

International Analysis of Students' Knowledge Structure Coherence

Douglas B. Clark, Cynthia M. D'Angelo, Sharon Schleigh, Muhsin Menekse
Arizona State University

Abstract: This international comparison investigates students' knowledge structure coherence in physics across five countries. In particular, this study investigates two possible hypotheses explaining the conflicting results obtained by Ioannides and Vosniadou (2002) and diSessa, Gillespie, and Esterly (2004) about students' understandings of force in Greece and the United States. Ioannides and Vosniadou's study in Greece demonstrated broad consistency in students' understandings of force. diSessa and colleague's quasi-replication in the U.S. demonstrated conflicting results supporting more elemental perspectives. One hypothesis focuses on differences in analytic methods. The other hypothesis focuses on semantic, cultural, or educational differences between the students in the two studies. The findings of this study suggest that differences in analytic methods do result in coding differences but that these relatively small differences would not account for the significant differences between the studies. This study, however, demonstrates significant differences for force meanings and knowledge structure coherence across the countries that might explain a larger percentage of the differences in findings between the studies.

Introduction

This study investigates students' understanding of the scientific concept of "force" in Turkey, Mexico, the Philippines, the U.S., and China. More specifically, this study applies the analytic schemes of both Ioannides and Vosniadou (2002) and diSessa, Gillespie, and Esterly (2004) to 187 students from the five countries in order to map in greater detail the specific connections, regularities, and irregularities demonstrated within the students' understanding of force. The five countries were chosen to allow comparisons across a range of language families and cultures to clarify possible differences in students' understanding of force. The study contributes to the resolution of a central controversy among researchers of conceptual change regarding the structure and coherence of students' science knowledge and clarifies the role of methodological and semantic/cultural differences in the findings of researchers on opposing sides of the controversy.

The Controversy

A core controversy within research on conceptual change is concerned with the structure of students' knowledge. 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)?

The statements above are simplifications of the actual perspectives, which are considerably more nuanced as a result of substantial research and ongoing debate among their respective proponents. Proponents of theory-like positions, for example, do not argue that students' knowledge is "theory-like" in the same fashion as the knowledge of scientists (e.g., including meta-conceptual awareness or availability to hypothesis testing). These proponents do argue, however, for an overarching hierarchical conceptual structure with theory-like properties that constrains a student's interpretation of subordinate models and ideas. Similarly, the elemental perspectives should not be incorrectly caricatured 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. These models implicate radically different pathways for curricular design to help students reorganize their understandings.

While the researchers in each camp also vary in terms of other important issues (e.g., conceptual grain-size, ages of students, methods, and scientific content areas) this debate remains highly visible and contested. Comparing findings between researchers in this debate has been difficult, however, because of the differences in research methodologies and contexts. Recently, two groups of researchers have begun to address these issues around the concept of "force" in science. Ioannides and Vosniadou (2002) conducted an initial study in Greece with four age groups showing that students express consistent answers about force across multiple contexts. diSessa, Gillespie, and Esterly (2004) conducted a quasi-replication in the United States showing that students' explanations lack ontological coherence and vary significantly across contexts.

Significant findings from Ioannides and Vosniadou (2002)

Ioannides and Vosniadou (henceforth frequently referred to as I&V) investigated the meaning of force

and its development among children across grade levels in Greece. In their standardized sets of questions, I&V showed children pictures with simple stick models and asked the students about forces on the objects in the pictures. I&V also asked comparison questions to further explore students' interpretations of force. I&V found that 88.6% of their subjects' responses could be categorized into seven internally consistent "meanings" of force. This included 87% of pre-k students, 80% of elementary students, 87% of middle school students, and 100% of 9th grade students. For example, a student who has a consistent *internal* force meaning believes that force is related to an objects' size or weight. According to I&V's findings, this student would make predictions and gives explanations consistent with this meaning across the question sets regardless of context.

Significant findings from diSessa, Gillespie, and Esterly (2004)

diSessa, Gillespie, and Esterly (henceforth frequently referred to as DG&E) performed a study among American students in the United States by quasi-replicating a condensed version of I&V's (2002) study and found that their students' meaning of force did not demonstrate the same consistency as reported by I&V for the Greek students. More specifically, DG&E found that only 16.6% of their 30 students were 100% consistent for a meaning. DG&E then broadened their criterion for consistency with the error allowance (which allowed a student to be categorized as consistent for a meaning if that student was coded for that meaning for at least 8 of the 10 question sets rather than 10 out of 10). Using this criterion, 10 of the 30 (33.3%) students could be counted as consistent, but 9 of these 10 were consistent for the *gravity and other* meaning.

Possible Explanations for the Differences in Findings

Two hypotheses have been proposed as the most likely causes for the contradictory findings between I&V's and DG&E's studies. The first possibility is that I&V's and DG&E's analytic schemes resulted in differential coding of subjects and thus resulted in different findings. The second possibility focuses on differences between the sample populations in terms of schooling, cultures, and/or languages. In the Greek language, the word for force, "dynamis," also means strength or power in everyday speech. Potentially, this might have resulted in higher levels of coherence in Greek students' explanations across interview contexts. While semantic and cultural differences have been shown to impact students' thinking about specific science concepts (Aikenhead & Jegede, 1999; Costa, 1995; George, 1999; Inagaki, 2002; Lubben, Netshisaulu, & Campbell, 1999), differences in methods are also critical candidates for further examination. Clearly even slight differences in analytic methods can profoundly impact interpretations (Burkhardt & Schoenfeld, 2003; Nisbett & Ross, 1980; Stigler, Gallimore, & Hiebert, 2000; van de Vijver & Leung, 1997).

Purpose of the current study

The current study investigates these conflicting results and a subset of the surrounding questions. The three interrelated questions at the heart of this study investigate: (1) whether the coding schemes characterize individual students similarly, (2) whether students are consistent in their meanings of force across question sets, and (3) whether students vary in meanings or consistency across countries. By addressing these questions, this study contributes to our understanding of students' knowledge structure coherence, students' conceptual change processes, and curricular approaches to support these processes.

Methods

This study applies DG&E's (2004) and I&V's (2002) coding methodologies across interviews using the same question sets in Turkey, the Philippines, the U.S., China, and Mexico.





<i>Drawing A</i>	<i>Question A</i>	<i>Drawing B</i>	<i>Question B</i>	<i>Comparison Question</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? Why?"		"This stone is standing on a hill. It is stable. That means it won't easily fall down. Is there a force on the stone? Why?"	"Is the force on this stone (A) the same or different than the force on this stone (B)? Why?"
	"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 (A) the same or different than the force on this stone (B)? Why?"

Figure 1. Question Sets 2 (Stable vs. Unstable Stones) and 10 (Throwing Small Stone vs. Big Stone)

Instrument

Students were asked the 10 sets of replication questions that DG&E (2004) condensed from I&V's (2002) questions. The question sets each include two drawings comparing stones and people of different sizes in different configurations to explore the contexts in which the participants would ascribe forces and how they would describe those forces. Two sample question sets are outlined in Figure 1. The contexts of all 10 sets of questions are outlined below:

1. Big vs. Small Stones Standing on the Ground
2. Unstable vs. Stable Similar Stones Standing on a Hill
3. Unstable Small vs. Unstable Big Stones Standing on a Hill
4. Falling Big vs. Standing Big Stones
5. Falling Big vs. Falling Small Stones
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
8. A Man Trying to Move a Big Stone vs. a Child Trying to Move a Big Stone but They Both Fail
9. A Man Throwing a Stone vs. a Similar Stone Standing on the Ground
10. A Man Throwing a Small Stone vs. Throwing a Big Stone

Each set consists of three questions, two simple questions and one comparison question. In each set, the typical opening question asked students, "Is there a force on this stone? Why?" After asking the same questions for the second drawing in each set, students were asked, "Is the force on this stone (in the first picture) the same or different than the force on this stone (in the second picture)? Why?" The comparison question provided more information in terms of relative strengths and contextual-related differences.

Subjects and Procedures

This study compared 187 students (37 students from the U.S., 32 students from Turkey, 39 students from Mexico, 40 students from mainland China, and 39 students from the Philippines). As with I&V's and DG&E's studies, the current study analyzed four different age groups of students including pre-k, elementary school, middle school, and high school. Approximately 9 students were interviewed at each age group in each country. The mean ages were approximately 5, 10, 13, and 16 years for each age group, respectively. Students in each country were selected as socioeconomically representative of middle class students as defined for their country. All students were interviewed individually for 20-25 minutes. Students were asked all questions in one session. All interviews were videotaped.

Coding Schemes

In the analysis, each student's responses were examined to determine if the student consistently applied the same meaning of force across the 10 question sets. To address the possibility that the differences in the findings between I&V (2002) and DG&E (2004) resulted from differences in the two coding schemes, this study separately applied both schemes to each student. I&V's scheme involves first distilling a student's response down to a basic response category and then mapping that response category with a rubric onto all possible force meaning matches for each question set. DG&E felt that they could not reliably distill students' explanations down to I&V's response categories. Instead, DG&E adapted the coding scheme to focus on a "coarse quantitative" analysis involving a rubric for each question set that considered (a) which stones had forces on them, (2) which stone had the greater forces, and (3) exemptions precluding specific force meanings if a student's answer mentioned specific terms (such as "gravity"). DG&E attempted to create a coding scheme that would more liberally attach force meanings to question sets so that their coding scheme would be more liberal in assigning overall consistency to a student. Due to space constraints, please see the two original studies for full descriptions of the two coding schemes.

Force Meanings

This study focuses on the seven force meanings identified by I&V and adopted by DG&E:

1. *Internal force.* Students were assigned to this meaning if they indicated that there is a force on all objects or only on big/heavy objects because the objects 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 if they indicated that the force is due to only to the size/weight of the object or if moving objects and objects that are likely to fall have less internal force than stationary objects.
3. *Internal and acquired.* Students were assigned to this meaning if they indicated that there is a force on stationary objects due to size/weight and that these objects acquire an additional force when they are set in motion. I&V included students in this meaning who were ambivalent about unstable objects and interpreted unstable objects as either lacking internal force or being likely to acquire additional force.

4. *Acquired*. Students were assigned to this meaning if they indicated that force is a property of objects that explains motion and potentially acts on other objects. These students answered that there is no force on stationary objects and that the force on moving objects disappears when the object stops moving.
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* meaning for question sets 7 and 8 even if they did not mention the word “gravity” in these sets.

Analysis, Inter-Rater Reliability, and Determination of Consistency

Every interview in the current study was coded individually by two different coders and then any differences were discussed and final codes were agreed upon for each student for each question set. The interrater reliability between the two coders before discussion was over 90%. After coding each interview, each student was assigned best-match force meanings based on how often the student matched each meaning across the ten question sets. A student was considered fully consistent if he or she matched for at least one force meaning across all 10 question sets. A looser criterion (the 20% error allowance) was also applied allowing students to be considered consistent if they matched for at least one of the same meanings on at least 8 of the 10 sets. This error allowance was not included in I&V’s study but was included in DG&E’s study and earlier studies by Vosniadou.

Results

Results for all U.S., Turkish, Mexican, Chinese, and Philippine students are presented here. This study addresses three fundamental questions about (1) whether the coding schemes characterize individual students similarly, (2) whether students are consistent in the meanings of force that they apply across question sets, and (3) whether students vary in terms of meanings or consistency across countries.

Do DG&E’s and I&V’s Schemes Result in Similar Codings of Students?

Our results for all 187 students demonstrate 84.0% level of agreement between the two schemes across the students in terms of the best-match meanings expressed by the individual students. Similarly, the two schemes agree in about 75% of their determinations of whether or not a specific student is consistent for a specific meaning or not. That said, the level of agreement varies by country and by age of the students (see Table 1). There is also a tendency as discussed in the next section for I&V’s scheme to code a slightly higher percentage of students as consistent than DG&E’s scheme (less than 10% higher). Overall, however, our results suggest that while I&V’s and DG&E’s schemes result in some differences in the coding of individual students, the differences in codings between the two coding schemes are not sufficient to account for the extreme differences in findings between I&V’s and DG&E’s studies. Overall, the pre-k and high school students had the highest percentages of best-match agreement (93.3% and 93.0%, respectively) and the elementary and middle school students had the lowest (78.7% and 78.8%, respectively).

Table 1: Best-match agreement (percentage) between I&V’s and DG&E’s schemes.

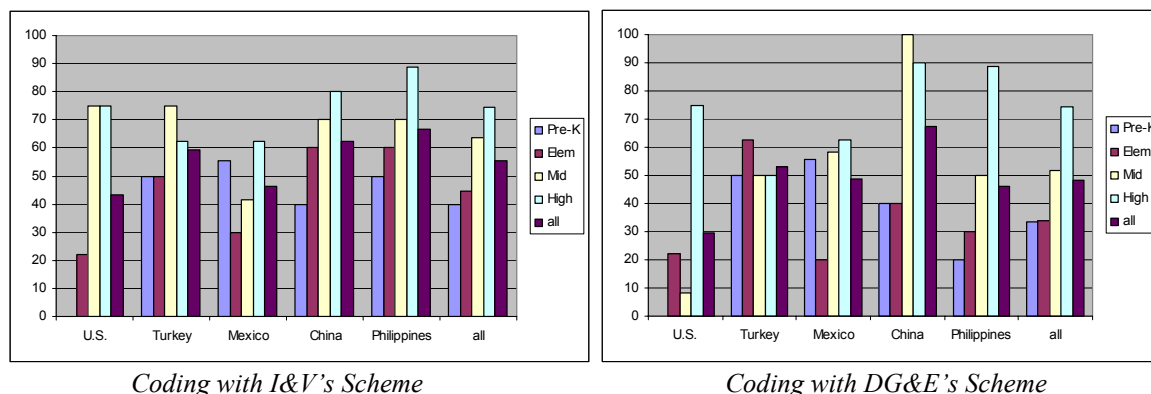
	U.S.	Turkey	Mexico	China	Philippines	All
Pre-K	75.0	100.0	100.0	100.0	90.0	93.3
Elem	88.9	75.0	80.0	100.0	50.0	78.7
Mid	58.3	75.0	91.7	80.0	90.0	78.8
High	100.0	100.0	100.0	90.0	77.8	93.0
All	78.4	87.5	92.3	92.5	76.9	84.0

We can also look at agreement between the schemes in terms of matching individual force meanings (not just best-match) for each student. This essentially involves checking the data “cell by cell” for agreement for every student for every question for every possible force meaning. The two schemes agree 84.8% of the time over all students, with little variation among countries and age groups. Question sets 1, 4, and 9 have the highest level of agreement, 96.0%, 94.0%, and 93.5% respectively. Question sets 3 and 10 have the lowest level of agreement, 75.6% and 69.5% respectively. The rest of the question sets have about 83% agreement. These trends are robust for all five countries, suggesting that the majority of the differences observed in the best-match

force meaning categories can be attributable to differences in age groups and countries rather than coding scheme issues.

How Consistent Are Students In Their Meanings?

In terms of the second question, I&V found that 88.6% of the 105 students in their study were consistent for a single meaning across 100% of the question sets. This included 87% of pre-k students, 80% of elementary students, 87% of middle school students, and 100% of ninth grade students. DG&E found that only 16.6% of their 30 students were 100% consistent for a meaning. DG&E then broadened their criterion for consistency with the 20% error allowance (which allowed students to be categorized as consistent for a meaning if the student was coded for that meaning on at least 8 of the 10 question sets rather than 10 out of 10). Using this error allowance criterion, 10 of the 30 (33.3%) students could be counted as consistent, but 9 of these 10 were consistent for the *gravity and other* meaning.



Figures 2 & 3. Percentage of Students Consistent with Error Allowance Using I&V's and DG&E's schemes

Our own study shows that approximately 36.5% of U.S. students are consistent using the error allowance according to both schemes (43% according to I&V's scheme and 30% according to DG&E's scheme), which is very similar to DG&E's findings for their U.S. students. The breakdown across grades also follows the pattern found in DG&E's study with our results showing 0% of the U.S. pre-k students, 20% of the elementary students, 45% of the middle school students (I&V's and DG&E's schemes diverge significantly on this group), and 75% of the high school students. Interestingly, the Turkish students are much more coherent according to both schemes (53% of Turkish students consistent according to DG&E's scheme and 60% of Turkish students consistent according to I&V's scheme with the error allowance). Turkish students do not show the same pronounced increase in coherence as students get older. For the most part, the Mexican students show fairly constant levels of consistency over age groups similar to the pattern for the Turkish students except that the Mexican elementary cohort has a much lower level of consistency, similar to that of the U.S. students. Overall, 47.5% of the Mexican students were consistent using the error allowance (46% according to I&V's scheme and 49% according to DG&E's scheme). Students in China and the Philippines demonstrate progressively higher levels of consistency as the age of the students increases. High school students in China and the Philippines demonstrate the highest levels of consistency of any students in the study. The charts above summarize these data using both I&V's (see Figure 2) and DG&E's (see Figure 3) schemes.

These results suggest the possibility that a significant portion of the differences in the findings of I&V and DG&E could result from differences in their student samples. The U.S. students in the current study seem quite similar in levels of consistency to the U.S. students in DG&E's study. The students from Turkey, China, and the Philippines, while not demonstrating levels of consistency as high as found in I&V's Greek students, do demonstrate significantly higher levels of consistency than the U.S. students in the current study. The Mexican students demonstrate levels of consistency somewhere between the U.S. and Turkish students. If the differences in the current study between student samples in each country can be attributed to some aspect of their language, culture, or schooling, then it is also possible that similar types of differences could result in the even higher levels of consistency observed in I&V's Greek students.

We created a set of tables in the same manner as DG&E, showing where individual students fall in both number of question sets matched and force meaning category. Space limitations unfortunately preclude their inclusion here, but these tables are available from the first author. Analysis of these tables show a remarkable degree of similarity between I&V's and DG&E's coding schemes for each country. However, comparing the tables among the five countries we see very different patterns. The tables containing the data from Turkey shows a wide spread of students consistent in many different force meanings while the data from the U.S. shows that most of the students that are consistent are in only one or two force meaning categories. The data from the

Philippines shows that consistent students are centered around a couple different force meaning categories, mostly *acquired/push-pull* meanings for the middle and high school students and *acquired*-related meanings for the pre-k and elementary students. The data from China shows that most students are consistent in the *gravity and other* force meaning (many more than in any of the other countries).

What Meanings do Students Express?

In terms of this third question, the general trends in terms of meanings by age group basically follow the findings of I&V and DG&E. Details for each country analyzed so far are in Table 2 (for I&V's scheme) and Table 3 (for DG&E's scheme) where the percentage in bold is the force meaning category that had the most students in that age group. Please note that columns can total above 100% because students that had two meanings "tie" for best-match are counted for both meanings in Tables 2 and 3. In tracking the best-match meanings expressed by individual U.S. students, the pre-k students are generally clustered across the *internal*-related and *push-pull* meanings, the elementary students are mostly clustered in the *acquired*-related meanings, the middle school students are mostly in either *acquired/push-pull* or *gravity and other*, and the high school students are primarily in *gravity and other*.

Table 2: Percentage of students in each force meaning category using I&V's scheme.

	U.S.				Turkey				Mexico				China				Philippines			
	K	E	M	H	K	E	M	H	K	E	M	H	K	E	M	H	K	E	M	H
Internal	-	-	8	-	38	-	-	-	67	10	-	-	10	-	-	-	-	-	-	-
Int/Mov	50	22	8	-	25	38	-	-	-	10	17	-	30	10	10	10	-	-	10	-
Int/Acq	25	22	25	-	38	-	-	13	11	10	25	-	30	-	10	-	10	10	20	11
Acquired	-	22	-	-	13	50	13	-	22	30	17	-	10	-	-	-	60	70	-	-
Acq/P-P	25	44	58	13	-	13	38	50	-	-	42	63	30	50	10	-	30	-	60	22
Push-Pull	25	11	-	-	13	-	-	-	22	20	-	-	30	40	-	-	30	10	-	-
Gravity	13	11	50	88	-	13	63	38	-	30	25	50	-	10	100	90	-	10	20	67

Table 3: Percentage of students in each force meaning category using DG&E's scheme.

	U.S.				Turkey				Mexico				China				Philippines			
	K	E	M	H	K	E	M	H	K	E	M	H	K	E	M	H	K	E	M	H
Internal	13	-	-	-	63	-	25	-	67	20	8	-	10	10	20	20	-	-	-	-
Int/Mov	38	22	-	-	25	25	-	-	-	10	8	-	20	10	-	-	-	-	10	-
Int/Acq	-	-	25	-	38	-	25	38	11	20	25	-	10	10	-	10	10	-	10	-
Acquired	13	33	8	13	-	50	25	-	11	20	42	-	20	20	-	-	50	30	20	-
Acq/P-P	25	56	50	13	-	25	25	50	11	40	50	63	30	50	30	10	30	60	70	44
Push-Pull	25	11	-	-	13	-	-	-	22	20	-	-	30	40	-	-	30	20	-	-
Gravity	-	11	33	88	-	13	63	50	-	20	17	38	-	10	70	90	-	10	10	56

There is a similar pattern of progression of force meaning categories that we observed in most of the countries. The pre-k students tend to be clustered around the *internal*-related force meanings, the elementary students are mostly in *acquired*-related meanings, and the middle and high school students code in both *acquired/push-pull* and *gravity and other* (with generally most high school students in *gravity and other*). Although each country has different characteristics at each age group, it is interesting to note how this progression happens to some degree for each. It is most evident in the U.S. and Turkish data and least in the Chinese and Philippine data.

One of the interesting things to note in the above tables is how spread out some of the groups are, so much so that no force meaning category is more than 40%. This happens for the pre-k Chinese students, the elementary Mexican students, the pre-k Turkish students (using I&V's scheme only), and the pre-k U.S. students (using DG&E's scheme only). On the other extreme, the Chinese middle and high school students and the U.S. high school students were almost all (more than 85% in the combined score) in one category, the *gravity and other* force meaning category. It is also interesting that not one student in middle or high school had a best-match code of the *push-pull* force meaning category. Also, not surprisingly, only one pre-k student was coded into the *gravity and other* force meaning category and only a few high school students were coded in any of the *internal*-related force meaning categories.

The Turkish and Mexican students have different relationships based on age group. The pre-k and high school students are very similar but the elementary and middle school students are quite different in terms of force meaning categories. For the most part, the Chinese students at all age groups are different from the other countries in terms of force meanings. Even though, for example, for the middle school age group the Turkish students look similar to the Chinese students in terms of the force meaning in which they are coded (*gravity and other*), the percentages vary (63% versus 85%). The Chinese and Philippine students are similar in that none of their age groups has more than 30% of the students in any of the *internal*-related force meaning categories.

If we focus only on students' best-match meanings that meet the error allowance criterion for consistency, we see a slightly different picture. These numbers are not included in the table due to space limitations. No U.S. pre-k students are consistent for any meaning but there are four consistent pre-k Turkish students for a mix of *internal*-related meanings and four pre-k Mexican students consistent in the *internal* meaning. The consistent Chinese pre-k students are in either *internal*-related or *push-pull*-related force meaning categories while the consistent Philippine pre-k students are in either *acquired*-related or *push-pull*-related force meaning categories. There are two or three (depending on the scheme) U.S. elementary students who are consistent for either the *acquired*-related meanings or *gravity and other*, there are four or five (depending on the scheme) Turkish elementary students who are consistent for a range of meanings spanning *internal/movement* to *acquired* to *gravity and other*, and there are two or three (depending on the scheme) Mexican elementary students who are consistent in either the *internal* or *gravity and other* meaning. The Chinese and Philippine consistent elementary students are all in either *acquired*-related or *push-pull*-related force meaning categories. For the middle school cohort, nearly all of the consistent students are coded in either one of the *acquired*-related or the *gravity and other* force meaning categories. However, there are two Chinese middle school students who are consistent in *internal*-related force meanings. Most of the high school students in the U.S., Turkey, and China and about half of the high school students in Mexico and the Philippines that code as consistent do so for *gravity and other* but a few Turkish and half of the Mexican and Philippine high school students are consistent for *acquired*-related meanings.

Summary, Implications, and Conclusions

Two primary hypotheses were proposed to explain the differences in students' knowledge structure coherence between I&V's (2002) Greek students and DG&E's (2004) U.S. students. One hypothesis focuses on differences in analytic methods. The other hypothesis focuses on semantic, cultural, or educational differences between the students. The current study investigated these two hypotheses by analyzing students in five countries using both analytic methods to further clarify the debate over knowledge structure coherence.

The results of this study have implications for educational policy and curriculum design. The two perspectives on knowledge structure coherence differ fundamentally in terms of top-down versus bottom-up instructional approaches for scaffolding conceptual change. For example, should curricula focus on helping students revise their existing ideas and connections (e.g., diSessa, in press) or should curricula focus on instilling new perspectives incommensurate with students' existing interpretations (e.g., Chi, 2005)? Equally important, findings about differences in how students from Turkey, Mexico, the Philippines, and China think about force in comparison to English-monolingual students (who are more frequently studied) can provide insights into developing curricula to better support diverse underserved student populations around the world.

The findings of this study suggest that differences in analytic methods do result in coding differences for some students but that these relatively small differences would not account for the significant differences in findings between the two studies. This study instead suggests differences between student populations as a more likely explanation. More specifically, the findings of the current study demonstrate higher consistency for the students in Turkey, Mexico, China, and the Philippines than documented by DG&E for their U.S. students and lower consistency than documented by I&V for their Greek students. The findings thus support the possibility that substantial differences in students' meanings and knowledge structures for force and motion might potentially result from cultural, semantic, or educational differences between student populations. The results also suggest that while the levels of consistency seen by I&V may not be common, there are certainly important systematicities in students' thinking that need to be explained. Our upcoming research will focus on (a) conducting semantic analyses in two or more of the languages to examine possible relationships between the languages and the students' understandings of force and (b) integrating our own approach for coding the components of students' conceptual ecologies from earlier work (Clark, 2006) with current coordination class efforts in the conceptual change research literature (e.g., diSessa & Sherin, 1998; Wagner, 2006; diSessa & Wagner, 2005; Dufresne, Mestre, Thaden-Koch, Gerace, & Leonard, 2005; Parnafes, in revision; Thaden-Koch, Dufresne, & Mestre, 2006) to examine these systematicities in greater detail.

References

Aikenhead, G., & Jegede, O. (1999). Cross-cultural science education: A cognitive explanation of a cultural phenomenon. *Journal of Research in Science Teaching*, 36, 269-287.

- Burkhardt, H., & Schoenfeld, A. H. (2003). Improving educational research: Toward a more useful, more influential, and better-funded enterprise. *Educational Researcher*, 32(9), 3-14.
- 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.
- Chi, M. T. H. (2005). Commonsense conceptions of emergent processes: Why some misconceptions are robust. *The Journal of the Learning Sciences*, 14(2), 161-199.
- Clark, D. B. (2000). *Scaffolding Knowledge Integration Through Curricular Depth*. Unpublished doctoral dissertation, University of California, Berkeley, CA.
- 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., & Jorde, D. (2004). Helping students revise disruptive experientially-supported ideas about thermodynamics: Computer visualizations and tactile models. *Journal of Research in Science Teaching*, 41(1), 1-23.
- Clark, D. B., & Linn, M. C. (2003). Scaffolding knowledge integration through curricular depth. *Journal of Learning Sciences*, 12(4), 451-494.
- Clark, D. B., Sampson, V., Han, J., Ozdemir, G., Jensen, J., Neakrase, J., et al. (in preparation). Conceptual change and knowledge structure coherence: Arguing both sides of coherence and fragmentation.
- Costa, V. (1995). When science is "Another World": Relationships between worlds of family, friends, school, and, science. *Science Education*, 79(3), 313-333.
- diSessa, A. A. & Sherin, B. (1998). What changes in conceptual change? *International Journal of Science Education*, 20(10), 1155-1191. [Read First]
- diSessa, A. A. (2006). A history of conceptual change research: Threads and fault lines. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences*. Cambridge, UK: Cambridge University Press.
- diSessa, A. A., & Wagner, J. F. (2005). What coordination has to say about transfer. In J. Mestre (ed.), *Transfer of learning from a modern multi-disciplinary perspective* (pp. 121-154). Greenwich, CT: Information Age Publishing.
- diSessa, A. A., Gillespie, N., & Esterly, J. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science*, 28, 843-900.
- Dufresne, R., Mestre, J., Thaden-Koch, T., Gerace, W., & Leonard, W. (2005). Knowledge representation and coordination in the transfer process. In J. Mestre (ed.), *Transfer of learning from a modern multi-disciplinary perspective* (pp. 155-215). Greenwich, CT: Information Age Publishing.
- George, J. (1999). World view analysis of the knowledge in a rural village: Implications for science education. *Culture and Comparative Studies*, 83, 77-95.
- 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.
- Inagaki, K., & Hatano, G. (2002). *Young Children's Thinking about Biological World*. Philadelphia, PA: Psychology Press.
- Ioannides, C., & Vosniadou, S. (2002). The changing meanings of force. *Cognitive Science Quarterly*, 2(1), 5-62.
- 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). Student's use of cultural metaphors and their scientific understandings related to heating. *Science Education*, 83, 761-774.
- Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice Hall.
- Parnafes, O. (In Revision). What does "fast" mean? Understanding the physical world through computational representations. Submitted to *The Journal of the Learning sciences*.
- Stigler, J. W., Gallimore, R., & Hiebert, J. (2000). Using video surveys to compare classrooms and teaching across cultures: Examples and lessons from the TIMSS video studies. *Educational Psychologist*, 35(2), 87-100.
- Thaden-Koch, T., Dufresne, R., & Mestre, J. (2006). Coordination of knowledge in judging animated motion. *Physics Education Research*, 2, 1-11.
- van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage Publications Inc.
- Wagner, J. F. (2006). Transfer in pieces. *Cognition and Instruction*, 24(1), 1-71.
- Wellman, H. M., & Gelman, S. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology*, 43, 337-375.