanations also included the *gravity and other* force meaning because the explanation of "force from the ground" implies the existence of the gravitational force. In her thinking, gravity or force of the ground applies only to big objects. Furthermore, she did not assign the same force to the other standing stones in Sets 2 (unstable vs. stable similar stones standing on a hill) and 3 (unstable small vs. unstable big stones standing on a hill). However, she did assign the same gravitational force to the standing stones in Sets 4 (falling big vs. standing big stones) and 9 (a man throwing a stone vs. a similar stone standing on the ground). It was not clear which stones she considered to be under the effect of a force from the ground in her explanations. An example from Sezen's interview demonstrates her *gravity and other* force meaning for the standing stone in Set 4 (falling big vs. standing big stones).

I: This stone is standing on the ground. Is there a force on this stone?

S: Yes.

I: Why?

S: Because, it is standing.

I: What is the force acting on it?

S: The line at the bottom.

I: Which line?

S: This one. (Showing the line at the stone.)

I: Do you mean the ground?

S: Yes; the ground is exerting a force on it.

*Summary: Elementary School Students.* Both schemes coded only one out of eight elementary students as fully consistent. I&V's scheme coded four students as consistent with the error allowance. DG&E's scheme coded five students as consistent with the error allowance. The students often assigned multiple forces to the contexts without an obvious pattern. Although the error allowance allows some students to be coded as consistent, their responses usually included inconsistencies as demonstrated in the transcripts of their interviews.

*Middle School Students*

Five middle school students expressed fully consistent force meanings according to I&V's scheme. Four of these were consistent for the *gravity and other* force meaning. In terms of I&V's scheme, one of the four students also expressed fully consistent *internal* and consistent *acquired/push–pull* force meanings. Another one of the five expressed a consistent *acquired* force meaning only for I&V's scheme.

With the error allowance, five of the eight middle school students coded as consistent for more than one force meaning according to I&V's scheme (Table 10). Veli demonstrated the *internal*, *acquired/push–pull*, and *gravity and other* force meanings. Dilek demonstrated the *acquired* and *acquired/push–pull* force meanings. Emine and Emre demonstrated the *acquired/push–pull* and *gravity and other* force meanings. Sibel demonstrated the *acquired/push–pull* and *push–pull* force meanings. In fact, with the error allowance, only one student did not demonstrate a consistent force meaning according to I&V's scheme.

Four middle school students expressed fully consistent force meanings according to DG&E's scheme. Applying DG&E's scheme with the error allowance results in force meaning consistency for Dilek, Sibel, and Erkan in addition to the four students.

*Discussion of Middle School Students.* As discussed for the preschool and elementary students, the middle school students' force meanings in this study may not be best categorized as highly organized theory-like systems regardless of how the schemes code them for consistency. For example, clear fragmentation of force meanings appeared in Pelin's case. She interpreted the pushing contexts through the *acquired/push–pull* and *push–pull* force meanings in Set 6 (a man trying to move a big stone vs. small stone) because she thought that ''the man is exerting a force to the stone.'' On the other hand, she interpreted the similar pushing contexts through the *acquired* force meaning in Sets 7 (a man trying to move a big stone vs. small stone but he cannot move either) and 8 (a man trying to move a big stone vs. a child trying to move a big stone but they both fail). She did not assign a force to the stones in Set 7 and 8 because she thought that ''the stones did not change their locations'' even when the stones were pushed by the agents. For example, the quotations below, which were taken from Pelin's interview about the pushing contexts in Set 7 (a man trying to move a big stone vs. small stone but he cannot move either) demonstrate how her reasoning is different compared to her reasoning for a similar context in Set 6 (a man trying to move a big stone vs. small stone).

I: This man is trying to move this stone and it won't move. Is there a force on the stone?

P: No.

I: Why?

Table 10
*Middle school participants' consistent table with and without error allowance*

| | Int | Int/Move | Int/Acq | Acq | Acq/P–P | Push–pull | Gravity | Fragmented |
|---|---|---|---|---|---|---|---|---|
| Emine | | | | | I* | | ID | |
| Pelin | | | | | | | | ID |
| Veli | I | | | | I | | ID | |
| Dilek | | | | ID* | I*D* | | | |
| Bilal | | | | | | | ID | |
| Sibel | | | | | I*D* | I* | | |
| Emre | | | | | I* | | ID | |
| Erkan | | | | | | | I*D* | |

I, I&V's scheme; I*, I&V's scheme with error allowance.
D, DG&E's scheme; D*, DG&E's scheme with error allowance.

P: When the objects change their location or position we call it a force. And in this case, the stone is stationary, which means there is no force.

Her causal explanations indicated a *gravity and other* force meaning in the other sets, but her interpretations of gravity were varied, non-normative, unusual, and inconsistent across sets. She did not assign a gravitational force to the object pushed by the human agents in Sets 6 (a man trying to move a big stone vs. small stone), 7 (a man pushing against, but not moving, a big stone vs. small stone), and 8 (a man pushing against, but not moving, a big stone vs. a child pushing against, but not moving, a big stone). Nor did she assign a gravitational force to the unstable stones standing on a hill due to the small surface area between the hill and the stone in Sets 2 (unstable vs. stable stones standing on a hill) and 3 (unstable small vs. unstable big stones standing on a hill). From her perspective in these contexts, there should be complete contact between the object and the ground in order for there to be a gravitational force. As she explained, ‘‘the surface area between the unstable stones and the hills is too small so the unstable stones cannot apply a force to the hill. In turn, gravity does not apply a force to the unstable stones.’’
Pelin’s statement suggests that she believed that the force of gravity was zero on the unstable stone, but that there was the force of gravity on the standing stone because the force of gravity worked against the stone’s pressure on the ground. On the other hand, she inconsistently assigned a gravitational force to the falling object in Sets 4 (falling big vs. standing big stones) and 5 (falling big vs. falling small stones) due to falling. If gravitational force requires full physical contact between an object and the ground or hill in her conceptualization, then she should not assign a gravitational force to the falling object—and yet she did. The quotations below, which were taken from Set 1 (big vs. small stones standing on the ground), demonstrate Pelin’s confusion regarding gravitational force and her use of terminology:

I: Is there a force on this stone (big stone)?

P: Yes, gravity. There is gravitational force because of its volume. Its volume and its mass change acceleration of gravity.

I: Is there a force on this stone (small stone)?

P: Gravity does not change by the size of the object. The object that has greater weight, mass, and volume has greater acceleration of gravity as well.

I: Is the force on this stone (big stone) the same or different than the force on this stone (small stone)?

P: Those stones may be different sizes. But, if their masses are equal, then the force of gravity does not change. If their masses are different, then there will be differences in the force of gravity on the stones. If their masses are different, then their sizes, weights, and volumes are different. Since the force from gravity is based on the size and weight, the object having greater mass has greater force on it.

In sum, while Pelin uses the term ‘‘gravity’’ often, she does not appear to maintain a clear, consistent, and coherent causal scheme for gravity’s impact.

Pelin’s interview highlights the under-specification of the coding categories regarding forces from the air and gravity in terms of the meanings that the middle school students tended to express (see Table 10). Because students’ specific interpretations of force from the air and gravity are significantly underspecified by the schemes, the students’ explanations appear to fit these ‘‘underspecified’’ categories more often than seems warranted. This is a problem for several reasons because the most important aspect of coherence from our perspective involves consistency of causality in students’ explanations.

According to I&V’s scheme, for example, students can be assigned to the *acquired/push–pull* force meaning category if they simply said ‘‘force from the air above the stones.’’ Unfortunately, there is no explanation to justify ‘‘force from the air’’ as an *acquired/push–pull* force meaning. It could possibly be explained with an Aristotelian view that all bodies by their very nature have a natural way of moving. Movement is not the result of the influence of one body on another. However, the students’ explanations were not that simple. The students expressed many inconsistent interpretations related to force from the air including: (1) air pressure above the stones, (2) air pressure below the stone against its weight, (3) air pressure shared by the stones, (4) greater air pressure on the big stone or small stone, (5) air force on the standing stone only, (6) frictional force from the air, and (7) air resistance. In spite of these multiple ideas related to air and the inconsistency of the air-related interpretations, they were classified under the *acquired/push–pull* force meaning as suggested by I&V's scheme, but it seems clear that I&V's scheme may be too soft to confidently conclude that half of the middle school students really demonstrate an *acquired/push–pull* force meaning because of these inconsistencies related to forces from air.

The under-specification of the *gravity and other* force meaning is also problematic. In both schemes, students were assigned to the *gravity and other* force meaning category if the students assigned the same gravitational force to both stones or a greater gravitational force to the big stones. I&V's rubrics, however, allowed students to match for the *gravity and other* force meaning in the "pushing" contexts (Sets 6–8) without actually referring to gravity or forces from the ground. Note that this softness in I&V's rubrics is independent from, and in addition to, the error allowance that we incorporated from DG&E's study. We applied the same scheme to classify the students' responses for the purposes of replication, but many of the students' explanations that coded for the *gravity and other* force meanings did not match the meaning projected by I&V particularly well. Those multiple interpretations of gravitational force and the inconsistencies between them have already been discussed in the previous sections.

*Summary: Middle School Students.* Several of the middle school students can be categorized as consistent for *gravity and other* without the error allowance and for the *acquired/push–pull* categories with the error allowance. Their causal responses, however, appear varied and inconsistent upon closer inspection. It seems possible that softness within the schemes leads to students being coded as consistent even when they do not express coherent force meanings.

### *High School Students*

According to both schemes, four of the eight high school students were coded as fully consistent for the *gravity and other* force meaning. With the error allowance applied to I&V's scheme, Sami and Ece were also coded as consistent for the *gravity and other* force meaning while Gul and Inci were coded as consistent for the *acquired/push–pull* force meaning. All high school students therefore were coded as consistent with the error allowance using I&V's scheme (Table 11). With the error allowance applied to DG&E's scheme, two additional students were coded consistent for the *gravity and other* force meaning for a total of six students.

*Discussion of High School Students.* Similar to the middle school students, the high school students usually expressed *acquired/push–pull* and *gravity and other* force meanings. The high school students usually assigned a greater gravitational force to the falling stones and the big stones. Most of the time, they ignored the gravitational force for all pushing contexts (Sets 6–8). They usually interpreted those contexts with the *acquired/push–pull* and *push–pull* meanings. As with the students at other grade levels, however, the high school students expressed conflicting answers across sets even for these meanings.

Ece, for example, demonstrated the most fragmented meanings across sets. She sometimes interpreted the contexts and was coded with the *acquired* and *acquired/push–pull* force meanings, sometimes with the

Table 11
*High school participants' consistent table with and without error allowance*

| | Int | Int/Move | Int/Acq | Acq | Acq/P–P | Push–pull | Gravity | Fragmented |
|---|---|---|---|---|---|---|---|---|
| Ece | | | | | | | I*D* | |
| Gul | | | | | I* | | | |
| Inci | | | | | I* | | | |
| Aysel | | | | | | | ID | |
| Remzi | | | | | | | ID | |
| Sami | | | | | | | I*D* | |
| Kamil | | | | | | | ID | |
| Sakir | | | | | | | ID | |

I, I&V's scheme; I*, I&V's scheme with error allowance.
D, DG&E's scheme; D*, DG&E's scheme with error allowance.

*gravity and other* force meaning, and sometimes with the *acquired/push–pull* and *push–pull* force meanings. Ece usually did not explicitly include ‘‘gravity’’ as a force. She did assign a gravitational force in Set 5 (falling big vs. falling small stones), but she did not assign a gravitational force to another falling context in Set 4 (falling big vs. standing big stones). Although she said, ‘‘the stone is falling with gravity,’’ she did not say that gravity was a factor in Set 4 (falling big vs. standing big stones) when she stated:

> There is no force. This stone is in the air. If something is in the air, there won’t be a force on it. It is falling, it has velocity. Is there a force from the air? I am not sure about that. There is no force. Because it is in the air, there is no movement. There is a movement actually. It is confusing. For example, you drop the stone from this height, it is falling. There is no effect on it. It is falling with gravity. Might there be a gravitational force on it? Anyway, I say no.

In Set 5 (falling big vs. falling small stones), she changed her mind and said that gravity is a for

> No force. Might gravity be a force here? I’m confused. I’ll ask my physics teacher when I get to school. This stone is in the air. That is why there is no force on the stone. It is gravitational force. This stone is in the air and there is not any effect on it.

In Set 4 (falling big vs. standing big stones) and Set 9 (a man throwing a stone vs. a similar stone standing on the ground), Ece assigned a force from the ground to the standing stone contexts, but she did not think in the same way for the standing stone in Set 1 (big vs. small stones standing on the ground). In Set 1 (big vs. small stones standing on the ground) she said that ‘‘there is no force on the stones because they are just standing there . . . If somebody put a hand on it or kicks it then there will be a force on it.’’

In Set 2 (unstable vs. stable similar stones standing on a hill) and Set 3 (unstable small vs. unstable big stones standing on a hill), which both involve standing stones at the top of the hills, Ece suggested that the hill applied a force to the stones in order to keep the stones in balance. This idea fit the criteria of the schemes as *gravity and other* force meaning even though this answer was not a perfect conceptual fit for the category. Ece’s response for the unstable stone in Set 2 (unstable vs. stable similar stones standing on a hill) indicated a ‘‘balance’’ idea:

> If it is not falling then there is a force . . . It is placed at the middle. I think there is a force but there might not be a force. The hill applies a force to the stone. The stone is in balance at the top of the hill so the sharp hill may apply a force to the stone.

For the unstable small stone in Set 3 (unstable small vs. unstable big stones standing on a hill), she expressed a similar reasoning saying:

> Yes. What might be the reason for that? Gravity? No, it is not related . . . I will say the same thing. If the top of hill were not sharp, I would say there is no force. If the top of hill were big, the stone would stay there like it is standing on the floor. The top of hill applies a force to the stone.

In summary, Ece expressed multiple fragmented contradictory ideas about force across the different contexts. She was confused about the force of gravity. It appeared that she knew that there was a force called ‘‘gravity,’’ but her understanding of ‘‘gravity’’ was inchoate, vague, and shifting. As seen from the quotations, she was easily confused in different contexts and she frequently changed her mind while she thought aloud.

Inci provides further examples of multiple and inconsistent responses and reasoning even though she coded as consistent with the error allowance for the *acquired/push–pull* force meaning. In Set 1 (big vs. small stones standing on the ground), Inci assigned a force from the air to both stones and concluded that ‘‘the small stone is affected more by the force of air so force from the air is greater on the small stone.’’ Based on this response, it is expected that she would always assign a greater force from the air to the small objects. However, in Sets 3 (unstable small vs. unstable big stones standing on a hill), 5 (falling big vs. falling small stones), and 6 (a man trying to move a big stone vs. small stone), she assigned a greater force from the air to the big stone. In

Set 10 (a man throwing a small stone vs. throwing a big stone), she thought that forces from the air were equal on the different sized stones because both of them were in the air. Inci expressed another explanation regarding force of air for the pushing contexts in Set 7 (a man trying to move a big stone vs. small stone but he cannot move either) and Set 8 (a man trying to move a big stone vs. a child trying to move a big stone but they both fail). In these sets, she thought that all forces, including the force from the air, were equal on the stones because the stones did not move.

After the first three question sets, Inci began assigning a gravitational force to the stones. In Set 4 (falling big vs. standing big stones), she said that gravity was greater on the standing stone because gravity applied more force to keep the standing stone on the ground compared to the falling stone. In Set 9 (a man throwing a stone vs. a similar stone standing on the ground), she did not assign a gravitational force to the thrown stone in the first context because she thought there would not be a gravitational force on the stones which are in the air. There would therefore be a greater force on the standing stone because of gravitational force on the standing stone. On the other hand, she did assign a gravitational force for the falling contexts in Set 4 (falling big vs. standing big stones) and she assigned an equal gravitational force to the thrown stones in Set 10 (a man throwing a small stone vs. throwing a big stone).

*Coherence in high school students' interpretations of force*. Can we conclude that high school students in this study have coherent understandings of force just because they coded as consistent with the error allowance for the *gravity and other* and *acquired/push–pull* force meanings according to the I&V's and DG&E's schemes? These students have already taken a number of physics and/or physics-related science courses. Most physics courses introduce the concept of force and motion within the *push–pull* context. Gravity is introduced later. These two types of forces are also extensively represented in force problems. Therefore, it is expected that high school students would mention gravity and forces from human agents when someone presents them with those contexts. However, perhaps because of the complexity of the domain, high school students also expressed many non-normative and context-dependent *gravity and other* and *acquired/push–pull* force ideas as discussed above. The high school students clearly knew that there were forces on the objects from gravity and air but they did not know how those forces work. Their interpretations changed from context to context as discussed previously. Ultimately, the softnesses of the coding schemes regarding the *acquired/push–pull* and *gravity and other* force meanings appear likely to have led to the over-classification of the students into those categories as was the case with the middle school students. The softnesses of the coding criteria along with the error allowance therefore seem problematic in terms of supporting claims of coherence.

*Summary: High School Students.* In summary, high school students demonstrated more consistent responses across question sets compared to the other group of students. They usually assigned a force from the ground and a force from the air to the question sets. They also talked about gravitational force. It seemed in many cases, however, that the students often maintained fragmented inchoate understandings of these concepts.

## Synthesis and Implications Across Grade Levels

We have thus far presented and discussed the results for each age group in terms of consistency of force meanings. We now discuss the implications of the results in terms of (1) whether differences in the coding schemes could have been responsible for the differences in the findings in the original studies, (2) whether the error allowance criterion for consistency supports claims of coherence, (3) how the structuring of the question sets may affect students' explanations, (4) how softness in the coding schemes affects claims of coherence, (5) how the Turkish students in this study parallel and diverge from the Greek and U.S. students in the original studies in terms of force meanings, and, finally, (6) what the previous five implications suggest about the overarching debate about student knowledge structure coherence and the original studies.

### *Implication #1: Similar Categorizations of Students Without the Error Allowance*

DG&E adapted their scheme from I&V's scheme, as outlined in the Methods Section of the current study, but there was concern that mechanical differences between the two schemes might have resulted in the conflicting findings between the two studies. While DG&E attempted to create a coding scheme that coded students as ‘‘consistent’’ more liberally than the original I&V qualitative scheme, it is entirely possible that the differences in the schemes contributed to the substantial differences in findings. If the two coding schemes coded students substantially differently in terms of consistency, the current study might conjecture that coding scheme differences contributed significantly to the differences in findings of the original studies. If the two coding schemes code students similarly, however, comparing the levels of consistency demonstrated by the Turkish students with the levels of consistency demonstrated by the Greek and U.S. students in the original studies will help us determine whether the findings for the Turkish students align more closely with DG&E’s or I&V’s findings in terms of consistency (and thus align more closely with the theoretical explanations proposed in those original studies).

The first implication of the results of the current study is that I&V’s and DG&E’s analytic methods categorize individual students very similarly in terms of consistency and force meaning when not allowing the error allowance (Table 12). In the current study, the two schemes coded individual students with the same classification in terms of consistency and force meaning for 97% (31/32) of the students when the error allowance was not allowed. This rate decreased to 63% (20/32) students with the error allowance. Considering that the agreement rate is 97% when not applying the allowance, it seems unlikely the large differences in findings between I&V and DG&E can be attributed solely to differences in coding schemes, particularly because I&V focused solely on the fully consistent criterion.

### *Implication #2: The Error Allowance Seems Inappropriate for Claims of Broad Consistency*

I&V did not need the error allowance because 88.6% of their students were consistent without it. DG&E added the allowance to their study to account for potential differences in how their scheme might diverge from I&V’s. As discussed in the preceding section, however, the results of the current study suggest that the two schemes code students very similarly when the fully consistent criterion is applied. This is helpful methodologically, because the error allowance does not seem to support claims of coherence. The error allowance seems questionable not only because it causes divergent codings between I&V’s and DG&E’s schemes but more importantly because some key contextual comparisons are only covered in two question sets, allowing them to be ignored. Essentially, while there are ten question sets in one sense, there are only three categories of sets in another sense. These categories focus on the nature of what is varied in the set. The first category focuses on varying the size of the stone in different contexts (Sets 1, 3, 5, 6, 7, and 10). The second category varies the motion state of the similar sized stones (Sets 4 and 9). The third category varies stability factors for similar sized stones (Sets 2 and 8). Because the second and third categories only involve two question sets each, allowing two sets to be dropped from the analysis with the error allowance can easily drop an entire category of variation from the analysis. We therefore suggest that the error allowance is not compatible with claims of broad coherence.

Table 12
*Number of students having at least one consistent force meaning by grade level*

| | Students consistent without 20% allowance | | | | Students consistent with error allowance | | | |
|---|---|---|---|---|---|---|---|---|
| | I&V consistent | DG&E consistent | Schemes disagree* | Schemes partially agree** | I&V consistent | DG&E consistent | Schemes disagree* | Schemes partially agree** |
| Pre | 2/8 | 2/8 | 0/8 | 0/8 | 7/8 | 5/8 | 2/8 | 0/8 |
| Elementary | 1/8 | 1/8 | 0/8 | 0/8 | 4/8 | 5/8 | 4/8 | 1/8 |
| Middle | 5/8 | 4/8 | 1/8 | 1/8 | 7/8 | 7/8 | 4/8 | 2/8 |
| High | 4/8 | 4/8 | 0/8 | 0/8 | 8/8 | 6/8 | 2/8 | 0/8 |
| Total | 12/32 | 11/32 | 1/32 | 1/32 | 26/32 | 23/32 | 12/32 | 3/32 |

*Disagreements include (1) individual students coded as consistent by one scheme and not by the other scheme (complete disagreement) and (2) individual students coded as consistent for different force meanings by the two schemes (complete disagreement).
**Partial Agreements include individual students coded by both schemes as consistent for at least one of the same force meanings but also coded as consistent for at least one additional force meaning by one scheme but not the other scheme.

*Implication #3: The Format of Sets may be Inappropriate for Claims of Broad Consistency*

Another methodological issue involves the structure of the question sets. Because DG&E's structuring of each question set asks students to compare two scenarios that only differ in terms of a single variable, it seems possible that this format leads students to attribute the existence and meanings of force in each question set to the specific focal variable in that question set. The first three sets focus on stone size, and may therefore lead students to reason that the size of the stone is what matters throughout the interview because of this initial priming. This might potentially inflate the number of *internal* force meanings expressed across the interviews. This is certainly not a fatal flaw, but it is worthy of consideration in the development of future instruments, especially because stone size defines six of the ten question sets.

*Implication #4: Softness of the Coding Schemes With Regard to Gravity and Air*

Although more than half of all high school and middle school students' interpretations of force coded consistent in terms of *gravity and other* and/or *acquired/push/pull* force meanings with the error allowance, these consistencies were frequently due to the softness of the coding schemes with regard to gravity and air. Essentially, the students' specific interpretations of force from the air and gravity are not considered in enough detail. DG&E discussed this in their study, but appropriately chose to follow I&V's lead in their replication of I&V's methods. Although we described what students said about ''gravitational force'' and ''force from the air,'' future research needs to account more carefully for the variety of force meanings students discuss related to the ground, gravity, and the air.

*Implication #5: The Turkish Students' Different Misconceptions About Force*

In this study, Turkish students came up with many force ideas that were not captured by I&V's or DG&E's schemes. As discussed above, many of the new ideas were related to gravity, which DG&E expressed concerns about as a category because their analysis suggested that the category appeared to include so many different ideas. Many new ideas were also expressed that related to air and other sources. In this section, we outline some of these meanings.

Students expressed a wide range of ideas related to gravity. True to I&V's scheme, we classified responses into the *gravity and other* force meaning as long as the students explicitly mentioned gravity or the Earth's attraction in their responses. Students, however, expressed some unusual *gravity and other* force meanings. First, some students assigned greater gravitational force to the small stone compared to the big stone or greater gravitational force to the standing stone compared to the falling stone. Second, a few students described a ''supportive force'' from the ground or hill. Third, one student assigned forces from the moon and the other planets. Finally, some students discussed a force meaning related to surface area between the objects and ground or hill as well as a frictional force from the ground.

Similar to the variety of meanings that we observed with regard to gravity, students also expressed a variety of air-related force ideas. Almost all of our unusual air-related explanations were observed in middle and high school students. To be consistent with the categorization, we classified all air-related force interpretations into the *acquired/push–pull* force meaning as I&V did. The Turkish students, though, expressed several air-related force interpretations that seem distinct. First, a few students described an air pressure below the stone that balances it. Second, several students thought that air pressure was related the sizes/weights of the objects. Third, some students assigned air pressure to the standing stone only in comparison to the falling stone. Finally, one middle school student thought that not all objects are under the same air pressure in all contexts. In his understanding, total air pressure is partitioned by the objects, for example, by the man and the standing stone in Set 4 (falling big vs. standing big stones). As he explains, ''The man may affect the stone because of air pressure; the man can change the air pressure. For example, think about a room; there is only one object inside and if we measure the pressure in this room and compare the result with the pressure in a room with full of stuff, we find different pressure values. In this case, the man and this stone share the air pressure; if the stone stands alone, air pressure can be higher.''

In addition to the new ideas related to gravity and air, the Turkish students also discussed a few other unusual ideas. First, a few students assigned a force to falling objects because they must have been pushed. Second, some students did not express an overt explanation of why there was a force. They simply said that

"that is the way things are." Third, one student assigned a force of living things only. Finally, some students thought that falling and unstable stones could be broken and lose their force. These students' force meanings imply an *internal* force that the stones possess, but these students also thought that the stones lose their force when they are broken into parts. For example, Faruk, a pre-school student, clearly demonstrated this force meaning in Set 2 (stable vs. unstable stones standing at the top of hills). After he expressed the unstable stone has an internal force due to its size, he stated that ". . . if this stone rolls down, it will break into parts and then its force disappears . . . now it has force but later this force disappears."

In summary, the Turkish students discussed a number of ideas related to force not discussed in the prior studies. Of these new force meanings, our subsequent analyses suggest that only the "force is a property of living things" meaning was applied by a student consistently across all question sets. Therefore, these new meanings do not suggest that the overall levels of consistency would have been much higher with additional formal coding categories. These new meanings do suggest, however, that the parsimony of I&V's scheme is misleading. Students apparently express a broader range of ideas in terms of force. These results also therefore suggest potential differences between Turkish, American, and Greek samples in terms of meaning of force. These differences might be due to either the small size of the samples in each country or these differences might result from educational, language, or cultural differences between the countries.

### *Overarching Implication: The Results Support Elemental Perspectives*

Synthesizing the discussion from the five preceding sections suggests a sixth overarching implication—The results most closely support elemental perspectives on knowledge structure coherence. Differences in coding schemes do not appear to be the cause of the differences in findings of the original studies. The findings of the current study according to both schemes suggest less fragmentation for the Turkish students' than documented by DG&E for the U.S. students and greater fragmentation than documented for the Greek students by I&V. These levels are closer, however, to those reported by DG&E overall. I&V found that 89% of their students were fully consistent. DG&E, however, found that only 17% of their students were fully consistent and 43% were consistent with the error allowance. Our study found that approximately 36% of our students were fully consistent and 73% were consistent with the error allowance (38% and 81% using I&V's scheme and 34% and 64% using DG&E's scheme).

The relationship between consistency of overall classifications of individual students and the coherence of those students' understandings, however, is brought into question by this study as discussed in the preceding sections. The coding categories are so broad and soft that many students expressed multiple different meanings across contexts that still fell within the broad categories. This issue was discussed by DG&E. In addition, the error allowance criterion for consistency seems not to support claims of coherence because entire contexts can be ignored in the question sets. Therefore, while there are clearly systematicities in students explanations and understandings of force, strong claims regarding the theory-like structure of students' understanding of force do not seem supported.

## Limitations and Next Steps

Our findings therefore suggest that Turkish students' understandings of force are fragmented in nature. However, we must still be careful in our interpretation of the nature of Turkish students' knowledge structure. Although the original and current studies applied the same specific codes to the data gathered from their respective samples, we should not ignore the possibility that the differences between the findings may have resulted from different interpretations or applications of the schemes. For example, although we scored students' responses based on the specific coding rules documented in the schemes, we still had to interpret how to categorize several of the responses given by students when the categories and the coding for those categories were underspecified or did not align well with a student's comments. Therefore, the differences between the interpretations or applications of the coding schemes may have contributed to the differences in findings. Future research that involved researchers or representatives from each of the original studies working on a shared data set would prove valuable in clarifying these issues.

That said, however, our findings diverge from I&V's findings in important ways. I&V's results indicated that "most of the children made use of a small number of relatively well-defined and internally consistent interpretations of force" (Ioannides & Vosniadou, 2002, p.5). We found that the Turkish students as a group employed a larger number of force meanings in our study. This assertion was confirmed both within and across age groups. Students in our study assigned more than one force meaning to individual contexts. Some meanings were present in one context but absent in another context. In addition, in many cases, students had contradictory causal explanations that were not internally consistent. Overall, our findings seem more in line with the knowledge-in-pieces perspective, suggested by the evidence that the Turkish students in this study used fragmented, varied, and context-dependent force meanings. The findings do suggest, however, the possibility that differences in student populations (whether due to differences in language, culture, educational system, or some other factor) may contribute to differences in force meanings expressed by students and in levels of knowledge structure coherence.

Although this study does not directly examine how cultural and linguistic differences influence students' ideas about force, it does support the possibility that some differences in students' thinking about force and motion might potentially result from cultural or semantic differences. In Turkish, as in Greek, the word for ''force'' is the same word for ''strength'' and ''power;'' while in English the word ''force'' is often used interchangeably with the term ''compel,'' among many other meanings. This could potentially lead the young Turkish and Greek students to be more consistent and the U.S. students to be more fragmented. Research focusing on the effects of culture and language on science learning supports this idea. A number of researchers have documented how cultural and language differences can influence how students think about natural phenomena (e.g., Cobern, 1993; Allen & Crawley, 1998; Hatano et al., 1993; Hewson & Hamlyn, 1984; Jegede & Okebukola, 1991a,b; Jones, Carter, & Rua, 2000; Kawagley, Norris-Tull, & Norris-Tull, 1998; Lubben, Netshisaulu, & Campbell, 1999). Therefore, while these results support elemental perspectives overall, more research is needed to examine how differences in cultural and linguistic contexts influence individual's conceptualizations of force and motion.

Several other factors such as educational system differences, formal/informal knowledge differences, and learning experiences in different countries might also account for the differences in students' knowledge structures. The degree and manner in which students are exposed to the concepts of mechanics in physical science, for example, potentially influence their understandings and explanations of gravitational force. In turn, this knowledge may influence their overall interpretation of force. Lastly, the relatively small sample sizes in each of the studies could possibly present different apparent patterns within each sample. Clearly more research is needed to examine these issues that may or may not influence students understanding of force and motion.

Lastly, while the results of this study support elemental claims over strong claims of naïve theory-like knowledge structures, the findings of the current study clearly illustrate significant systematicities in students' understandings. Further work investigating this continuum of coherence and the mechanisms underlying the apparent systematicities should focus on the nature of these systematicities. In our future work, we plan to continue the exploration of this continuum as well as the possible role of cultural or semantic differences on students' knowledge structures.

## Conclusions, Next Steps, and Final Thoughts

This study investigates the conflicting results obtained by Ioannides and Vosniadou's (2002) and diSessa, Gillespie, and Esterly (2004) studies about students' knowledge structure coherence. The results suggest that those differences were not a function of analytic methods. Both schemes coded individual students in the current study with the same overall classification in terms of the full consistency criterion for 97% of the students. Thus, there are differences in how the two schemes code students, but these differences could not account by themselves for the vast differences in the findings of I&V's and DG&E's studies, particularly if we focus on the full consistency criterion.

According to both coding schemes, the findings of the current study suggest less fragmentation for the Turkish students' than documented by DG&E for the U.S. students and greater fragmentation than documented for the Greek students by I&V. These levels are closer, however, to those reported by DG&E. I&V found that 89% of their students were fully consistent. DG&E, however, found that only 17% of their students were fully consistent and 43% were consistent with the error allowance. Our study found that ~36% of our students were fully consistent and 73% were consistent with the error allowance (38% and 81% using I&V's scheme and 34% and 64% using DG&E's scheme).

Further supporting elemental claims, the relationship between consistency of overall classifications of individual students and the coherence of those students' understandings is brought into question by closer analysis of students' interviews. The coding categories are so broad and soft that many students expressed multiple different meanings across contexts that still fell within the broad categories. This issue was discussed by DG&E. In addition, the error allowance criterion for consistency seems not to support claims of coherence because entire contexts can be ignored in the question sets. Therefore, while there are clearly systematicities in students explanations and understandings of force, strong claims regarding the theory-like structure of students' understanding of force do not seem supported.

Lastly, while the results of the current study strongly support elemental perspectives on knowledge structure coherence overall, the results of this study suggest that the Turkish students in this study thought about force in some ways differently than the Greek students in I&V's study and the U.S. students in DG&E's study in terms of new context-dependent force-related ideas. Future research should further investigate the nature of the observed systamaticities and students' understandings as a function of culture, language and educational system.

## References

Allen, N.J., & Crawley, F.E. (1998). Voices from the bridge: Worldview conflicts of Kickapoo Students of science. Journal of Research in Science Teaching, 35, 111–132.

Brown, D. (1995). Theories in pieces? The nature of students' conceptions and current issues in science education. Paper presented at the annual meeting of the National Association for Research in Science Teaching, San Francisco, CA.

Carey, S. (1985). Conceptual change in childhood. Cambridge, MA: MIT Press.

Carey, S. (1999). Sources of conceptual change. In E.K. Scholnick, K. Nelson, & P. Miller (Eds.), Conceptual development: Piaget's legacy (pp. 293–326). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Clark, D.B. (2006). Longitudinal conceptual change in students' understanding of thermal equilibrium: An examination of the process of conceptual restructuring. Cognition and Instruction, 24(4), 467–563.

Clark, D.B., D'Angelo, C., & Schleigh, S. (in preparation). Multinational comparison of students' knowledge structure coherence.

Clark, D.B., Menekse, M., Ozdemir, G., & D'Angelo, C. (in preparation). Comparison of knowledge structure coherence and force meanings across sites in Turkey.

Cobern, W.W. (1993). College students' conceptualizations of nature: An interpretive world view analysis. Journal of Research in Science Teaching, 30, 935–951.

diSessa, A.A. (1988). Knowledge in pieces. In G. Forman & P.B. Pufall (Eds.), Constructivism in the computer age (pp. 49–70). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

diSessa, A.A. (1993). Toward an epistemology of physics. Cognition and Instruction, 10(2&3):105–225.

diSessa, A.A., Gillespie, N., & Esterly, J. (2004). Coherence versus fragmentation in the development of the concept of force. Cognitive Science, 28, 843–900.

Harrison, A.G., Grayson, D.J., & Treagust, D.F. (1999). Investigating a grade 11 student's evolving conceptions of heat and temperature. Journal of Research in Science Teaching, 36(1), 55–87.

Hatano, G., Siegler, R.S., Richards, D.D., Inagaki, K., Stavy, R., & Wax, N. (1993). The development of biological knowledge: A multi-national study. Cognitive Development, 8, 47–62.

Hewson, M.G., & Hamlyn, J. (1984). The influence of intellectual environment on conceptions of heat. European Journal of Science Education, 6, 245–262.

Hunt, E., & Minstrell, J. (1994). A cognitive approach to the teaching of physics. In K. McGilly (Ed.), Classroom lessons: Integrating cognitive theory and classroom practice (pp. 51–74). Cambridge, MA: MIT Press.

Ioannides, C., & Vosniadou, S. (2002). The changing meaning of force. Cognitive Science Quarterly, 2, 5–61.

Jegede, O.J., & Okebukola, P.A. (1991a). The effect of instruction on socio-cultural beliefs hindering the learning of science. Journal of Research in Science Teaching, 28, 275–285.

Jegede, O.J., & Okebukola, P.A. (1991b). The relationship between African traditional cosmology and students' acquisitions of science process skill. International Journal of Science Education, 13, 37–47.

Jones, M.G., Carter, G., & Rua, M.J. (2000). Exploring the development of conceptual ecologies: Communities of concepts related to convection and heat. Journal of Research in Science Teaching, 37, 139–159.

Kawagley, A.O., Norris-Tull, D., & Norris-Tull, R.A. (1998). The indigenous worldview of Yupiaq culture: Its scientific nature and relevance to the practice and teaching of science. Journal of Research in Science Teaching, 35(2), 133–144.

Kuhn, T. (1962). The structure of scientific revolution. Chicago: The University of Chicago Press.

Linn, M.C., Eylon, B., & Davis, E.A. (2004). The knowledge integration perspective on learning. In M.C. Linn, E.A. Davis, & P. Bell (Eds.), Internet environments for science education. Mahwah, NJ: Lawrence Erlbaum Associates.

Lubben, F., Netshisaulu, T., & Campbell, B. (1999). Students' use of cultural metaphors and their scientific understandings related to heating. Science Education, 83, 761–774.

Schleigh, S. (2008). Format and sex in assessing the knowledge structure coherence of middle school student' understanding of the concept of force. Unpublished doctoral dissertation. Arizona State University.

Vosniadou, S. (2002). On the nature of naïve physics. In M. Limon & L. Mason (Eds.), Reconsidering conceptual change: Issues in theory and practice (pp. 61–76). Dordrecht: Kluwer Academic Publishers.

Wellman, H.M., & Gelman, S. (1992). Cognitive development: Foundational theories of core domains. Annual Review of Psychology, 43, 337–375.

Wiser, M., & Carey, S. (1983). When heat and temperature were one. In D. Gentner & A. Stevens (Eds.), Mental models (pp. 267–298). Hillsdale, NJ: Lawrence Erlbaum Accociates, Inc.

## Appendix: Coding of Students' Consistent Force Meanings With and Without Error Allowance Across Grade Levels

| | Int | Int/Move | Int/Acq | Acq | Acq/P-P | Push–pull | Gravity | F to living |
|---|---|---|---|---|---|---|---|---|
| Pre-school | | | | | | | | |
| Sena | ID | | | | | | | |
| Nazlı | | | | | | | | |
| Ali | | | I*D* | | | | | |
| Hasan | | I*D* | | | | | | |
| Aye | | | I*D* | | | | | |
| Meral | | | | | | | | I* |
| Faruk | | I* | | | | | | |
| Can | ID | | | | | | | |
| Elementary school | | | | | | | | |
| Kenan | | D* | | | | | | |
| Mert | | | | D* | | | | |
| Rana | I*D* | | | | | | | |
| Sezen | | I* | | | | | | |
| Nur | | | | | | | ID | |
| Senay | | | | | | | | |
| Engin | | | | | | | | |
| Utku | | | | I*D* | I* | | | |
| Middle school | | | | | | | | |
| Emine | | | | | I* | | ID | |
| Pelin | | | | | | | | |
| Veli | I | | | | I | | ID | |
| Dilek | | | | ID* | I*D* | | | |
| Bilal | | | | | | | ID | |

(*Continued*)

Appendix: (*Continued*)

| | Int | Int/Move | Int/Acq | Acq | Acq/P-P | Push–pull | Gravity | F to living |
|---|---|---|---|---|---|---|---|---|
| Sibel | | | | | I*D* | I* | | |
| Emre | | | | | I* | | ID | |
| Erkan | | | | | | | I*D* | |
| High school | | | | | | | | |
| Ece | | | | | | | I*D* | |
| Gül | | | | | I* | | | |
| Inci | | | | | I* | | | |
| Aysel | | | | | | | ID | |
| Remzi | | | | | | | ID | |
| Sami | | | | | | | I*D* | |
| Kamil | | | | | | | ID | |
| Sakir | | | | | | | ID | |

I, I&V's scheme; I*, I&V's scheme with error allowance.

D, DG&E's scheme; D*, DG&E's scheme with error allowance.