Comparing American and Chinese Students' Learning Progression on Carbon Cycling in Socio-Ecological Systems

J. CHEN[*], C. W. ANDERSON[†]

ABSTRACT: Previous studies identified a learning progression on the concept of carbon cycling that was typically followed by American students when they progress from elementary to high school. This study examines the validity of this previously identified learning progression for a different group of learners—Chinese students. The results indicate that American and Chinese students share similar learning progression from force-dynamics to scientific model-based reasoning. And there are interesting similarities and differences between American and Chinese students' performances. Whereas American students perform better on items assessing the environmental impact of human behaviours, Chinese students include chemical equations, named forms of energy, and mention the energy conservation principle more commonly than American students. This study suggests that these differences may be due to various aspects of science education between these two countries and have implications for improving science education in each country.

KEY WORDS: environmental science education, international comparison study, Item Response Theory (IRT), learning progression, science education

INTRODUCTION

Carbon cycling has a significant impact on global climate (Heimann & Reichstein, 2008) and the imbalances of carbon cycling processes are widely agreed to be the primary cause of global warming (Houghton, 2005). Excessive human-related emission of carbon dioxide into the atmosphere is a major contributor to global warming. America and China are two countries with high rates of carbon emission (Gregg, Andres, & Marland, 2008). Therefore, it is crucial for American and Chinese citizens to understand carbon cycling so they can make environmental friendly decisions to reduce carbon emission into the atmosphere. In addition, science education in these two countries needs to help students develop sophisticated understanding about carbon cycling. To achieve this, we first need to know American and Chinese students' understanding about carbon cycling and how their understanding progress over time.

---

* Corresponding Author: Educational Testing Service, USA, jchen003@ets.org
† Michigan State University, USA

Given the importance of the "carbon cycling" topic, previous studies (Mohan, Chen, & Anderson, 2009; Jin & Anderson, 2012) investigated students' understanding of the carbon cycle and developed a learning trajectory on the concept of carbon cycling that was followed by American students typically when they progressed from elementary to high school. Learning progressions were used to describe how students' understanding of a topic progressed over time and to classify the progress of students into steps or levels (Corcoran, Mosher, & Rogat, 2009).

In this study, we examine whether this pre-identified learning trajectory is also shared by a different group of learners — Chinese students. Chinese students are educated in another education system, and have a very different culture and language. The differences between American and Chinese students' learning trajectory may result from the differences in various aspects of science education between these two countries, such as curriculum, national standards or the examination systems. The comparison study informs us with ideas of science education from another country and fuels our new thinking on how to help students in both countries reach higher level understanding more efficiently.

Multiple studies have applied item response theory (IRT) based methods to evaluate learning progression frameworks and the assessments that are used to measure students' understanding over time (Adams, Wilson, & Wu, 1997; van der Linden & Hambleton, 1996; Songer, Kelcey, & Gotwals, 2009). In this study, we apply IRT based methods to examine whether Chinese students share a similar learning progression to that typically followed by American students. We also investigate the similarity and differences between American and Chinese students' performances.

The following research questions are put forward:

a) Are learning progression levels developed for U.S. students valid for Chinese Students?

b) How do American and Chinese Students' performance compare based on learning progression of carbon cycling in socio-ecological systems?

LITERATURE REVIEW

Learning Progression in Science Education

Learning progressions have become popular within the science education community because of their potential to build a bridge between research on how people learn and the methods for teaching and assessing science (Salinas, 2009; Corcoran et al., 2009). By tracing students' progress over time, researchers receive richer information about how students' understanding progress over time, and about common misconceptions. Meanwhile, research suggests that assessment instruments that are

developed in coordination with learning progressions can provide more information about a larger range of students than typical assessments (Songer, Kelcey, & Gotwals, 2009; Songer & Gotwals, 2012) and offer more discriminatory power than traditional items (Liu, Lee, Hofstedder, & Linn, 2008).

The use of a learning progression is a promising way to bridge research and standards (Duncan & Hmelo-Silver, 2009). It coordinates studies across subject topics and looks, over the years, on students of different grade levels and then integrates all the pieces together to give a reasonable sequence over time. This is very useful for developing educational standards which are across subject areas and across grade levels. The recently released Next Generation Science Standards (NGSS; Achieve Inc., 2013) is informed by learning progression research.

Learning progressions on different science topics have been proposed and verified. For example, Smith, Wiser, Anderson and Krajcik (2006) have built a learning progression for matter and atomic-molecular theory. Merrit, Krajcik and Shwartz (2008) have proposed a learning progression regarding the particle model of matter for sixth graders, while Alonzo and Steedle (2008) have developed a learning progression on force and motion. Fulmer (2014) has validated a force and motion learning progression using data from Singaporean students. Catley, Lehrer and Reiser (2004) have reported the construction of a learning progression for the understanding of evolution. With much learning progression research conducted in the US, one question of interest has been whether the learning progression identified from a sample of students in one country (e.g. American students) can also be shared by their counterparts in another country (e.g. Chinese students). In addition, continuous evaluation and verification of the pre-identified learning progression frameworks are needed.

Carbon Cycle Learning Progression and Two Progress Variables

Previous studies (Mohan et al., 2009; Jin et al., 2012) have identified four achievement levels in American students' learning progression of carbon cycling, based on data collected from written assessments and clinical interviews. These four identified achievement levels are presented in Table 1, where level 1 to level 4 represent increasingly sophisticated scientific understanding about carbon transforming processes. The "description" column describes the general characteristics of the responses at each level and the "Example Response" provides an example response at each level from one item (item label abbreviated as ENPEO). This item asks students "what are the energy sources enabling people to live and grow?" Students are required to choose Yes or No from a list of five things: water, food, nutrients, sunlight, oxygen, and then explain their answers.

At the lowest level, level 1, students' understanding is confined to the macroscopic event of human living and growth without recognizing the

underlying chemical changes and energy transformations in human living and growth. A typical level 1 response is: "People need water when they are exercising, so they can feel energized. You need food so you won't feel weak during the day…" At the highest level, level 4, students can use atomic-molecular models to trace matter/energy systematically in human living and growth. A typical level 4 response is presented in Table 1.

Table 1   Carbon Cycling Learning Progression (Mohan, et al., 2009)

Level	Description	Example Response to item ENPEO
1	Students describe the world in terms of objects and events rather than chemically-connected processes. Their understandings are confined to the macroscopic scale without recognizing the underlying chemical changes or energy transformations of events.	"People need water when they are exercising, so they can feel energized. You need food so you won't feel weak during the day. You need nutrients to help keep your body healthy and strong. When you exercise, it also helps your body stay strong. You need sunlight to feel refreshed."
2	Students continue to attribute events to the purposes and natural tendencies of actors, but they also recognize that macroscopic changes result from "internal" or "barely visible" parts and mechanisms that involve changes of materials and energy in general.	"We drink water and eat food. Food has nutrients and vitamins, which are converted into energy and pumped to your muscles. We do exercise to burn fat. Sunlight does not give us energy; plant use it but not us."
3	Students can reason about macroscopic or large-scale carbon transforming processes, but because of limited understanding at the atomic-molecular scale, they cannot trace matter and energy separately and consistently through these processes.	All living organisms need water; food contains glucose needed to make ATP for energy. Nutrients are a food source. Exercise controls the size of lipids. Sunlight is the energy source of all life.
4	Students can use atomic-molecular models to trace matter/energy systematically through carbon transforming processes connecting multiple scales. They use constrained principles (conservation of atoms and mass, energy conservation and degradation), codified representations (e.g. chemical equations, flow diagrams) to explain chemical changes.	Humans break down carbohydrates as well as fats. The body gets these from food. H_2O, nutrients, exercise, CO_2, and O_2 are all necessary for life, but are not energy sources. Humans don't undergo photosynthesis so sunlight isn't an energy source.

In learning progression research, progress variables are used to track students' increasingly sophisticated understanding of a given concept (Merritt & Krajcik, 2009). Progress variables summarize the important strands of student development that are intended by the curriculum developer (Wilson & Draney, 1999). They are aspects of science knowledge that are present at the achievement levels of the learning progression. In previous studies (Mohan, et al., 2009; Jin, et al. 2012), two progress variables are identified to mediate between big ideas (i.e. carbon cycling) and specific concepts and skills being learned in classrooms: 1) carbon transforming processes and 2) tracing matter and tracing energy practices.

Carbon-transforming processes

The key carbon-transforming processes include photosynthesis, biosynthesis (including digestion), cellular respiration, decomposition, combustion and cross-process events. Cross-process events are events that involve sets of related carbon transforming processes. For example, global warming is an event that involves a set of carbon transforming processes. Previous studies (e.g., Anderson, Sheldon, & Dubay, 1990; Songer & Mintzes, 1994) document a wide range of students' difficulties in understanding carbon-transforming processes. For example, Boyes' and Stanisstreet's (1993) findings show it takes time for students to understand the mechanism of global warming over the course of secondary education. Kempton, Boster, and Hartley (1995) find that many students confuse global warming with ozone depletion. Therefore, carbon transforming processes are key concepts that require time for students to develop sophisticated understanding, which is one progress variable this study uses to trace and contrast American and Chinese students' performances.

Tracing matter and tracing energy practices

Tracing matter practice requires students to use the matter conservation law as a conceptual tool to explain chemical changes, both in amount (quantitative conservation of mass) and in chemical identify of the materials or substances involved in chemical changes. Tracing energy requires students to use energy conservation and energy degradation as tools to explain chemical changes. Numerous studies have found that students intuitively focus on visible aspects of systems and do not use atomic-molecular accounts to explain macroscopic or large-scale events (Hmelo-Silver, Marathe, & Liu, 2007; Lin & Hu, 2003). Thus, they cannot trace matter successfully. A previous study (Wilson, Anderson, Heidemann, Merrill, Merritt, Richmond, Sibley, & Parker, 2006) finds that college students do not trace matter and energy separately to explain chemical changes. Instead, they think fat is "burned up" or "used for energy" when

people lost weight. Tracing matter and energy successfully is crucial for students to understand carbon cycling. Therefore, it is another progress variable this study uses to trace and contrast American and Chinese students' performances.

Science Education in China and in America

China has the biggest education system in the world with over 200 million elementary and secondary students. As the science curricula and science teaching focus on building strong foundational knowledge and mastery of core concepts (Asia Society, 2006), science classes in China are often dominated by teacher lectures and practices in solving assessment items. With large class size constraints (typically 40-50 students per class) and non-availability of equipment, hands-on activities are rare. In addition, science education in China is often criticized for its lack of connection between science and environmental or social impacts (Ding, 2000; Shen, Ying, & Chen, 2006).

On the other hand, science education in America has problems as well. Recent large scale international assessments such as the Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA) reveal that American students' science achievement is lagging behind their peers, especially those in East Asia (Ferraro & Van de Kerckhove, 2006; Mullis, Martin, & Foy, 2008; Provasnik, Gonzales, & Miller, 2009). The 2012 PISA results place the U.S. an unimpressive 27th out of 64 countries in science. Science education in U.S. is criticized for less well trained teachers compared to those in other countries and does not value intellect as much as in other countries (Goldstein, 2014).

Various studies investigated students' knowledge of environmental issues in different countries and their willingness to resolve them (Chhokar, Dua, Taylor, Boyes, & Stanisstreet, 2012; Mutisya & Barker, 2011; Trumper, 2010). For example, Mutisya and Barker (2011) evaluated Kenyan students' awareness of key environmental issues and their knowledge about the causes, effects and solutions pertaining to the environmental issues. Chhokar et al. (2012) investigated Indian students' knowledge about global warming and their willingness to take personal action in reducing global warming.

Through this study, we provide information about students' understanding of carbon cycling and its related environmental issues in another country, China.

METHODOLOGY

Participants

The sample used in this study was based on convenience of location. The American students were from Michigan public schools and the Chinese students were from public schools in Shanghai. There were 1200 American and Chinese students participated in this study from 2008 to 2010.

Table 2 lists the specific numbers of schools and students at each grade level from each country.

Table 2   **The Number of American & Chinese Students at Each Level Who Participated in this Study**

Grade Level	American		Chinese	
	No. of schools	No. of students	No. of schools	No. of students
Elementary	2	85	1	85
Middle	2	233	2	246
High	3	298	2	253
Total	7	616	5	584

The American schools are rural and suburban public schools. According to the great school overall rating, the four elementary and middle schools involved in this study are close or slightly above average schools (the overall ratings are 6 to 7), and the three high schools are relatively good schools (overall ratings are 9 to 10). The Chinese schools are all urban public schools, which are usually better than rural or suburban public schools in China (Park, 2008). Schools in Shanghai are classified as key or non-key schools. Around 5~10% of the schools in each district are key schools that have higher overall educational quality. Only one middle school in the study is a key middle school and the other four schools are non-key schools.

We emphasize that our samples are selected based on convenience and the samples are not statistically representative for students in either country. Our intention is to get a glimpse of students' science learning in both countries. Conclusions we draw based on the samples may not be applicable to all American and Chinese students since the samples are not representative of the whole population in each country.

Assessment Items

A carbon cycle test was designed to assess students' learning progression of carbon cycling[‡]. The test had 44 items in total at three grade levels with approximate twenty percent of the items acting as vertical linking anchor items across grade levels. This followed the recommendation of Angoff (1971) for the minimum number of anchor items in common-item linking. One student typically answered 10~12 items during the test.

Each test item was designed to measure one practice (i.e. tracing matter or tracing energy) and one carbon transforming process. Most items were 'constructed response' (see example items later in the paper) items, each of which requires around five minutes to develop a response. Other items are two-tier items having two parts; a multiple choice or multiple True or False part, followed by a constructed response part that requires students to explain their choice. The assessment items are initially developed in English and then translated by the first author into Chinese. The translated items are first administrated to three Chinese students in each grade level (i.e. elementary, middle and high) to see whether the test is easy for them to understand. After confirming the test is completely understandable, it is administered to all Chinese students[§]

Scoring Rubrics and Scoring Process

The American students' responses were scored using scoring rubrics that corresponding to the response characteristics described at each of the four learning progression levels (see Table 1). We used the same rubrics to score a small sample of Chinese students' responses first (i.e. 50 responses per item). We found that the initial patterns in this small sample of Chinese students' responses were similar to the patterns in American students' responses. These Chinese students' responses could be classified into four achievement levels previously identified. Thus, the rubrics were found to be useful to classify Chinese students' responses too and the rest of the responses were coded using the same rubrics. Ten percent of the responses were randomly selected to be double scored by a second-rater to examine

---

[‡] A refined version of this assessment has been developed and has been used in a subsequent study (see Doherty et al. (2015) for details).

[§] We did not undertake a formal piloting of the test in China because the English version of the test had been piloted and administered in U.S. The quality of the test items had been examined in a number of studies (Mohan, et al., 2009; Chen, Anderson, Choi, Lee, Draney, 2010). It is worth noting that although we tried to make the translated items keep the same meanings as the original items, the translated test may still not be considered as an equivalent test for Chinese students. The context of the items may be more familiar to American students.

the scoring quality and reduce scoring errors**. The inter-rater reliability between the first and second-rater was around 80% for all items. Discrepancies in scoring were discussed and final agreements were reached for each response. These final agreed scores were used in our data analysis.

Data Analysis

First, descriptive statistics and sample t-tests were used to analyse and contrast the general patterns of American and Chinese students' responses. More specifically, IRT based methods were used to analyse students' responses and examine the validity of the learning progression framework for Chinese students. ConQuest (Wu, Adams, Wilson, & Haldane, 2007) was used to estimate both the item and person ability parameters. Among all 44 items administered, 39 items were included for this study. Five items were excluded from this study because they had relatively fewer responses (i.e., fewer than 50 responses from one country or both countries).

RESULTS

Research Question 1: The Validity of Learning Progression Levels for Chinese Students

Evidence from IRT Analysis

If an item fits the partial credit model, the range of Mean-Square statistics (MNSQs) should be between 0.6 and 1.4 and the associated t statistics should be within the range from -2 to 2 (Wu, Adams, Wilson, & Haldane, 2007; Wright, Linacre, Gustafsson, & Martin-Loff, 1994). The results suggest that the MNSQs of all the items were within the acceptable range as are those for the t-statistics except two items that showed slight misfit. The Wright map (see Figure 1) shows the location of item Thurstonian threshold estimates on the person ability scale. The triangle, square and diamond dots in Figure 1 are the Thurstonian thresholds, which are indicators of "score difficulties". The Thurstonian threshold for a score category is defined as the ability at which the probability of achieving that score or higher reaches 50%. The diamond dots represent the difficulty for achieving a score of 2 and above. The square dots represent the difficulty for achieving a score of 3 and above. The triangle dots represent the difficulty for achieving a score of 4. From Figure 1, we can see that in general, if a student's ability increases, he or she is more likely to get higher scores across all the items.

** Because of the limit of time and human resource, only ten percent of the responses were double scored by two human raters.
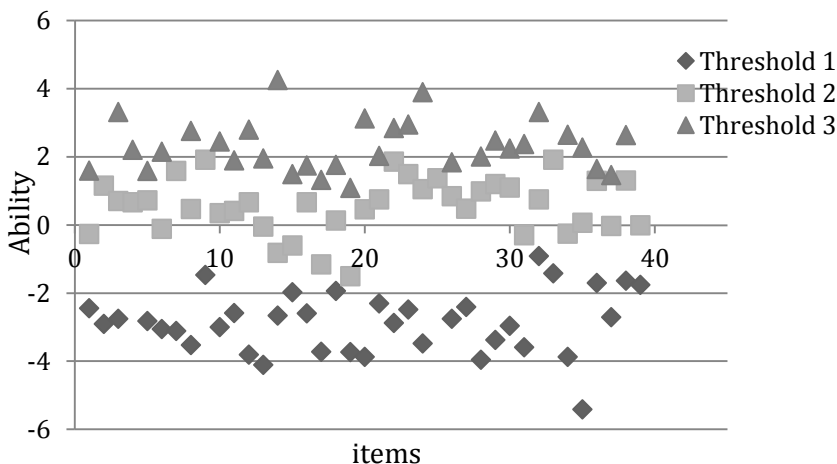
Figure 1   The Wright Map of the item Thurstonian threshold estimates.

Both the item fit statistics and the wright map suggest that most items fit well with the partial credit model. Using IRT models to fit the data is a way to examine the validity of the learning progression framework and the validity of the assessment (Alonzo & Gotwals, 2012; Briggs & Alonzo, 2012). Students' abilities can be indicated by both the achievement levels defined in the learning progression framework and by the IRT ability estimates. When these two different definitions of ability reconcile with each other, it provides evidence for the validity of the learning progression framework and the validity of the single latent construct defined by the four achievement levels.

Results in the Wright map suggest that students with higher ability are more likely to respond at higher levels. The results provide evidence that the four learning progression levels are valid to classify the responses from the Chinese student sample. Their responses are in the same range that can be classified using the four achievement levels previously identified.

Evidence from qualitative analysis

Table 3 presents example responses from both groups of learners at each level from one item. This item (TREET) measures students' understanding of matter transformation during photosynthesis.

TREET (photosynthesis, tracing matter)
A small tree grew into a big tree, where did the extra mass come from?
Choose from [Explain your choice(s)].

a. Soil	b. Air	c. Sunlight
d. Water	e. Minerals in soil	f. Other

Table 3 **Typical Responses from the Chinese and the American Student Samples at Each Learning Progression Level from TREET Item**

Learning progression level	*Responses from Chinese student sample*	*Responses from American student sample*
Level 1 *Simple force-dynamic accounts*	Chose sunlight only. Without sunlight, the tree won't grow. It will die. **[29.1%]**	Chose all. They all combined together and it gets big and strong. **[24.8%]**
Level 2 *Elaborated force-dynamic accounts*	Chose air, sunlight, and water. The tree absorbed water and conducted photosynthesis. **[43.6%]**	The tree uses the minerals from the soil to do photosynthesis. **[53.3%]**
Level 3 *Attempts to trace matter and energy but with errors*	Chose water only. First, there is a lot of water in the tree's root, stem, and leaves. Second, the tree also needs air to provide CO_2, sunlight is energy for photosynthesis, and soil provides elements to produce cellulose. **[10.9%]**	Chose air, sunlight: water, and minerals. It uses air to convert CO_2 into oxygen. It uses sunlight to make glucose. It uses water to get the nutrients from it. It uses the nutrients and minerals in the soil to make glucose. **[15.2%]**
Level 4 *Correct qualitative tracing of matter and energy through processes at multiple scales*	Chose air, sunlight and water. There is CO_2 and O_2 in air. During photosynthesis, chloroplasts absorb and use light energy. CO_2 and H_2O combined into organic materials and release O_2. Light energy is converted into chemical energy. The sugar produced by photosynthesis is converted to starch, involve in the synthesis of amino acid, protein, lipid. So the weight of tree increases. **[3.6%]**	The plants increase in weight comes from CO_2 in the air. The carbon in that molecule is used to create glucose, and several polysaccharides which are used for support. **[1.2%]**

Note: The numbers in **[ ]** are the percentages of responses at each level from students in each country group.

Similar to the responses presented in Table 3, we find broad similarities between students' responses from each country group to the other items. The similarities of the responses from these two groups of learners provide qualitative evidence that the four learning progression levels are valid to classify Chinese students' responses. Therefore, based on both quantitative

and qualitative analyses, we think the Chinese students in our sample share a similar learning trajectory that is followed by American student in the sample.

Research Question 2: American and Chinese Students' Performance on Two Progress Variables: Process and Practice

We compare the two groups' performances using two progress variables—process and practice introduced earlier. In this section, we summarize the similarity and differences between the performances of students in each country group on these two progress variables.

Carbon Transforming Processes

Similarities. The American and Chinese students in our sample have a similar confusion between greenhouse effect and ozone depletion. Data from this study indicate that both groups of learners have the same misconception that global warming is caused by the depletion of the ozone layer which allows more ultraviolet light to reach the Earth but does not contribute appreciably to global warming. For example, the GLOBW item asks students to describe what is global warming. Around 23% of the Chinese students (26/113) and 18% of the American students (18/102) in our sample responded that global warming was caused by the depletion of the ozone layer.

Differences. Table 4 presents some descriptive statistics of the scores of each item from the two groups of learners. It also presents the results from sample t-tests that compare the score difference of the two groups of learners. There are 12 items for which the scores from students in each country group are significantly different (see p-values highlighted in Table 4). For each process, one group of the students perform better on one or two items of that process, but not on all the items or most of the items of that process. So it cannot be concluded that one group of the students perform significantly better than the other group on any of these six carbon transforming processes. By looking more closely at the 12 items that showed significant differences, we find the following patterns.

Table 4   Independent Sample t-test of American and Chinese Students' Scores of Each Item

Process	Practice	Item	Sample size		Mean		SD		Mean Diff	T-test p-value
			US	CH	US	CH	US	CH		
Photo-synthesis	Matter	CARBM	91	102	1.98	2.22	0.83	0.88	-0.24	0.055
		TREET	165	165	1.82	1.64	0.80	0.96	0.18	0.063
		SEEDG	110	102	2.11	2.26	1.00	0.87	-0.15	0.229
	Energy	LIGHT	144	136	2.70	2.57	1.12	1.19	0.14	0.329
		ENPLT	260	265	1.81	1.96	0.82	0.79	-0.15	**0.028**
Cellular Respiration	Matter	AIRNB	138	135	1.48	2.08	0.70	1.05	-0.60	**0.000**
		AIRDIF	105	98	1.80	2.09	0.80	0.75	-0.29	**0.008**
		WTLOS	216	216	1.81	1.73	0.76	1.03	0.08	0.365
		ANIMW	57	74	1.18	1.22	0.57	0.91	-0.04	0.754
	Energy	BODYT	132	136	1.80	2.11	0.76	0.75	-0.31	**0.001**
		STORE	161	164	1.96	1.58	1.15	1.22	0.38	**0.004**
		FDFIN	213	176	1.71	1.77	1.03	1.08	-0.06	0.588
		ENPEO	217	235	1.82	1.68	0.88	1.10	0.14	0.136
		JARED	103	68	1.07	1.09	0.66	0.59	-0.02	0.838
Combustion	Matter	GASWA	111	103	1.68	2.14	0.94	0.94	-0.46	**0.001**
		CARGM	186	188	2.24	2.10	0.74	0.88	0.14	0.094
		WAXBU	100	100	2.13	2.23	0.91	1.01	-0.10	0.463
		BRNMA	78	69	2.33	2.51	0.60	0.68	-0.18	0.100

Table 4 (Continued)   Independent Sample t-test of American and Chinese Students' Scores of Each Item

	Energy	CARGE	349	332	1.83	2.18	0.83	0.87	-0.35	**0.000**
		WAXBU	100	100	1.80	1.77	1.13	0.85	0.03	0.832
		HOTTH	110	103	2.04	2.32	1.13	0.63	-0.28	**0.023**
Cross-Processes	Matter	ENRIC	100	99	1.98	1.80	0.91	0.83	0.18	0.143
		GRANJ	145	135	1.66	1.49	1.10	0.88	0.17	0.161
		TROPM	144	95	1.88	1.77	1.03	1.01	0.11	0.430
		CONNL	162	170	1.69	1.69	0.73	0.64	0.00	0.971
		GLOBW	104	113	1.89	1.86	0.85	1.15	0.03	0.943
	Energy	DIFFE	212	202	1.39	1.72	0.96	0.81	-0.33	**0.000**
		LAMPL	141	70	2.06	1.33	0.98	1.21	0.73	**0.000**
		LBULB	100	102	2.24	1.77	0.96	0.95	0.47	**0.001**
		ECOSP	158	166	1.80	1.95	0.95	1.03	-0.15	0.182
		TROPE	129	62	1.78	1.85	1.07	0.96	-0.07	0.640
Decom-position	Matter	TRDEM	201	203	1.89	1.72	0.81	0.81	0.17	**0.046**
		LEAVE	89	108	1.45	1.60	0.92	1.13	-0.15	0.297
		APPRM	105	97	2.00	2.08	0.93	0.62	-0.08	0.457
	Energy	APPRE	104	89	1.63	1.42	0.89	1.16	0.21	0.167
		TRDEE	351	336	1.66	1.66	0.78	0.90	0.00	0.966
Digest-ion	Matter	INFAN	162	205	1.70	1.87	0.68	0.94	-0.17	0.053
		GROWT	105	102	1.26	1.36	0.71	1.15	-0.10	0.429
		EATAP	295	310	1.94	1.98	0.96	0.91	-0.04	0.583

Note: Mean difference = Mean of Chinese students' scores-Mean of American students' scores; p-values in bold indicate that American students' scores and Chinese students scores of the item are significantly different.

First, the Chinese student sample performed slightly better than the American student sample on some of the combustion items. Independent sample t-test suggested that Chinese students' scores of three combustion related items (GASWA, CARGE, HOTTH) were significantly higher than American students' scores. Here, we took the GARGE item as an example. This item asked students "When you are riding in a car, the car burns gasoline to make it run. Eventually the gasoline tank becomes empty. What happens to the materials of which the gasoline is made when the car uses the gasoline?" More Chinese students were able to identify organic substances in gasoline were converted to gas and water. They also noticed that chemical energy of gasoline was converted into other forms of energy. In contrast, many American students provided vague responses such as the "chemicals" or "substances" in gasoline that were "flammable" were burned during combustion.

Second, the Chinese student sample performed better on items assessing matter transformations in breathing. Independent sample t-test suggested Chinese students' scores of the item AIRNB and item AIRDI were significantly higher than American students ($p <.001$ and $<.01$ respectively). These were two similar items that both assessed matter transformations in breathing. The AIRNB item asked students "Humans get oxygen from the air they breathe, and they exhale carbon dioxide. How does oxygen get used and how is the carbon dioxide produced in the body?" The majority of American students' responses to this item were at Level 1 (44%) or Level 2 (44%) and only 5% of the responses were at Level 3 or above. A typical answer was "oxygen is converted into carbon dioxide in lungs" without explanation about how oxygen was converted into carbon dioxide. On the other hand, Chinese students, especially at high school level, began to recognize that oxygen reacted with substances in the human body such as glucose and carbon dioxide was produced during the cellular respiration process. Around 25% of the Chinese students' responses reached level 3 and 10% of the responses were at level 4.

Third, the American student sample was more aware of the environmental impact of human behaviours than the Chinese student sample. They performed better on two items (LAMPL and LBULB) that assessed the environmental impact of using electrical appliances. The LBULB item asked why using high-energy efficient fluorescent light bulbs instead of incandescent light bulbs would help to slow down global warming. About one third of the American students identified CO_2 released from electricity generation, which caused global warming, while only 13% of the Chinese students were able to make connections between high-energy efficiency light bulbs and global warming. In addition, more American students had heard of global warming or noticed that certain human behaviours led to global warming though they were not very successful in explaining why those human behaviours caused global

warming. When answering the GLOBW item, 90% of the American elementary and middle school students indicated that they had heard of global warming, but only 37% of the Chinese elementary and middle school students replied in a similar manner. Approximately 83% of the American students recognized that "driving trucks on the highway" contributed to global warming while only 68.1% of the Chinese students gave a similar recognition. Around 75% of the American students recognized that cutting down forests contributed to global warming compared to only 54% of the Chinese students. These results indicated that American students were more aware of human social impact on environmental issues compared to Chinese students.

Tracing Matter and Tracing Energy Practices

Similarities. Many students in both groups were unable to trace matter and energy separately and confused matter with energy (e.g. fat burns into energy when people lose weight). For example, the WTLOS item asks students what happens to fat when people lose weight. Only 13% of the American students and 19% of the Chinese students knew that fat left the person's body as water and gas. The majority of the students, around 81% of the American students and 64% of the Chinese students, thought fat was converted to energy and disappears. The rest of the students, around 5% of the American students and 13% of the Chinese students, thought fat was broken down and left the person's body as faces and urine. In both groups, the majority of the students confused matter with energy and could not trace matter and energy separately.

Chinese and American students in our sample had the similar misconception that energy was released when chemical bonds were broken rather than when they were formed. Around 5% of the students in both groups mentioned energy released when chemical bonds were broken. For example, a Chinese student's response to a photosynthesis item was that "Sunlight was used in the photosynthesis. Light energy was converted into chemical energy. It broke the chemical bonds and released a lot of energy." An American student's answer to an item that asked how the grape you ate could help you move your little finger was "in the grape there are carbon bonds which were high energy bonds; when these were broken down by the enzymes in our stomach, energy was released. This energy was then used to move our fingers."

Differences. The Chinese student sample was better at naming energy forms, especially at the middle school level. Many middle school Chinese students were able to name different forms of energy such as chemical energy, thermal energy, kinetic energy; these terms were rarely included in middle school American students' responses. This explained why Chinese students received higher scores on items such as DIFFE, CARGE and ENPLN. For example, when answering to the CARGE item, the

percentages of Chinese middle school students who mentioned "chemical energy" "kinetic energy" and "thermal energy" in their responses were 30%, 31% and 12% respectively. However, none of the middle school American students named these energy forms when answering this item. Although Chinese students performed better in naming energy forms, they did not explain how energy was converted from one form to another significantly better than American students. For example, there were two items about energy transformation in photosynthesis. One item (ENPLN) asked students what was the energy source of tree growth, and the other item (LIGHT) asked students to explain how light helped plants to grow. Chinese students did significantly better than American students on the first item ($p < .05$) but their scores on the second item were not significantly different from those of the American students' (p-value $= .329$).

Many Chinese students in our sample memorized the energy conservation principle, but they only applied this principle in physical systems but not in biological systems. For example, when answering the CARGAS item, many Chinese students included the sentence "energy cannot be created or destroyed" in their responses, which was an important sentence they learned from their textbook. Around 40% of Chinese high school students mentioned the energy conservation principle and applied this principle to explain what happened to the energy in gasoline when gasoline was burnt. Only around 10% of American high school students mentioned the energy conservation principle when answering the same item. However, Chinese students did not apply the energy conservation principle successfully to reason about biochemical processes.

Chinese students in our sample included codified representations (e.g. chemical equations, flow diagrams) to explain chemical changes much more often compared to the American students in our sample. About 10% of the Chinese high school students used chemical equations in their explanations of carbon transforming processes, while only 2% of the American high students included chemical equations when answering the same items. Chinese students at Level 3 or 4 often included correct or partially correct chemical equations when responding to questions. However, this was not usually the case for level 3 or 4 American students.

CONCLUSIONS & IMPLICATIONS

The results suggested that the American and Chinese student samples shared similar general trends in their learning progressions from a simple force-dynamics account to scientific model-based reasoning. Our framework describing four learning progression levels was empirically valid for Chinese students' responses as it was for American students' responses. Results from the IRT analysis showed that the PCM model fitted well to the items, which provided quantitative evidence for the validity of

the learning progression framework for the Chinese student sample. The typical responses from each group of learners provided qualitative evidence of the general similarity of students' learning trajectory. Therefore, we found that although the students in this study were from two different countries and had very different science education systems and cultures, they still shared similar patterns in the development of their scientific knowledge and practice. Very few studies investigated Chinese students' science learning progression and no generalizable conclusions were made about Chinese students' science learning progression. Findings from this study needed to be verified by future studies that investigate the learning progression of Chinese student with a more nationally representative student sample.

Similarities and Differences

There were some similarities and differences between the performances of the students in each country group in terms of the two progress variables: carbon transforming processes and tracing matter and energy practices. With respect to similarities, a certain proportion of the students in both groups had the misconception that greenhouse effect was caused by ozone depletion. This was consistent with Kempton, et al.'s (1995) finding about students' confusion between global warming and ozone depletion. Many students in both groups were unable to trace matter and energy separately. Less than 20% of the students in each group were able to trace matter and energy separately when they answered the question about what happens to fat when people lose weight. Students in both groups had the same misconception that energy was released when chemical bonds were broken rather than when they were formed. A small proportion of responses from both groups mentioned this misconception. Educational systems in both countries were, at present, failing to correct these misconceptions.

In terms of differences, the Chinese group performed better on some combustion items and some items assessing matter transformations in breathing. The American group performed better on some items about the environmental impact of human behaviours. The Chinese students in our sample named energy forms at a younger age than the American students in our sample. Many middle school Chinese students were able to name different forms of energy such as chemical energy, thermal energy, kinetic energy; these terms were rarely included in middle school American students' responses.

More Chinese students in our sample mentioned the energy conservation principle than American students. Around 40% of the Chinese high school students mentioned this principle and applied it to explain what happened to the energy in gasoline when gasoline was burnt. But only around 10% of American high school students mentioned this principle when answering the same item. However, the Chinese students applied the

energy conservation principle mainly to explain events in physical systems, not in biological systems and hence did not explain energy transformations more successfully than American students in biochemical processes. Finally, Chinese students included chemical equations and chemical identities of substances more often in their responses than American students.

Hypotheses of the Causes of the Differences

These differences may result from factors such as curriculum, standards, or teaching focus between two countries. Science subjects are less integrated in Chinese science education compared to those in American science education (Lewin, 1987). At the secondary level, the sciences are taught throughout as separate subjects: physics, chemistry, and biology with some health science. Chinese students may have difficulty connecting knowledge from different disciplinary areas to reason about carbon transforming processes. Chinese students learn energy concepts and energy conservation principles in their physics classes. Though they perform better for items about energy transformations in combustion, they often fail to apply their knowledge of energy into biochemical processes. Chinese national biology curriculum materials mainly focus on matter transformations during processes such as photosynthesis and cellular respiration, with energy flow in ecosystems only introduced in a short section. The transformations of energy in biochemical processes are less emphasized.

As pointed out in the literature, the science curriculum in China may lack a connection between science and environmental or social impacts (Ding, 2000; Shen, et al, 2006). In both middle and high school national biology textbooks, the relationship between humans and ecosystems is briefly introduced in the last chapter. In Michigan's state biology standards, understanding human impact on ecosystems is listed as core knowledge. Students need to understand that, "humans can have tremendous impact on the environment", and be able to "describe the greenhouse effect and list possible causes and list the possible causes and consequences of global warming" (MDE, 2006). This may explain the difference between Chinese and American students' performances for items about environmental impact of human behaviours.

Though more American students indicated that they had heard of global warming, and recognized that human behaviour, such as cutting down trees and driving cars, contributes to global climate change, they still need to learn more about this important topic. Research shows that both American and Chinese citizens are by and large uninformed, or misinformed about environmental science (Leiserowitz, Maibach, Roser-Renouf, & Smith, 2011; CEAP, 2008). The Chinese public is also generally unaware about environmental problems. The Chinese national environmental awareness survey conducted in 2007 shows that among the 3000 participants from

over 20 provinces, only 31.6% of the participants understood the concept of global warming. Over half of the people didn't know that using energy saving appliances would help slow down the global climate change (CEAP, 2008). This is consistent with our data that showed only 13% of the Chinese students were able to make connections between high-energy efficiency light bulbs and global warming. The widespread environmental "illiteracy" among American and Chinese citizens makes it imperative for science education in both countries to emphasize science topics that have important environmental and social impact.

Memorization. Chinese students memorize more chemical identities, chemical equations, energy forms and fundamental energy principles, which are emphasized in Chinese science education. The entire education system in China is examination-driven (Yu & Suen, 2005). Science teachers usually focus their teaching on chemical identities, equations and scientific principles, which are tested by standardized assessments. Students are highly motivated to memorize chemical identities, chemical equations and fundamental principles to get high science scores. This may explain why Chinese students remember and include more chemical representations and fundamental principles in their responses.

Suggestions Based on the Similarities and Differences

The similarities and differences between the performances of the American and Chinese student samples provide us with ideas on how to make improvements for science education in each country. For instance, the common misconception that greenhouse effect is caused by ozone depletion needs to be corrected in both countries. Science education in China needs to place more emphasis on connections between science and environmental or social impacts. In addition, science education in China needs to develop students' science interdisciplinary knowledge to help them connect knowledge from different disciplinary areas. American science education need to pay more attention to developing students' chemical understanding and mastery of fundamental principles.

REFERENCES

Asia Society. (2006). *Math and Science Education in a Global Age: What the U.S. can learn from China*. New York: Education Division, Asia Society.

Adams, R. J., Wilson, M., & Wang, W-C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21,*1-23.

Alonzo, A., & Steedle, J. (2008). Developing and assessing a force and motion learning progression. *Science Education,* 93(3), 389-421.

Anderson, C. W., Sheldon, T. H., & Dubay, J. (1990). The effects of instruction on college non-majors' conceptions of respiration and photosynthesis. *Journal of Research in Science Teaching, 27*, 761-776.

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 508-600). Washington, DC: American Council on Education.

Boyes, E., & Stanisstreet, M. (1993). The greenhouse effect: Children's perceptions of causes, consequences and cures. *International Journal of Science Education*, 15 (5), 531-552.

Briggs, D.C. & Alonzo, A.C. (2012). The psychometric modelling of ordered multiple-choice item responses for diagnostic assessment with a learning progression. In A.C. Alonzo & A.W. Gotwals (eds). *Learning progressions in science: Current challenges and future directions.* Rotterdam, The Netherlands: Sense Publishers.

Catley, K., Lehrer R., & Reiser, B. (2004). *Tracing a prospective learning progression for developing understanding of evolution*, Paper Commissioned by the National Academies Committee on Test Design for K–12 Science Achievement, Washington, DC: National Academy of Science, 67.

Chen, J., Anderson, C. W., Choi, J., Lee, Y., & Draney, D. (2010, March). *Assessing K-12 students' learning progression of carbon cycling using different types of items.* Paper presented at National Association for Research in Science Teaching annual meeting, Philadelphia, PA.

Chhokar, K., Dua, S., Taylor, N., Boyes, E. & Stanisstreet, M. (2012). Senior secondary Indian students' views about global warming, and implications for education. *Science Education International, 23*(2), 133-149.

China Environmental Awareness Project. (2008). 2007 National public environmental awareness survey. *Journal of world environment, 2, 72-77.*

Corcoran, T., Mosher, F. A., & Rogat, A. (2009, May). *Learning progressions in science: An evidence based approach to reform* (CPRE Research Report #RR-63). Philadelphia, PA: Consortium for Policy Research in Education.

Ding, B. (2000). Comparison between Chinese and American science education. *Journal of the Chinese society of education, 2, 52-54.*

Doherty, J., Draney, K., Shin, H., Kim, J., & Anderson, C. W. (2015). *Validation of a learning progression-based monitoring assessment.* Manuscript submitted for publication.

Duncan, R. G., & Hmelo-Silver, C. E. (2009). Learning progressions: Aligning curriculum, instruction, and assessment. *Journal of Research in Science Teaching, 46*(6), 606-609.

Ferraro, D., & Van de Kerckhove, W. (2006). *Trends in International Mathematics and Science Study (TIMSS) 2003: Nonresponse bias analysis.* Washington, DC:

National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Fulmer, G.W. (2014). Validating proposed learning progressions on force and motion using the force concept inventory: Findings from Singapore secondary schools. *International Journal of Science and Mathematics Education.* 01/2014; DOI: 10.1007/s10763-014-9553-x.

Goldstein, D. (2014). *The teacher wars: A history of America's most embattled profession*. New York: Doubleday.

Gregg, J.S., Andres, R.J., & Marland, G. (2008). China: Emissions pattern of the world leader in CO2 emissions from fossil fuel consumption and cement production. *Geophys. Res. Lett,* 35, L08806, doi: 10.1029/2007GJ032887,1-5.

Heimann, M. & Reichstein, M. (2008) Terrestrial ecosystem carbon dynamics and climate feedbacks*, Nature,* 451, 289–292.

Hmelo-Silver, C.E., Marathe, S., & Liu, L. (2007). Fish swim, rocks sit, and lungs breathe: Expert-novice understanding in complex systems. *Journal of the Learning Sciences, 16(3),* 307-331.

Houghton, J. (2005). Global warming. *Rep. Prog. Phys.* 68 1343 doi:10.1088/0034-4885/68/6/R02

Jin, H., & Anderson, C. W. (2012). A learning progression for energy in socio-ecological systems. *Journal of Research in Science Teaching,* 49(9), 1149-1180.

Kempton, W., Boster, J. S., & Hartley, J. A. (1995). *Environmental values and American culture*. Cambridge, MA: MIT Press.

Leiserowitz, A., Maibach, E., Roser-Renouf, C., & Smith, N. (2011). *Global Warming's Six Americas, May 2011*. Yale University and George mason University. New Haven, CT: Yale Project on Climate Change Communication. Retrieved June 1, 2011, from http://environment.yale.edu/climate/files/SixAmericasMay2011.pdf

Lewin, K. (1987) Science education in China: transformation and change in the 1980s. *Comparative Education Review,* 31(3): 419-441

Lin, C.-Y., & Hu, R. (2003). Students' understanding of energy flow and matter cycling in the context of the food chain, photosynthesis, and respiration. *International Journal of Science Education,* 25(12), 1529-1544.

Liu, O. L., Lee, H-S., Hofstedder, C. & Linn, M. C. (2008). Assessing knowledge integration in science: Construct, measures and evidence. *Educational Assessment*. 13, 33-55.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-173.

Merritt, J., & Krajcik, J. (2009, June). *Developing a calibrated progress variable for the particle nature of matter.* Paper presented at the Learning Progressions in Science (LeaPS) Conference, Iowa City, IA.

Merritt, J. D., Krajcik, J., & Shwartz, Y. (2008). Development of a learning progression for the particle model of matter. In *Proceedings of the 8th International Conference for the Learning Sciences* (Vol. 2, pp. 75-81). Utrecht, The Netherlands: International Society of the Learning Sciences.

Michigan Department of Education. (2006). High school content expectations (biology).

Mohan, L., Chen, J., & Anderson, C. W. (2009). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. *Journal of Research in Science Teaching*. 46 (6), 675-698.

Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades.* Chestnut Hill, MA: Boston College.

Mutisya, S. M. & Barker, M. (2011). Pupils' environmental awareness and knowledge: A springboard for action in primary schools in Kenya's Rift valley. *Science Education International, 22*(1), 55-71.

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States.* National Academies Press; Washington, DC.

Park, A. (2008). Rural-urban inequality in China. In: S. Yusuf and T. Saich, (Eds). *China urbanizes, eds. 41–64.* Washington, DC: The World Bank.

Provasnik, S., Gonzales, P., & Miller, D. (2009). *U.S. performance across international assessments of student achievement: Special supplement to "The Condition of Education 2009"* (NCES 2009-083). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Salinas, I. (2009). *Learning progressions in science education: Two approaches for development*. Paper presented at the Learning Progressions in Science (LeaPS) Conference. Iowa City, IA.

Shen, X., Ying, Y. & Chen, B. (2006). The comparison of Chinese and American science educational standards (in Chinese). *Journal of foreign education research,* 5, 1-5.

Smith, C. L., Wiser, M., Anderson, C. W., & Krajcik, J. (2006). Implications on research on children's learning for standards and assessment: A proposed learning for matter and the atomic molecular theory. *Measurement: Interdisciplinary Research & Perspective*, 4(1), 1-98.

Songer, N. B., & Gotwals, A.W. (2012). Guiding explanation construction by children at the entry points of learning progressions. *Journal for Research in Science Teaching, 49*, 141-165.

Songer, N. B., Kelcey, B., & Gotwals, A.W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning in biodiversity. *Journal of Research in Science Teaching. 46*(6): 610-631.

Songer, C. J., & Mintzes, J. J. (1994). Understanding cellular respiration: An analysis of conceptual change in college biology. *Journal of Research in Science Teaching 31,* 621-637.

Trumper, R. (2010). How do learners in developed and developing countries relate to environmental issues? *Science Education International*, *21* (4), 217~240.

van der Linen, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of modern item response theory.* New York: Springer.

Wilson, C. D., Anderson, C. W., Heidemann, M., Merrill, J. E., Merritt, B. W., Richmond, G. Sibley. D. F., & Parker, J. M. (2006). Assessing students' ability to trace matter in dynamic systems in cell biology. *CBE Life Science Education,* 5, 323-331.

Wilson, M., & Draney, W. (1999). *Developing maps for student progress in the BEAR Assessment System,* University of California Press, Berkeley, CA.

Wright, B. D., Linacre, J. M., Gustafsson, J. E., & Martin-Loff, P. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions* 8(3): 370.

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). ACER *ConQuest Version 2.0: generalised item response modelling software*. Camberwell, Australia: Australia Council for Educational Research.

Yu, L., & Suen. H. K. (2005). Historical and contemporary exam-driven education fever in China. *KEDI Journal of Educational Policy, 2*, 1, 17-33.