International Electronic Journal of Elementary Education
Vol. 5, Issue 1, October, 2012.

IEJEE

International Electronic Journal of Elementary Education

International Electronic Journal of Elementary Education is an international, multi-disciplinary, peer-reviewed journal that is online published three times in a year. IEJEE aims to provide a platform for the publication of the most advanced high-quality research in the areas of learning, development, instruction and teaching at elementary level.

International Electronic Journal of Elementary Education publishes three issues per year (March, July, and October). IEJEE provides immediate open access to its content on the principle that making research freely available to the public to support a greater global exchange of knowledge.

ISSN: 1307-9298

Abstracting and Indexing:
Scopus, ProQuest, DOAJ, Education Research Complete™, Google Scholar, EDNA, Index Copernicus, Norwegian Social Science Data Services, NewJour, Open J-Gate.

ISSN:1307-9298

www.iejee.com

Table of Contents

Editorial Letter v

Introduction to the Special Issue: Learning and Instruction in the Natural Sciences 1-4
Benjamin D. JEE & Florencia K. ANGGORO

The reflective experimental construction of meanings about the shape of the Earth and the alternation of day and night 5-26
Paulo VARELA

Parent-child conversations about evolution in the context of an interactive museum display 27-46
Andrew SHTULMAN & Isabel CHECA

Elementary Children's Retrodictive Reasoning about Earth Science 47-62
Julie C. LIBARKIN & Matthew H. SCHNEPS

The Role of CLEAR Thinking in Learning Science from Multiple-Document Inquiry Tasks 63-78
Thomas D. GRIFFIN, Jennifer WILEY, M. Anne BRITT & Carlos R. SALAS

How do Openers Contribute to Student Learning? 79-92
Amber ZERTUCHE, Libby GERARD & Marcia C. LINN

Cognitive factors that influence children's learning from a multimedia science lesson 93-108
Florencia K. ANGGORO, Nancy L. STEIN & Benjamin D. JEE

Dear IEJEE readers,

This special issue of the *International Electronic Journal of Elementary Education (IEJEE)* is devoted to one of the most current and important educational topics of our time: ***Learning and Instruction in the Natural Sciences*** *(K-10 focus).* ***Drs. Florencia K. Anggoro*** *and* ***Benjamin D. Jee*** of the College of the Holy Cross, USA, have done a great job as special issue editors. I am very grateful for the scientific editorialship they have accomplished as two active scholars in their field.

Our generation has experienced many technological developments, innovative solutions and challenging aspects of information society and knowledge society. As contemporary educators we have been witnesses to the debates on learning and instruction in the Natural Sciences in our countries. Since the 1990s comparative international studies, i.e. the *Trends in International Mathematics and Science Study (TIMSS)* and the new demands in the national and globalized labor markets revealed that the educational systems in all countries have to reconsider their existing approaches to the teaching of natural sciences.

As the world is transforming to a knowledge society, our educational systems necessitate appropriate perspectives on and approaches to curriculum development, teaching methods, and the creation of learning environments for natural sciences. All are for the sake of our children, our environmental consciousness, and our future.

Dr. Anggoro and Dr. Jee have succeeded in bringing together papers on this challenging topic from prominent researchers. I am sure, as a reader, you'll be delighted to have the opportunity to read the papers that were written with the aim of contributing to enhancing learning and instruction in the natural sciences. You'll find creative papers and papers that present new ideas and trends and papers that inform us about the new approaches in this area. Regardless of whether we are teachers, teacher educators, and/or educational researchers, this special issue of the International Electronic Journal of Elementary Education (IEJEE) will give us several useful ideas originating from applied cognitive theory.

This well presented special issue hopefully will be one of our sources of inspiration to continue to be a part of the amazing educational field of *learning and instruction in the natural sciences*. This affiliation include some commitments: Working for a better learning and teaching environments in natural sciences in all stages of our educational systems. This is due to the fact that the natural sciences as a school subject suffers alack of qualified teachers and interest from young students. I hope this special issue will make its contribution and will inspire more and more youngsters in our countries and environments to be a part of our circle of natural scientists.

This is the fourth special issue of IEJEE since 2006. Without tireless efforts of Dr. Anggoro and Dr. Jee, it would not be possible to present this issue. I want to express my gratitude to both of them. I also want to thank all the contributors to this special issue. It's a great honor for me to make their knowledge accessible to the readers in the entire world.

ISSN:1307-9298

www.iejee.com

I also want to thank Dr. Karen Michele Zabrucky, Dr. Turan Temur and Dr. Gökhan Özsoy for all the academic advising, professional coordination, and the tireless technical contribution they have done for the realization of this special issue.

Dr. Kamil Özerk
Editor in Chief

International Electronic Journal of Elementary Education, 2012, 5(1), 1-4.

Introduction to the Special Issue: Learning and Instruction in the Natural Sciences

Benjamin D. JEE
College of the Holy Cross, Worcester, MA, USA

Florencia K. ANGGORO
College of the Holy Cross, Worcester, MA, USA

Education in the natural sciences is receiving increased interest worldwide. The president of the United States, Barack Obama, has consistently affirmed his commitment to a renewed focus on science education, and other world leaders have also expressed the need to re-imagine science education to better prepare students for the jobs and important challenges of the future. An understanding of science will be vital as the next generation of global citizens confronts complex problems such as climate change, sustainable energy, food production, and the control of disease and illness. Indeed, the educational systems of the future must not only prepare the next generation of scientists, but also produce an informed citizenry, capable of understanding and using scientific evidence to inform their opinions and choices.

A number of factors complicate the growing need to educate students in the natural sciences. For one, increasing urbanization has widened the disconnection between humans and nature (e.g., Atran, Medin, & Ross, 2004; Birnbaum, 2004; Wolff, Medin, & Pankratz, 1999). Students may have very little exposure to the natural world, and thus may have difficulty understanding the scale, properties, and behavior of natural objects (Birnbaum, 2004). The ethnic makeup of many countries is also changing rapidly; yet, relatively few minority students pursue degrees in natural science. For example, of the over 23,000 Bachelor's of Earth Science degrees awarded in the U.S. between 1996 and 2001, only about 5% were awarded to minorities (Stokes, Baker, Briner, & Dorsey, 2007). Given the growing minority population in the United States and other countries, this lack of minority representation poses a significant threat to the future of industries and institutions that depend on science graduates. Another challenge is that many students have negative attitudes toward science (e.g., Atwater, Wiggins, & Gardner, 1995; Simpson & Oliver, 1985). Several researchers have found that students' attitudes toward science decline as they progress through school (Atwater et al., 1995; Cannon & Simpson, 1985; Hill, Atwater, & Wiggins, 1995; Simpson & Oliver, 1985). These attitudes covary with science achievement (e.g., Freedman, 1998), such that students with the most negative attitudes tend to also perform the worst. Poor achievement, in turn, leads to increasingly negative feelings (e.g., Mattern & Scheau, 2002).

As cognition and learning researchers, our background leads us to think about these educational issues in terms of student learning, and the instructional approaches that can

ISSN:1307-9298

www.iejee.com

enhance this learning. If more students can be given the support to achieve in the natural sciences, this may promote positive attitudes, and encourage a wider range of students to pursue a career in science and fulfill the important needs of the future. Our aim as editors of this special issue was to bring together a variety of research approaches that share the common goal of understanding and improving learning and instruction in the natural sciences.

The articles in this special issue address some of the diverse factors involved in children's learning in the natural sciences. One important factor is the *pre-existing belief structures* that children hold during instruction. Children possess naïve theories about the natural world—preconceptions that often deviate from scientific theories and are resistant to change. Varela presents an in-depth study of 1st grade students learning about fundamental topics in astronomy. His paper describes the class discussions that supported students' transition to a more sophisticated and coherent understanding. Varela suggests that discussions involving an instructor and fellow students can make children reconsider their naïve theories and promote conceptual change toward scientific models. The paper by Shtulman and Checa also considers how discussions with others, in this case a parent, can help children to avoid and resolve misconceptions. The research captured children's reasoning in the context of an interactive museum display about biological evolution. Children were less likely to express common misconceptions when the conversation between parent and child was more back-and-forth. Both studies portray a dynamic learning process in which children are exposed to limitations in their naïve theories and presented with alternative ideas from an instructor, parent, or peer.

Another important factor that affects children's learning of natural science is the *thinking and reasoning skills* that they bring to bear. Libarkin and Schneps present the findings from interviews in which children were asked to explain various phenomena related to Earth science. This task required children to perform retrodictive reasoning—describing possible causes of observed phenomena. All children showed the ability to reason retrodictively, yet they generated a wide range of explanations by drawing on different beliefs about domain-specific and domain-general mechanisms in Earth science as well outside-domain analogies. Griffin, Wiley, Britt, and Salas present evidence that individual differences in children's commitment to logic, evidence, and reasoning (CLEAR thinking) predicts their learning from a multiple-document inquiry task in science. Children who were more committed to the use of scientific evidence to inform their beliefs, for example, were generally more effective at comprehending and integrating the content of a series of science texts. These articles speak to an apparent bidirectional relationship between reasoning skills and science learning: general reasoning skills can shape a child's science knowledge, and science knowledge can inform a child's explanations of natural phenomena.

In addition to naïve conceptions of the natural world and scientific reasoning skills, children's learning in the natural sciences is affected by the structure and content of formal instruction. As Anggoro, Stein, and Jee discuss, formal instruction on the molecular properties of the states of matter is often incomplete and incoherent. To support learning, children must be presented with a causally coherent lesson, and visual models must be used to show children the invisible molecular properties of the states. Yet, even when the lesson's content is arguably ideal, instruction can place too great a demand on a child's ability to interpret, maintain, and integrate verbal and visual representations. Anggoro et al. found that children who were read a science text by a tutor showed greater learning gains than students who read the same text by themselves, especially younger children. The finding suggests that supporting the integration of verbal and visual information is a critical consideration when a science lesson is comprehensive and causally coherent. Another way to enhance students' understanding of a science lesson is by leading the lesson with brief activities (Openers) that emphasize knowledge integration. Zertuche, Gerard, and Linn found that when students were given opportunities to make connections among ideas and reflect upon these ideas, they were more likely to form coherent hypotheses and explanations about chemical reactions. These articles offer some important guidelines in the design of the instructional conditions that support optimal learning of natural science.

The papers in this special issue highlight some of the many considerations relevant to children's learning in the natural sciences. The articles also speak to the many contexts in which science learning takes place—from informal conversations in a museum, to searching through online texts, to formal technology-assisted instruction. Indeed, children develop their knowledge of natural science from many places. To engineer the educational systems of the future, educators must not only think beyond conventional modes of instruction, but also beyond the classroom.

• • •

References

Atran, S., Medin, D., & Ross, N. (2004). Evolution and devolution of knowledge: A tale of two biologies. *Journal of the Royal Anthropological Institute*, *10*(2), 395-420.

Atwater, M. M., Wiggins, J., & Gardner, C. M. (1995). A study of urban middle school students with high and low attitudes toward science. *Journal of Research in Science Teaching*, *32*(6), 665-677.

Birnbaum, S. (2004). Overcoming the limitations of an urban setting through field-based earth systems inquiry. *Journal of geoscience education*, *52*(5), 407-410.

Cannon Jr, R. K., & Simpson, R. D. (1985). Relationships among attitude, motivation, and achievement of ability grouped, seventh-grade, life science students. *Science Education*, *69*(2), 121-138.

Freedman, M. P. (1998). Relationship among laboratory instruction, attitude toward science, and achievement in science knowledge. *Journal of Research in Science Teaching*, *34*(4), 343-357.

Hill, G., Atwater, M., & Wiggins, J. (1995). Attitudes toward Science of urban seventh-grade life science students overtime, and the relationship to future plans, family, teacher, curriculum, and school. *Urban Education*, *30*(1), 71-92.

Mattern, N., & Schau, C. (2002). Gender differences in science attitude-achievement relationships over time among white middle-school students. *Journal of Research in Science Teaching*, *39*(4), 324-340.

Simpson, R. D., & Steve Oliver, J. (1990). A summary of major influences on attitude toward and achievement in science among adolescent students. *Science Education*, *74*(1), 1-18.

Stokes, P. J., Baker, G. S., Briner, J. P., & Dorsey, D. J. (2007). A multifaceted outreach model for enhancing diversity in the geosciences in Buffalo, NY. *Journal of Geoscience Education*, *55*(6), 581.

Wolff, P., Medin, D. L., & Pankratz, C. (1999). Evolution and devolution of folkbiological knowledge. *Cognition*, *73*(2), 177-204.

International Electronic Journal of Elementary Education, 2012, 5(1), 5-26.

The reflective experimental construction of meanings about the shape of the Earth and the alternation of day and night*

Paulo VARELA**
CIEC-Institute of Education, University of Minho, Portugal

Received: 2 September 2012 / Revised: 5 November 2012 / Accepted: 11 November 2012

Abstract

The purpose of this paper is to describe and analyze the process of construction of meaning about the shape of the Earth and the alternation of day and night, which is inherent to the practice of experimental science teaching. This teaching practice was gradually done by the researcher in a 1st grade class of a Portuguese primary school. The class was composed of 18 students, ten girls and eight boys, with ages ranging from six to seven years old. The analysis of the meaning construction process focused on the class diary prepared by the researcher, based on the field notes and audio recordings made during the participant observation in the classroom. The goals of the interpretive analysis of the diary were as follows: a) identifying the students' initial ideas expressed during class about the shape of the Earth, b) characterizing the processes that promote the construction of knowledge about the topics under study; c) and presenting the learning that takes place during class. These instances of learning described in the class diary, combined with the results of a true or false questionnaire, suggest that most students developed a good understanding about the shape of the Earth and the alternation of day and night.

Keywords: Conceptual Development, Evolution Understanding, Parent-Child Conversation, Informal Learning Environments, Science Education.

Introduction

The identification of the children's intuitive ideas about various science topics was, over the last three decades, a powerful research guideline in the field of cognitive science and science education. Several studies have demonstrated that children construct, from an early age, intuitive mental models about the shape of the Earth and the alternation of day and night that diverge from the scientific model (Nussbaum, 1985; Vosniadou & Brewer, 1992, 1994; Fleer, 1997; Siegal, et. al, 2004; Blown & Bryce, 2007; Özsoy, 2012). As an example, Vosniadou and Brewer (1992) identified, in primary school children in the U.S., five alternative mental models of the Earth: the

* This manuscript and the case study about the shape of the Earth and the alternation of day and night are an integral part of the author's doctoral thesis.

** Paulo Varela, Centro de Investigação em Estudos da Criança, Instituto de Educação, Universidade do Minho, 4710-057 Braga, Portugal, Phone: +351 253601221. E-mail: pibvarela@ie.uminho.pt

ISSN:1307-9298

rectangular earth, the disc earth, the dual earth, the hollow sphere, and the flattened sphere.

Nevertheless, the vast knowledge produced by such lines of research, especially within science education, has garnered criticism from some constructivists as regards its diminished influence on the improvement of the children's learning and teaching. Matthews (2000) claims that this theoretical knowledge *"*offers very little guidance for teachers who are in the classroom trying to teach Science contents" (2000, p.270). In the same sense, White states that "...although the research on alternative conceptions has sparked interest in the content, it has not yielded clear advice about how to teach different topics" (1994, p.255).

In many countries, the primary school science syllabuses, while they recommend activities based on inquiry methods, hands-on activities, dialogues, discussion and collaborative working, do not include any information on how teachers should implement such activities in their approach to the various teaching contents (Eurydice, 2011). Despite all efforts, in many countries these curricular guidelines still do not bear the necessary influence to change the pedagogical practices of teachers. For example, Martínez and Díaz (2005), when referring to the Spanish educational context, mention that the curricular guidelines for primary education have been advising on more innovative Science teaching, based on an active and constructive role for the student. However, the reality in that country's classrooms is quite different, as the authors point out: "studies and polls carried out show that Science teaching as a transmission is still predominant, and it is based on blackboard master-classes, on the school textbook and the solving of closed problems related to the studied themes" (2005, p. 243).

The promotion of inquiry-based Science teaching is a highly demanding challenge, which calls for major changes in the teaching practice (Harlen, 2010; Brand & Moore 2011). Indeed, most primary school teachers have insufficient scientific knowledge and are not familiar with these teaching strategies, thus depriving their students from the opportunity to engage in relevant and meaningful learning (Harlen, 1999; Lee, et. al, 2004; Appleton, 2003; Brand & Moore, 2011). Harlen (1999), in analyzing several studies about the teachers' understanding of Science conducted during the 1990s in countries such as the USA, England and Scotland, concludes that primary school teachers have a low level of confidence about teaching Science and understanding science concepts. These insufficiencies have implications for the students' learning opportunities, and are commonly associated "with restricting classroom activities to following instructions and inhibiting creativity and questioning" (Harlen, 1999, p.81). Faced with those limitations, Appleton (2003), in a study performed with Australian pre-service teachers, states that some teachers avoid teaching Science or rely on activities with little scientific content, which are usually conducted as a demonstration. Other teachers, however, find support in Science textbooks and worksheets, whose hands-on activities are typically presented with step-by-step instructions (Huber & Moore, 2001). The way these activities are generally introduced in textbooks and developed by the teachers does not foster intellectual engagement on the part of the students. Instead, it leads to much physical action and little mental activity (Harlen, 2007), and it tends to give students a distorted and fallacious view of the nature of Science (Huber & Moore, 2001; Levinson, 2002). Teachers should build learning opportunities and encourage students to engage in genuine inquiry activities or, as mentioned by Jorgenson (2005) in "hands-on minds-on activities".

In Portugal, the experience of more than three decades has shown that the introduction of Science themes and new approaches to school knowledge construction in the Primary Education programs has failed to produce any significant effects in the renovation of pedagogical practices and subsequent improvement of the quality of student learning (Sá, 2002a). Children in primary schools still do not have opportunities to develop the "experimental attitude", highly praised in the Science component of the Environmental Studies curricular area (Ministry of Education, 2004), which leads to the neglect of important domains of knowledge construction and skill development that are necessary all across the different curricular areas (Varela, 2012). Memorization and repetition activities are overrated, and the students keep performing stereotyped, meaningless tasks (Roldão, 2003). In this learning environment, the student takes on a passive role, fundamentally limited to the accumulation of knowledge. Learning loses relevance, and its personal and social use becomes ineffective.

The contact with schools through interventions that we have been conducting in primary school classrooms (Sá & Varela, 2004, 2007) has allowed us to verify that children do not usually have opportunities to conduct practical or experimental science activities which: potentiate their intellectual, personal and social development; stimulate thinking and conceptual understanding (Zohar, 2006); promote language use as a tool for constructing and sharing knowledge (Aleixandre, 2003; Rivard, 2004; Ibáñez & Alemany, 2005); stimulate discussion and argumentation around the students' ideas and the experimental evidence they produce (Naylor et. al, 2007); promote in students an active, autonomous regulator and reflective role on their own learning (Cleary & Zimmerman, 2004); and develop positive attitudes towards Science learning in children (Harlen, 2007). This reality is perceived also by Costa, who referring to practical and experimental Science activities states the following:

> "(…) the way that they have been used has contributed nothing at all to the learning of scientific concepts by students, or to the understanding of the processes used by Science or even to the acquisition of transversal skills. (…) they are rarely used in ways that develop skills of observation, inference, communication, interpretation and planning. Instead, they are more often used as a treat for students "if there is time left" (which rarely happens), or, at best, as an attempt to engage the less motivated students" (2006, p.33).

Thus, primary school students are rarely involved in a genuine process of scientific meaning construction and development of cognitive resources, which are based on a direct relation to concrete objects, manipulating, feeling, experiencing them and reflecting on the observations they make and the actions they carry out with them. Without this knowledge and resources, the students will lack the foundations on which to build essential skills for new forms of learning, which are necessary all throughout the different curricular areas and that will ensure them a participating and informed citizenship in the future (Sá & Varela, 2007).

In this context, we have been developing for over a decade and a half a research and intervention work in the classroom, geared towards an experimental reflective approach of Sciences in the early years of schooling (Sá, 2002a; Sá & Varela, 2004, 2007). This paper is situated in the continuity and deepening of this perspective of teaching and research, and its content is a small part of a study conducted by the author (Varela, 2012).

Objectives

In this paper the process of meaning construction is described and interpreted through the study of "the shape of the Earth and alternation of day and night", aiming at specific purposes: (a) identifying the students' initial ideas about the shape of the Earth, (b) identifying and characterizing processes that stimulate classroom construction of meanings about the topic under study, (c) and presenting the learning that takes place during class.

Reflective Experimental Science Teaching – R EST

REST places great emphasis on the stimulation of the student's reflective thinking skills, integrating and intensifying, in an interdependent manner, the development of cognitive processes and conceptual comprehension (Miras, 2001; Sá 2002a; Zohar, 2006; Harlen, 2007). It is an approach to Science teaching in which:

> "(...) experimental activities are not simple manipulations executed mechanically by imitation, or following instructions provided by the teacher or described in a textbook. On the contrary, they are actions with strong intentionality, closely associated with the student's mental processes. It is this combination of thought and action that leads to higher quality learning" (Sá, 2002a, p.47).

Learning takes on a dynamic and evolving nature of (re)construction of socially constructed meanings, which depart from the ideas that students construct in their personal and sociocultural experiences. When explained in the social context of the classroom, these are subject to a generative and reconstructive process of new meanings with greater power to explain physical and natural phenomena (Sá, 2002a; Harlen & Qualter, 2005; Harlen, 2007). Learning starts from:

> "relevant problems and personal ideas that describe and interpret them, in order to gradually construct, through a process of critical contrast with other ideas and with reality phenomena, a school knowledge that is socialized and shared by means of processes of conceptual change and evolution" (Porlán, 1998, p.101).

In the teaching and learning process, students confront their ideas and expectations with the experimental evidence produced (Harlen & Qualter, 2005; Harlen, 2007) in a methodical, organized and intentional way. The student thus becomes gradually skilled in the process of coordinating personal theories with evidence (Kuhn, et. al, 1988), aiming for a progressive harmonization and conformity of the new theories with the physico-natural world. However, the perspective of conformity between theories and experimental evidence is different for each subject, i.e. "the same experience or the same observation are experienced, seen and understood very differently by different children" (Charpack, 2005, p.29). For this reason, the meanings constructed by way of physical interaction with materials and objects are the subject of discussion and reflection in small and large groups, so that the critical selection and negotiation leads to higher-level meanings, shared by a growing number of students (Naylor et. al, 2007; Domínguez & Stipcich, 2009). It is in the process of social interaction that the different interpretations of physical experience are confronted, negotiated and reconstructed and it is in that interactive process that the different meanings are defined and refined (Candela, 1999).

In REST, the creation of collaborative contexts has a particular importance, as they facilitate the emergence and exchange of different meanings and explanatory interpretations for the various learning situations (Larkin, 2006) and stimulate the joint construction of scientific meanings (Palincsar & Herrenkohl, 2002). We thus recognize the importance of promoting spaces for collaborative mediation and negotiation of meanings, which stimulate students to share opinions among them and with the teacher, to defend their points of view and to justify and/or refute the arguments presented (Henao & Stipcich, 2008). The discussion generated in the classroom provides children with the awareness of their own ideas, the different ideas and ways of thinking that exist in the group (Larkin, 2006; Harlen, 2007) and the need to review and/or restructure their ideas, in face of other more plausible and consensual ones that appear in the social context of the class (Varela, 2012). Through this intense collaborative activity, children also learn, by the action of others and the teacher, to monitor and auto-regulate their own thought and gain access to a wider range of problem-solving strategies (Mercer & Littleton, 2007).

Thus, the teaching and learning process is aimed at encouraging students to reach the highest limit of their potential, i.e. their “zone of proximal development”, proposed by Vygotsky, allowing for the awakening of “a variety of internal development processes that are able to operate only when the child is interacting with people in his environment and in cooperation with his peers” (Vygotsky, 1978, p.90).

From sociocognitive activity, which takes place alternately in small and large groups, emerges the need for more refined observations of the evidence, as well as for the repetition of experimental procedures, which are accompanied by a more reflective attitude from the student. This attitude brings forth new ideas, propelling the discussion to higher quality thinking levels, inducing metacognitive and learning self-regulation skills in students (Larkin, 2006), while also favoring a high degree of transference of the learning acquired to new contexts (Georghiades, 2006) and the autonomy of students (González, & Escudero, 2007). Individual meanings, when explained, reflected upon, contradicted and negotiated will result in a smaller number of meanings, now enriched and shared by a large group of students (Sá & Varela, 2004).

REST lends special importance to the role of oral language as an instrument of communication and construction of scientific meanings (Català & Vilà 2002; Aleixandre, 2003; Maloney & Simon, 2006). Students often resort to written language, which requires greater awareness of the mental operation one executes, developing a process of inner speech within the subject himself (Vygstsky, 1987). Writing implies thinking about whatever is the object of the writing, organizing ideas, establishing connections between them, selecting the best words and articulating them correctly. Indeed, when we encourage students to develop the regular habit of writing about the experimental activities, we are simultaneously giving continuity to the reflective process, promoting the highest level of learning within their reach.

REST implies renewed roles for students and teachers. In this teaching practice, students:

- explain their ideas and ways of thinking about questions, problems and phenomena;
- argue and counterargue among themselves and with the teacher regarding the validity of their ideas and strategies;
- mentally construct simplified research plans with their peers;
- carry out the resolution plans and strategies for the problem situations they are confronted with;
- submit personal ideas and theories to the critical confrontation of their peers and to the test of evidence by resorting to the scientific processes;
- keep written records of their observations and evidence data, as an integral part of the exploration of practical and experimental situations;
- critically assess the conformity level of their theories, expectations and predictions with the ideas of others and with the experimental evidence they produce;
- negotiate different personal perspectives about evidence, questions or problems, aiming for the construction of enriched and socially shared meanings (Sá & Varela, 2004).

The teachers, for their part, take on a role of high activity, reflection and strong pedagogic intentionality: a) in the interpretation of the actions carried out by the students and in the meanings that are generated and reconstructed in the classroom, in order to regulate and re-feed the students’ mental constructive activity; b) in the stimulation and mediation of the students’ interactions with the experimental evidence they produce with their peers; c) in the promotion of an active participation by the students, providing them with the necessary stimulus for verbalization, action and reflection; d) in the valorization and regulation of the discussions that arise around the students' interventions; e) in the creation of an environment of collaboration, accountability and freedom of communication; f) through continuous and recurrent reflective questioning, which stimulates the students' thoughts and actions (Varela, 2012). This questioning will provide, at each moment, adequate help to the needs expressed by students in order to escalate to progressively higher levels of thought and learning (Rojas-Drummond & Mercer, 2003; Chin, 2006; Molenaar, et. al, 2011).

Method

The study is developed according to an action research approach within the theoretical framework of interpretive research, applied to the study of teaching and learning processes in a classroom context (Erickson, 1986; Guba & Lincoln, 2000).

A 1st grade class from a Portuguese primary school located in the outskirts of the city of Braga, composed of ten girls and eight boys (n=18) with an average age of 6.25 years, was subjected to a process of REST. Distributed over one school year, 20 lessons were taught addressing various science topics within the curricular area of Environmental Studies, amounting to a total of 40 hours of intervention in the classroom.

Each lesson, which corresponds to one action research cycle, begins with a teaching and learning plan that takes the form of a starting “curricular hypothesis” (Porlán, 1998) to be implemented flexibly, according to the teaching and learning processes that are generated and promoted in the class reality. The teaching and learning plan pertaining to the curricular topic on the shape of the Earth and the alternation of day and night was prepared according to the didactic sequence suggested by Vosniadou (1991) and Vosniadou et. al, (2004)[1]. According to these authors, in learning basic concepts of astronomy there should be an interrelation between the understanding of the spherical shape of the Earth, the rotation of the Earth with regard to the apparent movement of the Sun and the explanation of the alternation of days and nights.

At the time of the pedagogic intervention on "the shape of the Earth and the alternation of day and night", the students had already benefited from the cumulative effect of 32 hours of REST. The classes were taught by the researcher, who, in collaboration with the class teacher, played the role of both a researcher and the teacher. Thus, there was an attempt to capture and understand the processes of generating and (re)constructing scientific meanings promoted in the classroom, in a social learning context. The researcher-teacher's attention was especially focused on the interpretation of the meanings manifested by students in the moments of communication, action and interaction with their peers and the researcher, and on how these meanings were being reconstructed and negotiated within the class.

The data generated in the action were collected using two complementary methods, namely the fieldnotes made by the researcher and the audio recordings of the lesson. This raw data were later materialized in the form of detailed narratives of the most relevant events that occurred in the classroom – the class diary. These constituted the principal method of recording data and, simultaneously, a strategy of reflection and modeling of the teaching and learning process (Sa, 2002b; Zabalza, 2004).

The lesson, as a referential unit of analysis represented in the class diary, is therefore composed of a sequence of learning moments that correspond to more particular units of analysis. Each unit of analysis is the bearer of a specific sense that distinguishes it

[1] See the full teaching and learning plan, in Portuguese, in Sá and Varela (2007) or part of the plan, in English, in Varela (2012).

from other units – "unit of meaning" (Ratner, 2002), in the evolving and interactive process of constructing scientific meanings. In each diary we began by identifying the sequence of units of meaning. We then carried out the interpretive analysis of the meaning of the data concerning each identified unit and the definition of its central theme, based on that analysis.

The data contained in the diary represent a sample of the diversity of meanings that the students construct in the classroom, while interacting with their peers and the teacher, as well as during the activities conducted. A true or false questionnaire about the "shape of the Earth and the alternation of day and night" was therefore applied in order to attain a more reliable perception concerning the level of individual learning achieved by the students. In view of the dynamic character of the constructive process of meanings - the range and depth of learning only occurs after some time (Coll & Martín, 2001) -, the questionnaire was applied at two different moments, i.e., immediately after the lesson and after three weeks.

The interpretive content analysis of this class diary is the starting point to interpret the process of teaching and learning promoted in the classroom about "the shape of the Earth and the alternation of day and night".

Interpretive quality criteria adopted in this study

The regular presence of the researcher with the children in the classroom for various periods of time, in an accumulated total of 32 hours until the lesson about "the shape of the Earth and the alternation of day and night", ensures a "prolonged involvement" (Guba & Lincoln, 2000). The researcher's prolonged presence affords a progressive construction of a relationship of empathy and openness with the students. This relationship is essential so that the research subjects may share their views with the researcher. In these circumstances, the researcher can access multiple perspectives of meanings from the students' perspective, and can linger on them, analyzing them in depth and detail in order to derive a better clarification and understanding in the context in which they occur (Pérez Gómez, 2005).

"Listening to participants", proposed by Guba and Lincoln (1989), can hardly be sustained in this study. The subjects in this study are 1st grade students. From our point of view, it is not feasible to ask six and seven-year-old children to audit the representations that the researcher generates from the meanings they constructed in the teaching and learning process. Thus, while maintaining a certain parallelism to "listening to participants", a process of recurrent "validation" was adopted in the identification of the students' constructions. This consisted of ascertaining the meaning of what children say and do, at the time and in the context, through a systematic interaction with them. Thus, the researcher may put his/her inferences to the test, in a close and situated manner, because he/she is an active subject in the observation context (Erickson, 1986).

The use of audio recording is a procedure that lends more credibility and veracity to the qualitative data collected. Audio recordings afford the researcher greater availability to reflect and interact with the subjects in order to ensure that the meanings referred by them are correctly interpreted and represented (MacLean, et. al, 2004). Moreover, in situations of ambiguity or inconsistency (uncertainty) regarding the meanings inferred in the classroom, the recourse to a later hearing of the recording can clarify the intended meaning from the original source (Fasick, 2001). From our point of view, revisiting the original data, by hearing the recordings at a later time, allows for a novel outlook, distanced from the data itself and the interpretations made in the course of the participant observation, in order to construct a representation of the studied reality as accurate as possible.

The combination of the preceding techniques allows for the collection of “abundant information” (Carrasco & Hernández, 2000). After class, to best take the advantage of fresh memory, the class diary was written on the basis of the audio recordings and the field notes. The diary includes verbatim transcripts of what the children say in meaningful moments of interaction, events of non-verbal nature and emotional aspects not captured by audio recording. It represents the events generated in the classroom by means of a thick description (Denzin, 1989). A rich and detailed narrative of the observations increases the credibility and plausibility to an external reader regarding the data and the inferences made. Furthermore, the act of creating a narrative increases the confirmability of the study, since it allows us to distance ourselves from our judgments, premature interpretations and provides an opportunity to open our work to the inspection of others (Newman, 2000).

The analysis of the class diary includes segments of raw data so that an external reader can judge the credibility and neutrality of the inferences made from their meanings (Lincoln & Guba, 1989; Ratner, 2002). This constitutes a relevant factor for the confirmability criterion, i.e. to verify that we are not in the presence of arbitrary constructions imagined by the researcher.

Results

Interpretive content analysis of the class diary about the shape of the Earth and the alternation of day and night

The teaching and learning activities begin by identifying what students think about the shape of the Earth. Students are asked to draw the shape of the Earth.

What ideas do children present about the shape of the Earth?

The interpretation of the meaning of the drawings takes place in two moments: i) in class, through observation, communication and discussion generated around the intended meaning of the drawings; ii) after class, through a more detailed analysis of the content of the drawings and the arguments presented to the class by the students. We identified three categories, whose content represents qualitatively different ideas about the shape of the Earth:

A. Most students' drawings evidence the idea that the Earth is flat. At the bottom of the drawing, the Earth extends down and to the sides. On the surface, the students drew houses, trees, people, etc. Above that, there is the sky and/or space, with some birds, clouds, stars and the Sun. This is the most primitive conception of the shape of the Earth identified in the class (11/18; 61.1%).

Figure 1. Gabriel; 6.5 years

B. A second category of drawings apparently considers the idea that the Earth is round. However, what is relevant about the sphere on the drawing is what is inside it: i) a well-defined area at the bottom, which is round on the bottom and flat at the surface. That is where the trees, houses and people are; ii) the top part corresponds to the sky and/or space, where some birds and flying insects, the stars and the Sun are drawn. Although these students claim that the Earth is round, it seems very plausible that this model results from the incorporation of the scientific information

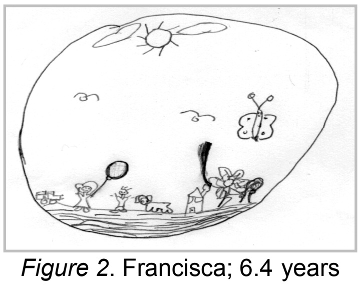


Figure 2. Francisca; 6.4 years

regarding the sphericity of the planet into the previous model. The Earth itself would be the bottom part, with the flat surface, whereas the top part would correspond to the sky and/or space. Phrases like "our country is inside the Earth" suggest that the word "Earth" can either take on the meaning of the cosmic body we inhabit, or that of a cosmic entity that contains the Earth and the space inside it. Reinforcing this interpretive hypothesis is the fact that the expression "inside the earth" does not, in any way, mean "below the Earth's crust", but rather "inside" the sphere on the drawing (3/18; 16.7%). This interpretation is further validated by other authors who have identified the same concept in children belonging to the same age group (Nussbaum, 1985; Vosniadou, et. al, 2004). Vosniadou et. al, (2004) call this model "hollow Earth", as it is a synthetic model derived from the children's attempts to incorporate the scientific information that says the Earth is a sphere into the initial concept that the Earth is a supported and stable plane.

C. In a third category of drawings, the Earth appears as a spherical body surrounded by space, where the stars and the Sun are drawn. On its surface there are countries, continents and oceans. For these students, people live on the surface and not "inside" the Earth: "it's on the outside"; "people walk up here on the land"; "they also ride boats on the sea"; "and swim and ride water scooters" (4/18; 22.2%). The meaning of the drawing is communicated to the class as follows: "I made the Earth round, seen from Mars. In space I drew the Sun and the stars, and here (on Earth) I drew the islands, the seas and people's lands (countries)".

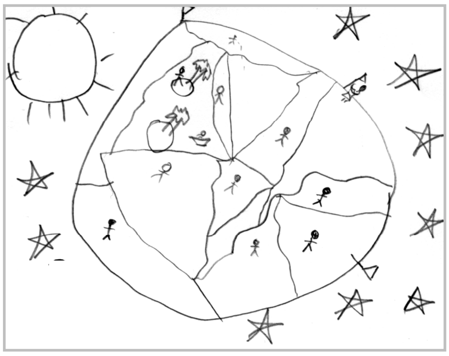


Figure 3. Sara; 7.1 years

Such ideas are in line with those identified by Nussbaum (1985) in Israeli students, aged between 8 and 14 years, about the Earth concept: the shape of the Earth, space and gravity. The author identified five notions that, from 1 to 5, correspond to a conceptual progress, from the most egocentric and primitive vision to the most de-centered and scientific one. Notions 1, 2 and 3 consider only the shape of the Earth and the nature of the sky/space:

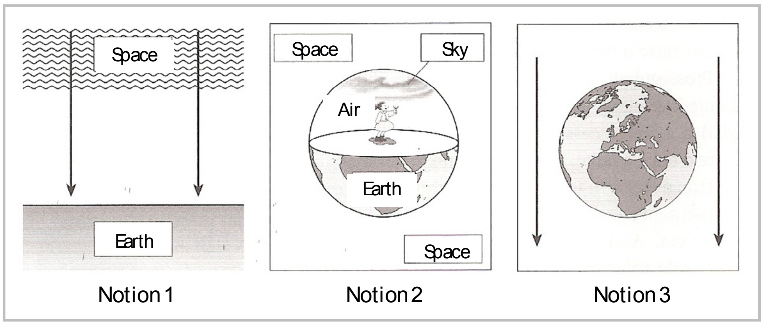


Figure 4. Notions about the shape of the Earth and the nature of the sky/space – adapted from Nussbaum (1985).

Our drawing categories (A, B, C) suggest a high level of parallelism with notions 1, 2 and 3 identified by Nussbaum (1985). The author found that approximately 80% of children aged eight years, in the 2nd grade, are distributed among notions 1 and 2.

A. Development of ideas about the shape of the Earth: from the flat model to the spherical model.

From the previous ideas that emerged in class it is intended that students, in small and large group, reflect on them and submit them, by way of discussion, to critical confrontation before colleagues and before the scientific model (photographs of Earth taken from space and earth globe), in order to develop a better knowledge and understanding of the shape of the Earth.

A_1. Communication and discussion of the meanings expressed in the drawings.
Passage from class diary:

> “I did the Earth, which is round” Francisca (6.2 years). "Francisca put things inside it" (Lionel; 6.9 years). "She put butterflies, the Sun and the clouds inside it" (Júlia; 6.9 years). Francisca's drawing falls into category B. Mafalda (6.4 years) clearly expresses the meaning of her drawing (category C): "I made the Earth round, seen from Mars. In space I drew the Sun and the stars, and here (on Earth) I drew the islands, the seas and people's lands (countries)". Gabriel (6.3 years) drew a flat Earth (Category A): "I did the clouds, a boy, the Sun..." He also says: "the Earth is under the boy." I ask him if the shape of the Earth in his drawing is flat. With a sad look, he says “yes”, but he recognizes that it is not flat: "No. I made it flat like this, but now I think it's round". Others also become aware that their drawings were not in agreement with the idea that the Earth is round: "Luís said that the Earth was round, but he didn't make it round" (Pedro; 6.3 years); "I didn't either, I thought it was flat" (Susana; 6.6 years), "Mine is not round" (Lionel).

Oral communication is aimed at sharing and confronting the meanings expressed in the three previous models, so as to subject them to critical review and promote their re-elaboration by others. In this process we observe the following:

i) the model of pseudo-sphericity of the Earth is strongly criticized with a hint of irony: "Oh look, she put things inside it; she put in butterflies, the Sun and clouds";

ii) the flat Earth model, confronted with the spherical model, generates some dissatisfaction and a critical attitude towards their own drawings and those of others, now rendering the notion of sphericity of the Earth far more plausible: “I made it flat like this, but now I think it's round"; "Luís said the Earth is round, but he didn't make it round".

A_2. The class faced with the spherical model of the Earth.

The students are unanimous in admitting that the spherical model (Category C) is the one that best represents the shape of the Earth. However, conflicting thoughts emerge between the most primitive meaning (flat Earth model) and the socially accepted meaning (spherical model), which demonstrate the difficulty in reconciling sphericity with the perception of the Earth's flat surface resulting from direct observation.

B. The Earth: development of a more comprehensive and richer meaning.

The students' initial ideas and ways of thinking are now subject to confrontation with empirical evidence: a photograph of the Earth taken from space and the Earth globe.

B_1. The photograph of the Earth and the Earth globe.

The comments about the observation of the photo begin by focusing on the Earth's spherical shape. However, color is a piece of information that stands out in their comments. The shade of blue is identified as the "seas" and, surprisingly, the white spots are associated with clouds and the ice at the poles. This inference requires a certain level of abstraction: in everyday life, students see the clouds from the Earth, but now they are identified from another perspective – that of someone who sees the Earth from a given location in space.

When the students' attention is focused on the globe, most of them know its name and prove to understand that the Earth globe represents a miniature of the Earth.

***B_2. What are the* similarities between the photograph of the Earth and the globe? Passage from the diary:**

> Children state without hesitation: "The shape is the same" (Júlia; 6.8 years), "they are both round" (Sara; 6.9 years); "it's the shape"; "it's the same" (other children). Some also recognize a few differences: "it's just that, in the picture, the blue is darker and there (globe) it's lighter" (Gabriel; 6.3 years). "This was taken from very far away" – argues Lionel (6.9 years), referring to the different shades of blue. When asked about the differences, they realize the following details: "The lands here (the globe) are neater and we can see them better" (Sara); "and here it seems they are more spread out, here (the photo) we can see the clouds" (Lionel); "it's as if it were space" – adds Sara.

The groups easily identify the similarity between the spherical shape of the Earth on the photo and on the globe. In that comparison, they also identify some differences between the reality of the Earth in the photograph and its representation on the globe model. In the photograph, the Earth is distinguished by: i) the darker shade of blue; ii) the lack of identification and contouring of the "lands", i.e. the continents and the countries contained therein; iii) the presence of clouds and the cosmic space around the Earth seen in the photograph.

***B_3. A renewed outlook on the drawings of the shape of the Earth*. Passage from the diary:**

> The children's attention is again focused on their drawings. Those who drew a flat Earth recognize once again that the Earth is round: "mine is not round" (Lionel; 6.9 years); "the Earth here is round (photo) and here it isn't" – says Gabriel, pointing at his drawing. I ask the class what they now know about the shape of the Earth. They assertively answer that "it is round" and Gabriel states that it looks like a ball. "It looks like a ball, but it's always spinning" says Júlia.

The idea of the Earth's sphericity is very mature in the class. For some students, the contrast between that knowledge and the less evolved ideas expressed in the drawings promotes greater awareness of their own learning. Only a few verbalize that increased awareness, but by doing it in a social context they are not only consolidating their ideas, but also promoting the intra-personal processes of assimilation of that learning in the other children. Verbalization favors the construction of more elaborate formulations of those same ideas, as in Júlia's case: "It looks like a ball, but it's always spinning".

C. Day and Night

C_1. What is day?

Within the small groups, students are encouraged to think about what day is[2]. There is a reference to *morning* as being daytime, an idea that can stem from the Portuguese morning greeting "bom dia" (good day). It is by opposition to that idea that they realize that the concept of day includes morning, noon and afternoon. After being questioned again, they now acknowledge that the elements Sun and light are subsumed in the

[2]It is intended that children develop the notion of day as the period of time during which a given location on Earth is illuminated by the Sun, i.e., the period between sunrise and sunset (natural day). That notion and the previously acquired meaning of the sphericity of the Earth will support the later development of the comprehension of the day-night alternation, as a result of the Earth's rotation movement in the presence of the Sun.

definition of day. It is said that the day begins with the sunrise. Some answers seem to contemplate the idea that the Sun is always visible during the day. This idea is subject to discussion. Students demonstrate an understanding of day as corresponding to the period of time during which the Sun illuminates a location on Earth, even if it is covered by clouds for whole days.

C_2. What is night?

In answers to questions about what night is, the following ideas emerged: i) they begin by making reference to darkness; ii) the darkness is a consequence of the absence of the Sun; iii) the absence of the Sun during the night is explained by some children with ideas of an animist nature – the need for the Sun to *"go away to rest or sleep"*[3]; and iv) in contrast, others claim that the Sun stays in space, in a different relative position, illuminating other parts of the planet. This is quite an evolved idea: it acknowledges the simultaneity of day and night in different locations, as a result of the Sun's relative position to those places.

D. Day and night in the Earth model – The Sun, without the Earth's rotation.

D_1. Identification by analogy of what the globe and the flashlight represent.
Passage from the diary:

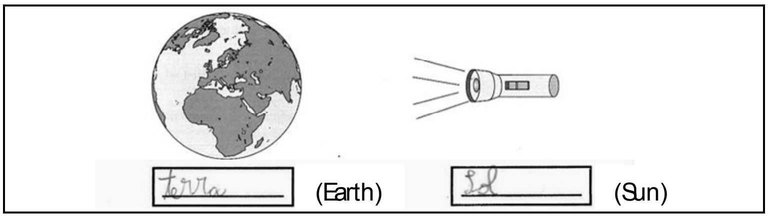
> The children's attention is again focused on the globe on the desk. They have no difficulty in recognizing that "it's the Earth in miniature". I show them a flashlight and some immediately associate it with the Sun: "It's the Sun" (Pedro; 6.2 years); "if you turn it on, it seems like the Sun" (Lionel; 6.9 years). Others also refer to the Sun and Sara adds: "it will give light to the Earth".

The students identify by analogy what each of the objects represents: the globe – "it's the Earth in miniature" – and the flashlight – "it's the Sun"; "if you turn it on, it seems like the Sun"; "it will give light to the Earth".

D_2. Elaboration of records.

The students proceed to the individual recording of what each object drawn on their record sheet intends to represent, i.e., the Earth and the Sun.

- I write in the boxes what each object represents.



Francisca (6.2 years)

Figure 5. Learning record: the globe-flashlight versus Earth-Sun analogy.

[3]It is interesting to verify that this idea has also been identified by other authors. For instance, Fleer (1997), in a study with Australian aboriginal children aged four to eight years, found a similar conception when, during the interview, she posed the following question to the children: why is it dark at night? Some answers explained the occurrence of night with the fact that the Sun went away or went to sleep, also relating night to the appearance of the stars and the Moon. However, in most of the answers given by children, according to the author, there is an animistic view of the Sun. According to the author, this conception tends to reflect expressions that children hear in everyday life ("the Sun has already gone away"), or the perception that the Sun descends over the horizon as night falls.

D_3. It is daytime on the part of the Earth that is lit by the Sun and nighttime on the part of the Earth that is not illuminated by the Sun. Passage from the diary:

> "Imagine the Earth is in the dark, what must we do to have daytime on the Earth?" – I ask. Everyone agrees that we have to turn on the flashlight and some state: "now it's daytime". "In what part of the Earth is it daytime?" – I ask. Children answer that it is in the part of the globe that is facing the flashlight: "on the part that is lit" (Bruno; 6.9 years); "it's the one that has the light" (Several). "And on the other side of the Earth, what will it be?" – I ask. Without hesitation some answer: it is nighttime. Mafalda looks pleased to see her previously expressed idea confirmed and states: "On the Sun's side it is daytime and on the other it's nighttime, the Sun does not get there and it goes dark".

D_4. Collective construction of a sentence about day and one about night.

In the discussion, the students show a good understanding of the notion that it is daytime on the part of the Earth facing the Sun, hence receiving light; and that it is nighttime on the part of the Earth that is opposite the Sun, which is in the dark. In class discussion, the following sentences are agreed upon about day and night, and are then written on the individual records: "it's daytime on the part of the Earth that is facing the Sun"; "it's nighttime on the part of the Earth that is not facing the Sun".

E. Earth Globe: from day to night in Portugal.

Our country is situated in the part of the globe illuminated by the flashlight (Sun) and therefore it is day in Portugal.

E_1.What must we do to have nighttime in Portugal? Passage from the class diary:

> The children answer: "We must try to turn the Sun to the other part" (Mafalda; 6.4 years); "turn the Sun" (Emanuel; 7.2 years); "the Sun has to go to the other side" (Júlia; 6.8 years); "we must turn the Sun" (Rui; 6.6 years); "we have to turn off the flashlight and put it on the other part" (Joana; 6.7 years); "when it is too sunny, it tilts a little" (João; 6.8 years). Among these answers, Gabriel and Bruno are the only students to contemplate the possibility of rotating the Earth: "We must rotate the Earth, the Earth is always spinning" (Gabriel; 6.3 years); "we must turn the Earth" (Bruno; 6.9 years). Mafalda argues: "we must put the Sun on the other part so that night comes over here" – the side where it was daytime. Gabriel does not accept Mafalda's or the other classmates' ideas and insists: "it's spinning the Earth, it's always spinning". I point out that there are two different ideas in the classroom: some say that the Sun (flashlight) must move around the Earth (globe) and others say that the Sun stands still and the Earth must spin around itself. Gabriel defends his idea before the class once more. For a few moments they remain in silence. "What do the others think?" – I ask. Bruno insists: "it's the Earth". Other children begin to support that idea: "it's the Earth that rotates, it's like a ball. Sometimes we kick it and it starts to spin" (Júlia; 6.8 years). But others again state that it is the Sun: "I think that it's the Sun that goes around the Earth" (Sara; 6.9 years); "me too" – says Rui. The students are divided. Asked to raise their fingers, ten children think that it is the Sun that must rotate around the Earth and eight think that it is the Earth that rotates around itself.

The answers are mostly supportive of moving the flashlight (the Sun) around the globe (Earth) so as to turn day into night in Portugal. This idea is spontaneously constructed by the students as a result of the observation of the changing position of the Sun throughout the day in relation to where they are – the apparent motion of the Sun. Only two children say that it is necessary to rotate the globe around itself – the Earth's rotation movement. These differing opinions generate intense discussion and promote participation by other children, who explicitly favor the notion of the Earth's rotation movement, as in Júlia's case. After the discussion, the class is divided (Earth's rotation: 10;55.6% vs. Movement of the Sun: 8; 44.4%).

F. Apparent movement of the Sun.

F_1. The illusion of movement of a static body when carried by another body in motion.

The students show an understanding of the illusion of backward movement of the "trees", "ground" and "houses" in relation to the automobile they are riding in. These family situations, evoked and recreated in class, enhance the comprehension of what is apparent movement. But will students be able to mobilize that knowledge and apply it to the apparent motion of the Sun? The answer appears in the following pedagogical approach.

F_2. The illusion of movement of the Sun as we are carried by the moving Earth. **Passage from the diary:**

> The children are encouraged to think about the following question: "so, is the Sun moving?" – I ask. Again, no one supports the idea of the Sun moving around the Earth. However, some evolve into a mixed idea: "It's the Sun and the Earth" (Pedro; 6.2 years), "I think the Earth moves, but the Sun moves everywhere" (João; 6.8 years). Others are now beginning to support the idea of the Earth's rotation: "it's the Earth" (Several) "it's the Earth that is always spinning" (Francisca; 6.2 years); "the Earth is a ball and it's always spinning and it seems like it's the Sun that is moving" (Mafalda; 6.4 years). Sara intervenes and states: "as the Earth moves slowly and we are here in our place, it seems like the Sun is moving. But it's not, it's the Earth". Sara is very excited at this point and continues to explain her idea: "because if the Earth did not spin, we would always be the same".

The idea that the movement of the Sun is only apparent gains momentum. In the process of social interaction, students evolve to meanings that reveal different levels of conceptual development:

- Some evolve into a construction that combines the spontaneous idea of the Sun moving around the Earth with the idea of the Earth's rotation, conveyed in the social context of the class: "It's the Sun and the Earth"; "I think the Earth moves, but the Sun moves everywhere".
- Others, who were previously partial to the movement of the Sun, now show an understanding of the Earth's rotation as the cause of the apparent motion of the Sun, as in the cases of Sara and Mafalda: "it's the Earth"; "it's the Earth that is always spinning"; "the Earth is a ball and it's always spinning and it seems like it's the Sun that moves".
- There are also those who develop a conceptual formulation of a higher level than the previous, with the generalization of the Earth's rotation movement as a cause of the day and night alternation: "as the Earth moves slowly and we are here in our place, it seems that the Sun is moving. But it's not; it's the Earth (...) because if the Earth did not spin, we would always be the same".

G. *The Earth's rotation movement in the presence of the Sun: the day and night alternation*

G_1. The day and night alternation in Portugal in the Earth– Sun model. **Passage from the diary:**

> "If it's the Earth that rotates, what must we do for it to be nighttime in Portugal?" – I ask. Without hesitation, the children recognize the need to rotate the Earth globe: "we have to rotate the Earth" – some say; "we turn the Earth" – others; "we must rotate the Earth" (Rui; 6.6 years). I slowly rotate the globe and ask them what had happened. They say that now it is nighttime in Portugal. "And on the other side of

> the Earth, what is it now?" – I ask. "It's daytime" – the children answer. When asked about what they must do for it to be daytime in Portugal again, the children answer that "they must rotate the Earth".

When applied to the Earth – Sun (globe – flashlight) model, the comprehension that the Earth revolves around itself promotes the acknowledgement that the alternation between day and night is a consequence of the Earth's rotation movement.

G_2. Being daytime in Portugal, could there be night in our country if the Earth stopped spinning?

The question raises the level of reflection in the class, translating into more elaborate answers. The students understand that, for it to be night, the globe, i.e. the Earth, must keep spinning until Portugal is again on the non-illuminated part: "it cannot be. If the Sun were always in Portugal and the Earth didn't spin, it would always be daytime and there would be no night. If it were nighttime and it wasn't spinning, it would always be nighttime" (Sara); "there couldn't, it would always be daytime" (Gabriel); "it would always be the same" (Mafalda).

G_3. Generalization of the day and night alternation. **Passage from the diary:**

> "So, why is there day and night?" - I ask. Gabriel states: "because the Earth is always spinning. It never stops". Other children intervene: "because the Earth is always spinning" (Sara); "if it didn't spin, there would only be day" (Lionel); "it's the Earth that rotates, and then there is day and night" (Bruno); "if it were daytime and the Earth stopped, there would never be night again. If it were night, when the Earth stopped, there would be no more day" (Mafalda). Rui adds: "only if the Earth moved again". The children who did not answer agreed with those answers and apparently understood that the succession of day and night was a result of the Earth's rotation.

The students' thoughts about the succession of days and nights have focused on the globe, based on the concrete situation of Portugal. When they are asked why there is day and night, the answers point towards the generalization of the idea of the alternation of day and night as a consequence of the Earth's rotation movement: "because the Earth is always spinning. It never stops"; "it's the Earth that rotates, and then there is day and night"; "if it were daytime and the Earth stopped, there would never be night again. If it were night, when the Earth stopped, there would be no more day".

H. Alternation of day and night: the Sun as a cause vs. consequence of the Earth's rotation movement

H_1. Confrontation between the intuitive ideas and the newly acquired learning.
Passage from the diary:

> At the beginning of the class the idea sprang up that, in the evening, the Sun would go "away" or that it "had gone to sleep". I remind them of those ideas and ask them what they have to say now. Sara begins by saying: "The Sun never sleeps, it never goes out, it never travels. It is always still, in one place." Other interventions follow: "It didn't go away" (Filipa; 6.2 years), "it's like a statue" (Gabriel; 6.3 years), "it never does anything, it is always still" (Lionel; 6.9 years); "it's as if it were glued to a wall" (Mafalda; 6.4 years). "If it didn't go away, then why don't we see it during the night?" – I ask. Sara answers again: "because the Earth is always rotating and the Sun stayed in the part where it was. It's on the other part" – says Mafalda, in the meantime. Júlia adds: "We are not the only ones who need the Sun, other people also need it. They also have plants to grow and they need the Sun"; "the Earth spins and then the other part gets the Sun and the part that had the Sun gets night" – says Joana (6.7 years).

This confrontation reveals remarkable progress in the comprehension of the apparent motion of the Sun. The conceptual level of development achieved by the students allows them a critical look at those ideas, and therefore the absence of motion of the Sun now makes more sense to them. The child who previously claimed that the Sun went to sleep at night is now the first to answer, in a critical tone: "The Sun never sleeps, it never goes out, it never travels. It is always still, in the same place". Other children now present solid arguments for the fact that we do not see the Sun during the night which rely on the Earth's rotation movement.

Analysis of the assessment results of the acquired learning

At the end of the class, students answered individually to a questionnaire with true or false items about the alternation of day and night. After three weeks, the students answered the same questionnaire. The following table shows the results obtained at the two moments.

Table 1. *Results Obtained in the Two Moments of Assessment of the Student Learning*

Items	Correct answers	
	M1 (%) after class	*M2 (%)* after 3 weeks
1. The Earth is round like a ball	16 (88.9%)	18 (100%)
2. The Sun stops shining during the night.	9 (50%)	10 (55.6%)
3. It is daytime in the part of the Earth that is facing the Sun.	14 (77.8%)	16 (88.9%)
4. When it is nighttime in Portugal, it is also nighttime in the whole world.	11 (61.1%)	14 (77.8%)
5. The Earth never stops spinning.	14 (77.8%)	14 (77.8%)
6. When it is daytime in Portugal, it is nighttime in other countries.	14 (77.8%)	16 (88.9%)
7. There is day and night because the Earth is always rotating.	12 (66.6%)	13 (72.2%)

These results suggest that the learning acquired by the students was meaningful because it is long-lasting, as opposed to memorized learning, which is soon forgotten (Coll & Martín, 2001).

Discussion

This study did not aim to assess the impact of Reflective Experimental Science Teaching (REST) on improving the understanding of the topics under study. However, the combination of the students' learning described in the class diary with the results obtained from the questionnaire suggests that the process of teaching and learning occurred in the classroom by means of the practice of REST may have had a highly positive effect on the students' learning. Thus, the combination of such data is indicative that the majority of students have developed a good learning about the Earth's shape and alternation day and night, as a result of the rotation of the Earth.

The construction of this learning started from the students' initial ideas, whose identification is an integral part of the teaching and learning process. Despite differences in method, it can be verified that ideas about the shape of the Earth are convergent with some ideas identified by other authors (Nussbaum, 1985; Vosniadou & Bewer, 1992) in children from other countries, cultures and similar age groups.

Through the interpretative content analysis of the class diary it is also possible to identify and characterize some of the processes that promote the quality of the students' thought and learning. The following stand out:

- the communication of ideas and ways of thinking to the class allows the students to contrast their own ways of thinking with the thoughts of others. In this process of verbalization, the students become more aware of their own ideas and the ideas of others. This increased awareness promotes, in some children, the need to restructure their ideas when confronted with other more plausible and consensual ones that appear in the social context of the class. Take, as an example, the communication to the class of the meanings implied in the drawings of the Earth;

- the discursive activity generated around the ideas that spring up in the classroom through the conjoint influence of their peers and the teacher's action improves the quality of those ideas, allows for the participation of other students and favors the development of more elaborate meanings;
- the students' more evolved meanings and the teacher's action direct and support the conjoint cognitive activity, allowing the slower students to elaborate new reconstructions and approximations to those meanings, which, after being verbalized in the social context of the class, are then shared by a growing number of students;
- the students' sociocognitive activity generated around experimental evidence introduces a considerable increase in the development of scientific meanings – the shape of the Earth; what is day and what is night; the alternation of day and night as a consequence of the Earth's rotation movement;
- the teacher's action, through continuous and recurrent reflective questioning (scaffolding), not only helps students to become aware and regulate their cognitive activity, but it also promotes their ability to escalate to progressively higher levels of cognition and learning;
- the introduction of significant and familiar analogies related to their day-to-day contexts facilitates the comprehension of particularly difficult situations, as was the case of the apparent motion of the Sun;
- the contrast between the learning acquired and their less evolved initial ideas triggers in the students a heightened awareness of their own learning – metacognitive knowledge;

The promotion of an experimental Science teaching practice in primary school has proved a difficult and complex task, as it requires that teachers assimilate and develop not only scientific knowledge, but also specific didactic knowledge about how to teach the subjects of specific curricular areas. The processes of teacher training, in our perspective, should be shaped by the practical and theoretical knowledge emerging from the holistic understanding of the teaching and learning processes, promoted and experienced in the classroom context. Thus, the present paper may prove a valuable resource for the initial and continuous teacher training process in order to endow these professionals with a specific knowledge on how to elicit and promote, within the classroom context, identical processes in approaching the curricular topic on the shape of the Earth and the alternation of day and night.

Finally, it can be argued that children are able to overcome complex cognitive challenges when they are approached in a collaborative context of stimulation and freedom of expression of their thoughts. Thus, the interaction with other more developed children, or with the teacher, and the domain of language promote higher levels of learning, which is an important factor for the development of thought (Vygotsky, 1978).

● ● ●

Paulo VARELA received his Ph.D. in Child Studies, in the field of Science Education, from the University of Minho in 2010. He is an Assistant Professor at the Institute of Education of The Department of Integrated Studies of Literacy, Didactics and Supervision at the same university since 1998. His research interests focus on the study of in-class experimental Science teaching and learning processes with children.

References

Aleixandre, J. M. P. (2003). Comunicación y lenguaje. In J. Aleixandre (Ed.), *Enseñar ciencias* (pp. 55-71). Barcelona: Editorial Graó.

Appleton, K. (2003). How Do Beginning Primary School Teachers Cope with Science? Toward an Understanding of Science Teaching Practice. *Research in Science Education, 33*, 1-25.

Blown, E. J. & Bryce, T. G. K. (2007). Knowledge Restructuring in the Development of Children's Cosmologies. *International Journal of Science Education*, *28*(12), 1411-1462.

Brand, B. R. & Moore, S. J. (2011). Enhancing Teachers' Application of Inquiry□Based Strategies Using a Constructivist Sociocultural Professional Development Model. *International Journal of Science Education*, *33*(7), 889-913

Candela, A. (1999). *Ciencia en el aula. Los alumnos entre la argumentación y el consenso.* México: Editorial Paidós.

Carrasco, J. B. & Hernández, J. F. C. (2000). *Aprendo a Investigar en Educación*. Madrid: Ediciones Rialp.

Català, M. & Vilà, N. (2002). Las funciones lingüísticas en el proceso de adquisición de los conocimientos científicos. In M. Catalá, et. al, (Eds.). *Las ciencias en la escuela. Teorías y prácticas* (pp. 89-103). Barcelona: Editorial Gráo.

Charpack, G. (2005). *As Ciências na Escola Primária: Uma Proposta de Acção*. Mem Martins: Editorial Inquérito.

Chin, C. (2006). Classroom Interaction in Science: Teacher questioning and feedback to students' responses. *International Journal of Science Education, 28*(11), 1315-1346.

Cleary, T. J. & Zimmerman, B. J. (2004). Self-regulation empowerment program: a school-based program to enhance self-regulated and self-motivated cycles of student learning. *Psychology in the Schools*, *41*(5), 537-550.

Coll, C. & Martín, E., (2001). A avaliação da aprendizagem no currículo escolar: uma perspectiva construtivista. In C. Coll, (Eds.), *O construtivismo na sala de aula. Novas perspectivas para a acção pedagógica* (pp. 196-221). Porto: Edições ASA.

Costa, F. (2006). As actividades práticas na educação em ciência: uma oportunidade perdida? In G. Hamido; H. Luís; M. C. Roldão & R. Marques (Eds.), *Transversalidade em educação e em saúde* (pp. 31-38). Porto: Porto Editora.

Denzin, N. K. (1989). *Interpretive interactionism*. Newbury Park, CA: Sage.

Domínguez, M. A. & Stipcich, M. S. (2009). Buscando indicadores de la negociación de significados en clases de Ciencias Naturales. *Revista Electrónica de Enseñanza de las Ciencias, 8*(2), 539-551. Retrieved from http://www.saum.uvigo.es/reec/

Erickson, F. (1986). Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 119-161). New York, NY: Macmillan.

Eurydice (2011). *Science Education in Europe: national Policies, Practices and Research.* Brussels. Education, Audiovisual and Culture Executive Agency - P9 Eurydice

Fasick, F. A. (2001). Some uses of untranscribed tape recordings in survey research. *Public Opinion Quarterly*, 41, 549-552.

Fleer, M. (1997). A Cross-cultural Study of Rural Australian Aboriginal Children's Understandings of Night and Day. *Research in Science Education, 27* (1), 101-116.

Georghiades, P. (2006). The Role of Metacognitive Activities in the Contextual Use of Primary Pupils' Conceptions of Science. *Research in Science Education*, 36, 29-49.

González, S. & Escudero, C. (2007). En busca de la autonomía a través de las actividades de cognición y de metacognición en Ciencias. *Revista Electrónica de Enseñanza de las Ciencias, 6*(2), 310-330. Retrieved from http://www.saum.uvigo.es/reec/

Guba, E. G. & LIncoln, Y. S. (1989). *Fourth Generation Evaluation.* Thousand Oaks, CA: Sage Publications.

Guba, E. G. & Lincoln, Y. S. (2000). Competing Paradigms in Qualitative Research. In N. Denzin, & Y. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105-117). Thousand Oaks, CA: Sage.

Harlen, W. & Qualter, A. (2005). *The Teaching of Science in Primary Schools*. London: David Fulton Publishers.

Harlen, W. (1999). *Effective Teaching of Science. A Review of Research*. Edinburgh: The Scottish Council for Research in Education (SCRE).

Harlen, W. (2007). *Enseñanza y aprendizaje de las ciencias*. (2ª ed. atualizada). Madrid: Ediciones Morata.

Harlen, W. (2010). *Implementing inquiry-based learning in science education (IBSE).* Fibonacci Project Presentation on September 22nd 2010. Retrieved from http://fibonacci.uni-bayreuth.de/index.php?id=203

Henao, B. L. & Stipcich, M. S. (2008). Educación en ciencias y argumentación: la perspectiva de Toulmin como posible respuesta a las demandas y desafíos contemporáneos para la enseñanza de las Ciencias Experimentales. *Revista Electrónica de Enseñanza de las Ciencias, 7*(1). Retrieved from http://www.saum.uvigo.es/reec/

Huber, R. A. & Moore, C. J. (2001).A Model for Extending Hands-On Science to be Inquiry-Based. *School Science and Mathematics*, *101*(1), 32-42.

Ibáñez, V. E. & Alemany, I. G. (2005). La interacción y la regulación de los procesos de enseñanza-aprendizaje en la clase de ciencias: análisis de una experiencia. *Enseñanza de Las Ciencias*, *23*(1), 97-110.

Jorgenson, O. (2005). What K-8 Principals Should Know About Hands-On Science. *Principal-Effective Intervention- Special Section, 85*(2), 49-52.

Kuhn, D., E. Amsel & M. O'Loughlin (1988). *The development of scientific thinking skills.* Orlando, FL: Academic Press.

Larkin, S. (2006). Collaborative Group Work and Individual Development of Metacognition in the Early Years. *Research in Science Education*, *36,* 7-27.

Lee, O.; Hart, J. E.; Cuevas, P.; Enders, C. (2004). Professional Development in Inquiry-Based Science for Elementary Teachers of Diverse Student Groups. *Journal of Research In Science Teaching, 41*(10), 1021-1043.

Levinson, R. (2002). *O professor, o aluno e o livro didático oficial de Ciências: será que deveríamos renunciar ao livro escolar?* Brasília: M. A. Marfan. Retrieved from http://portal.mec.gov.br/seb/arquivos/pdf/vol2c.pdf

MacLean, L. M., Meyer, M., & Estable, A. (2004). Improving accuracy of transcripts in qualitative research. *Qualitative Health Research*, *14*(1), 113-123.

Maloney, J. & Simon, S. (2006). Mapping Children's Discussions of Evidence in Science to Assess Collaboration and Argumentation. *International Journal of Science Education, 28*(15), 1817-1841.

Martínez, j. M. O. & Díaz, J. A. A. (2005) La enseñaza de las ciencias en primaria y secundaria hoy. Algunas propuestas de futuro. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias, 2*(2), 241-250.

Matthews, M. R. (2000). Construtivismo e o Ensino de Ciências: Uma Avaliação. *Cad. Cat. Ens. Fís.*, *17*(3), 270-294.

Mercer, N. & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach.* London: Taylor and Francis.

Ministry of Education (2004). *Organização Curricular e Programas Ensino Básico – 1º Ciclo*. Mem Martins. Departamento de Educação Básica.

Miras, M. (2001). Um ponto de partida para a aprendizagem de novos conteúdos: os conhecimentos prévios. In C. Coll; E. Martin; T. Mauri; M. Miras; J. Onrubia; I. Solé & A. Zabala (Eds.). *O construtivismo na sala de aula. Novas perspectivas para a acção pedagógica* (pp. 54-73). Porto: Edições ASA.

Molenaar, I.; van Boxtel, C. A. & Sleegers, P. J. C. (2011). Metacognitive scaffolding in an innovative learning arrangement. *Instructional Science*, *39,* 785-803.

Naylor, S.; Keogh, B.; Downing, B. (2007). Argumentation and Primary Science. *Research in Science Education*, *37*, 17-39.

Newman, J. M. (2000). Action research: A brief overview. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 1*(1). Retrieved from http://qualitative-research.net/fqs.

Nussbaum, J. (1985). The Earth as a Cosmic Body. In R. Driver, E. Guesne & A. Tiberghien (Eds.), *Children´s ideas in science* (pp. 170-192). Philadelphia: Open University Press.

Özsoy, S. (2012). Is the Earth Flat or Round? Primary School Children's Understandings of the Planet Earth: The Case of Turkish Children. *International Electronic Journal of Elementary Education, 4(2),* 407-415.

Palincsar, A. S. & Herrenkohl, L. R. (2002). Designing collaborative learning contexts. *Theory Into Practice, 41*, 26-32.

Pérez Gómez, A. I. (2005). *Comprender y transformar la enseñanza* (pp.17-137). Madrid: Ediciones Morata,S. L..

Porlán, R. (1998). *Constructivismo y escuela: hacia un modelo de enseñanza-aprendizaje basado en la investigación*. Sevilha: Díada Editora S. L..

Ratner, C. (2002). Subjectivity and Objectivity in Qualitative Methodology. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *3* (3). Retrieved from http://www.qualitative-research.net/fqs/fqs-eng.htm

Rivard, L. P. (2004). Are Language-Based Activities in Science Effective for All Students, Including Low Achievers? *Science Education, 88*, 420-442.

Rojas-Drummond, S. & Mercer, N. (2003). Scaffolding the development of effective collaboration and learning. *International Journal of educacional Research, 39*, 99-111.

Roldão, M. C. (2003). *Gestão do Currículo e Avaliação de Competências. As questões dos professores*. Barcarena: Editorial Presença.

Sá, J. & Varela, P. (2004). *Crianças Aprendem a Pensar Ciências: uma abordagem interdisciplinar.* Porto: Porto Editora.

Sá, J. & Varela, P. (2007). *Das Ciências Experimentais à Literacia: Uma proposta didática para o 1º ciclo*. Porto: Porto Editora.

Sá, J. (2002a). *Renovar as Práticos no 1º Ciclo Pela Via das Ciências da Natureza*. Porto: Porto Editora.

Sá, J. (2002b). Diary Writing: An Interpretative Research Method of Teaching and Learning. *Educational Research and Evaluation*, *8*(2), 149-168.

Siegal, M., Butterworth, G., & Newcombe, P. (2004). Culture and children's cosmology. *Developmental Science, 7(*3) 308-324

Varela, P. (2012). *Experimental Science Teaching in Primary School: Reflective Construction of Meanings and Promotion of Transversal Skills.* Saarbrücken, Germany: Lap Lambert Academic Publishing.

Vosniadou, S. & Brewer, W. F. (1992). Mental Models of the Earth: A Study of Conceptual Change in Childhood. *Cognitive Psychology, 24*, 535-585.

Vosniadou, S. (1991). Designing Curricula for conceptual restructuring: Lessons from the study of knowledge acquisition in astronomy. *Journal Curriculum Studies*, *3*(23), 219-237.

Vosniadou, S., & Brewer, W. F. (1994). Mental mod s of the day/night cycle. *Cognitive Science*, 18, 123-183.

Vosniadou, S.; Skopeliti, I.; Ikospentaki, K. (2004). Modes of knowing and ways of reasoning in elementary astronomy. *Cognitive Development*, *19*, 203-222.

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. London: Havard University Press.

Vygotsky, L. S. (1987). *Pensamento e Linguagem*. São Paulo: Martins Fontes Editora.

White, R. T. (1994). Dimensions of Content. In P. Fensham, R. Gunstone & R. White (Eds.), *The Content of Science. A Construtctivist Approach to its Teaching and Learning* (pp. 255-262). London: Falmer Press.

Zabalza, M. A. (2004). *Diarios de clase: un instrumento de investigación*. Madrid: Narcea.

Zohar, A. (2006).El pensamiento de orden superior en las clases de ciencias: objetives, medios e resultados de investigación. *Enseñanza de Las Ciencias, 24*(2), 157-172.

www.iejee.com

This page is intentionally left blank

International Electronic Journal of Elementary Education, 2012, 5(1), 27-46.

Parent-child conversations about evolution in the context of an interactive museum display

Andrew SHTULMAN*
Occidental College, Los Angeles, USA

Isabel CHECA
Occidental College, Los Angeles, USA

Received: 14 August 2012 / Revised: 16 October 2012 / Accepted: 16 October 2012

Abstract

The theory of evolution by natural selection has revolutionized the biological sciences yet remains confusing and controversial to the public at large. This study explored how a particular segment of the public – visitors to a natural history museum – reason about evolution in the context of an interactive cladogram, or evolutionary tree. The participants were 49 children aged four to twelve and one accompanying parent. Together, they completed five activities using a touch-screen display of the phylogenetic relations among the 19 orders of mammals. Across activities, participants revealed similar misconceptions to those revealed by college undergraduates in previous studies. However, the frequency of those misconceptions was attenuated by the level of parental engagement, particularly the frequency of turn-taking between parents and children. Overall, these findings suggest that evolutionary reasoning may be improved by the kinds of collaborative discussions fostered by interactive museum displays, so long as the affordances of those displays encourage multi-user interactions.

Keywords: Conceptual Development, Evolution Understanding, Parent-Child Conversation, Informal Learning Environments, Science Education.

Introduction

In 1996, the U.S. National Academy of Sciences identified evolution as one of five "unifying concepts and processes" that should be taught in all grades, from K through 12 (National Research Council, 1996). The rationale behind this recommendation was that "evolution is the central organizing principle that biologists use to understand the world. To teach biology without explaining evolution deprives students of a powerful concept that brings great order and coherence to our understanding of life" (p. 3, National Research Council, 1998). Despite the force of this recommendation, many schools continue to fail to teach evolution in *any* grade (Griffith & Brem, 2004), and many Americans continue to deny the very fact of evolution, particularly human evolution (Miller, Scott, & Okamoto, 2006; Newport, 2010).

* Andrew Shtulman, Department of Psychology, Occidental College, Los Angeles, CA 90041. Phone: 323-259-2633. E-mail: shtulmanoxy.edu

ISSN:1307-9298

One of the problems underlying, if not exacerbating, public denial of evolution is that most people fail to understand what evolution actually is and how evolution actually occurs (Shtulman & Calabi, 2012). Recent research has found that individuals of all ages and educational backgrounds tend to construe evolution as a kind of holistic transformation, by which organisms are predisposed to produce offspring more adapted to the environment than they were at birth (Shtulman, 2006; Shtulman & Calabi, in press). On this erroneous theory, the sole mechanism behind species adaptation is *need*: if a species *needs* to adapt, then it *will* adapt (Bishop & Anderson, 1990; Brumby, 1984; Southerland, Abrams, Cummins, & Anzelmo, 2001). Selection plays no causal role on this view, which, as a result, bears more resemblance to pre-Darwinian theories of evolution than post-Darwinian ones (Mayr, 1982). Developmentally, this view appears to be an outgrowth of an early emerging "essentialist" construal of biological kinds, in which an organism's outward appearance and behavior is determined by some kind of internal force, or "essence," conferred from parent to child at birth (Gelman, 2003; Hatano & Inagaki, 1994; Solomon & Zaitchik, 2012). While essentialism may be useful for reasoning about the properties of individual organisms (e.g., Gelman & Coley, 1990; Waxman, Medin, & Ross, 2007), it has been shown to be detrimental for reasoning about the properties of entire species, as it leads students to overvalue variation *between* species and undervalue variation *within* a species (Nettle, 2010; Shtulman & Schulz, 2008). As a consequence, students have difficulty understanding mechanisms of change that operate specifically over the variation within a species, namely, natural selection.

To date, numerous studies have documented the nature of students' essentialist, need-based views of evolution (for reviews, see Gregory, 2009; Sinatra, Brem, & Evans, 2008). Less attention, however, has been paid to the ways in which alternative views of evolution manifest themselves in everyday discourse and everyday interactions. The present study attempted to explore this issue in the context of parent-child conversations at a natural history museum. In particular, we sought to elicit conversations about the phenomena represented by one of the most canonical depictions of evolutionary change: the *cladogram*. More commonly referred to as an "evolutionary tree," cladograms are branching diagrams that depict patterns of common ancestry among three or more groups of organisms, or *taxa*. A sample cladogram, depicting the evolutionary relations among primates, can be seen in Figure 1.

Even though cladograms are ubiquitous in biology textbooks (Catley & Novick, 2008) and natural history museums (Torrens & Barahona, 2012), they are notoriously difficult to interpret, partly because they contain unfamiliar notational conventions (Novick & Catley, 2007) and partly because they are amenable to inaccurate, essentialist interpretations of evolutionary change (Shtulman, 2006). Drawing on recent empirical investigations of "tree thinking" in introductory biology students, Gregory (2008) outlined 10 such misconceptions:

1. Interpreting taxa on one side of a cladogram as "higher" or "lower" than those on the other side.

2. Interpreting the longest line in a cladogram as the "main line" from which other taxa have deviated or side-tracked.

3. Inferring information about relatedness from the ordering of a cladogram's terminal nodes rather than from its branches.

4. Interpreting cladograms as representations of morphological similarity rather than common ancestry.

5. Interpreting some taxa in a cladogram as the ancestors of other taxa rather than interpreting all taxa as "siblings" or "cousins."

6. Interpreting the length of the branches in a cladogram as measures of evolutionary change (or lack thereof).

7. Inferring that the taxa on one side of a cladogram appeared, in their current form, earlier than those on the other side of the cladogram.

8. Interpreting the length of the longest branch of the cladogram as a measure of time.

9. Interpreting the number of intervening nodes between two taxa as a direct measure of their relatedness.

10. Interpreting internal nodes in a cladogram as representing precise moments of speciation, with little to no change occurring before or after that point in time.

Many of the misconceptions are overlapping (e.g., 1 and 7) and some are mutually exclusive (e.g., 6 and 10), but all represent illegitimate inferences from the information at hand.

As an illustration, consider the cladogram depicted in Figure 1. The only information this diagram provides is information about common ancestry – namely, that humans share a common ancestor with chimpanzees more recently than with any of the other primates, that humans and chimpanzees share a common ancestor with gorillas more recently than with any of the other primates, and so forth. Nevertheless, most people are prone to infer that:

1. Humans are more highly evolved than other primates.

2. Human evolution represents the "main line" of evolution, whereas the evolution of other primates represents sidetracks from this main line.

3. Chimpanzees are related to gorillas more closely than humans are related to gorillas (because the nodes of the former pair are adjacent but the nodes of the latter pair are not).

4. Humans are most similar to chimpanzees (the closest node to humans) and least similar to new world monkeys (the farthest node from humans).

5. Each primate is the descendent of the primate on its left and the ancestor of the primate on its right.

6. Humans have undergone the more evolutionary change than other primates (because their connection to the root node is longest).

7. Each primate appeared, in its current form, earlier than the primate to its right.

8. Each primate is older than the primate to its right.

9. Humans are related to orangutans less closely than chimpanzees are related to orangutans (because of differences in the number of intervening nodes).

10. Chimpanzees came into being instantaneously at the point denoted by the rightmost node.

These inferences are not just logically unwarranted; they are also empirically incorrect. Indeed, inferences like 4 and 8 are not even meaningful on a scientific understanding of speciation, let alone correct or incorrect.

Misconceptions of this nature have been documented both in the classroom (Baum, Smith, & Donovan, 2005; Meir, Perry, Herron, & Kingsolver, 2007) and in carefully controlled

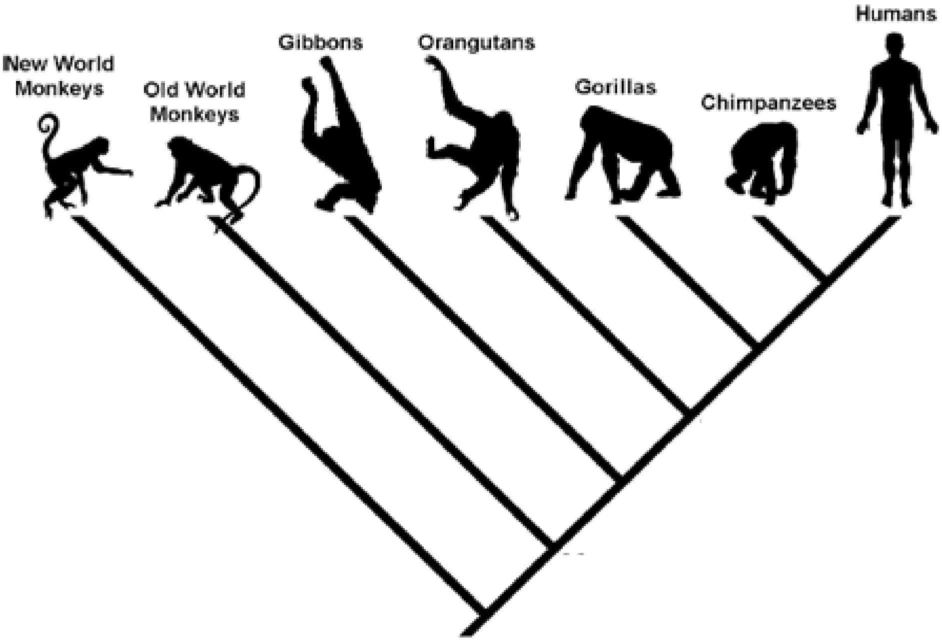


Figure 1. Cladogram depicting the phylogenetic relations among seven primates (adapted from Gregory, 2008).

laboratory studies (Catley, Novick, & Shade, 2010; Novick & Catley, 2007; Novick, Shade, & Catley, 2010). However, all such studies have involved college undergraduates, and no studies, to our knowledge, have explored the prevalence of macroevolutionary misconceptions in a non-college population (though see Berti, Toneatti, & Rosati, 2010, and Samarapungavan & Wiers, 1997, for research on children's misconceptions about other aspects of evolution). To address this gap in the literature, we explored the prevalence of macroevolutionary misconceptions among preschool- and elementary-school-aged children visiting a natural history museum with their parents. This sample served to broaden the scope of inquiry not only in terms of age but also in terms of context, as each child was interviewed as part of a dyad with his or her parent. In other words, children's reasoning about evolution was explored within the naturalistic context of a parent-child conversation.

Previous research on parent-child conversation suggests that conversations of this type can be a double-edged sword, with some aspects of the conversation scaffolding learning and other aspects obscuring, or even obstructing, learning. For instance, Jipson and Callanan (2003) found that parents of preschool-aged children typically use the word "grow" in a literal sense, to refer to biological changes in size (e.g., "the mushroom grew taller"), but also occasionally use the word in a metaphorical sense, to refer to non-biological changes in size (e.g., "the rock grew bigger"), yielding a linguistic signal that is reliable yet noisy nonetheless. In a similar vein, Rigney and Callanan (2011) found that parents at a marine science center ascribed biological properties to typical animals (e.g., sharks) no more often than they ascribed biological properties to atypical animals (e.g., anemones), potentially reinforcing the inclusion of atypical animals in the category of *living things*, a notoriously difficult concept to acquire (see, e.g., Anggoro, Waxman, & Medin, 2008). However, these same parents ascribed intentional states, like beliefs and desires, to both typical and atypical animals significantly more often than their children did, thus modeling a scientifically inappropriate form of reasoning. In short, parental input can serve as a source of accurate reasoning but by no means guarantees accurate reasoning (see also Gleason & Schauble, 2000; Gunderson & Levine, 2011).

In the present study, we explored how parents converse with their children about evolution, which we anticipated would be a difficult topic for both parties. We elicited these conversations by recruiting parent-child dyads from the floor of the Los Angeles Natural History Museum to complete a series of activities centered around an interactive cladogram. Two questions were of primary interest. First, how well do parent-child dyads interpret the information contained in cladograms, given that they are perhaps the most prevalent representation of evolutionary change in modern culture (Torrens & Barahona, 2012) yet are largely misunderstood by most biology students (Catley et al., 2010)? Research by Evans et al. (2010) and Spiegel et al. (2012) suggests that museum visitors hold a variety of preconceptions about *micro*-evolutionary change, some consistent with the principle of natural selection (e.g., need-based reasoning) and some inconsistent with it (e.g., creationist reasoning). Still, it remains an open question as to how museum visitors interpret displays representing *macro*-evolutionary phenomena, such as speciation, extinction, and common descent.

Second, what factors influence the accuracy of dyads' reasoning? Three candidate factors were identified from prior research on shared scientific thinking: (a) the child's gender, (b) the child's age, and (c) the dyad's overall style of interaction. In terms of gender, we predicted that dyads with male children would outperform dyads with female children, owing to the finding that parents are more likely to explain scientific phenomena to their sons than to their daughters (Crowley, Callanan, Tenenbaum, & Allen, 2001; Diamond, 1994) and might thus devote more attention to their sons in the activities at hand. In terms of age, we predicted that dyads with older children would outperform dyads with younger children, owing to the finding that older children are generally more familiar with evolutionary ideas than younger children (Berti et al., 2010; Legare, Lane, & Evans, in press; Spiegel et al., 2012) and might thus comprehend the purpose of the activities more thoroughly. Finally, in terms of interaction style, we predicted that dyads exhibiting higher levels of collaboration would outperform dyads exhibiting lower levels of collaboration, owing to the finding that parents at a science museum tend to hone their children's exploration of the exhibits in conceptually constructive ways (Crowley, Callanan, Jipson, Galco, Topping, & Shrager, 2001; Tare, French, Frazier, Diamond, & Evans, 2011), even if parents do occasionally provide conflicting or confusing input.

To preview our results, we found that parent-child dyads espoused the same kinds of misconceptions documented among college-level biology students. However, the frequency of such misconceptions varied by dyad, with dyads exhibiting low levels of collaboration espousing more misconceptions than those exhibiting higher levels of collaboration. This effect of dyad interaction was larger and more consistent than any of the other effects documented and thus has potentially important implications for both evolution education and informal science learning.

Method

Participants

The participants were 49 parent-child dyads recruited from the "Age of Mammals" exhibit at the Los Angeles Natural History Museum (see Figure 2A). We chose the Age of Mammals exhibit because it is thoroughly grounded in evolutionary findings and evolutionary principles and thus served as an ideal venue for eliciting conversations about evolution. All parents accompanying children between the ages of four and twelve were approached by

A

B

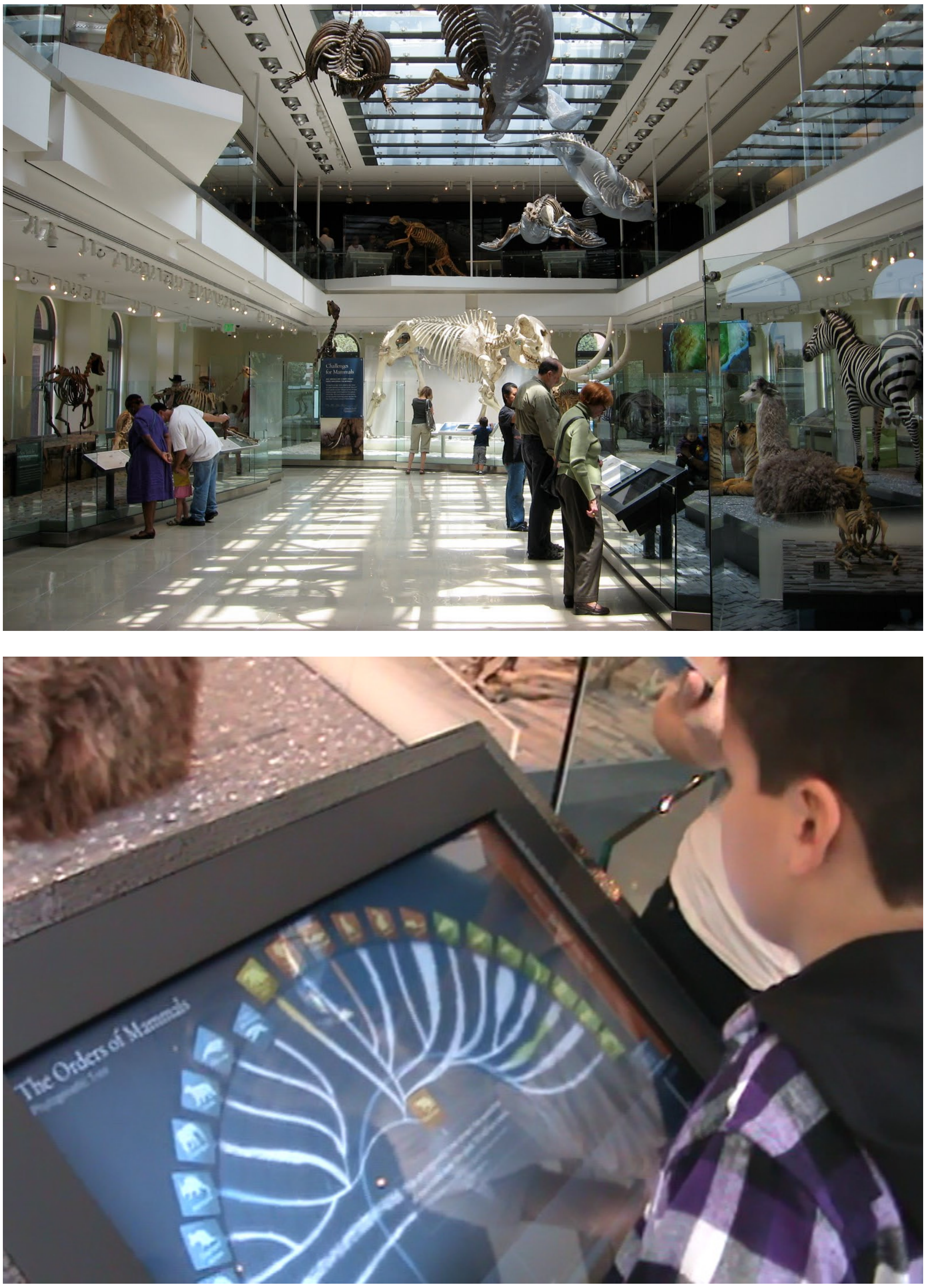


Figure 2. (A) The "Age of Mammals" exhibit at the Los Angeles Natural History Museum; (B) the interactive cladogram used to elicit participants' evolutionary reasoning.

the research team as they entered the exhibit and were invited to participate in a study. They were informed that the study consisted of how parents and children communicate about complex biological concepts, like evolution and common descent. Fifty-five parent-child dyads consented to participate, but only 49 completed the entire study. The ages of the participating children were evenly distributed across the age range sampled; half were between the ages of four and eight (M = 6.8, SD = 1.3, n = 24) and half were between the ages of nine and twelve (M = 10.3, SD = 1.0, n = 25). As for gender, 46% of the children were female, and 46% of the accompanying parents were female. Visitors who took part in the study did not receive any monetary compensation for their participation.

Procedure

The Age of Mammals exhibit features a total of 240 biological specimens, including both articulated skeletons of extinct species and taxidermied specimens of extant species. The specimens are organized into collections intended to illustrate three key principles: "Continents move. Climates change. Mammals evolve." These principles are also illustrated in the form of various touch-screen, computerized displays. One such display, an interactive cladogram, comprised the focus of the present study (see Figure 2B). All parent-child dyads who consented to participate were taken directly to the interactive cladogram, briefly familiarized with its features and functions, and then led through a series of five activities exploring the information contained within.

Technically, the display was not a "cladogram," in the biologist's sense of the word, because it attempted to represent more than just patterns of common ancestry; it also represented the times that taxa diverged and whether the taxa are extant or extinct. Nevertheless, we will refer to the display as a cladogram because its *primary* function was to depict patterns of common ancestry. Five display-based activities were designed to probe participants' understanding of macroevolution and the ways in which macroevolutionary relations are represented in cladograms. These activities, which were derived from previous research on students' macroevolutionary misconceptions (e.g., Catley et al., 2010; Gregory, 2008), are described below in relation to participants' actual responses. All responses were video recorded and transcribed at a later date. At the completion of the five-activity interview, participants were encouraged to explore other areas of the exhibit, but their conversations at those other areas were not recorded.

Coding

Measures of engagement. Each interview was transcribed by two independent coders, the first producing a written record of all utterances and actions and the second editing or embellishing that record to account for details the first coder appeared to have missed. The coders then summed the number of distinct utterances and actions for each dyad member and each activity, resulting in 490 engagement scores (5 scores per participant for 98 participants). Any self-contained thought, question, or response was coded as a distinct utterance, even if that thought did not take the form of a complete sentence (e.g., "yes" or "OK" were counted as distinct utterances). Likewise, any attempt to deliberately manipulate the display (via tapping or scrolling), or manipulate a partner's view of the display (via pointing or waving), was coded as a distinct action. Extended sequences of dialogue or activity were broken into distinct utterances and actions, with the exception of experimenter requested actions (e.g., "point to the monkey, the tree shrew, and the flying lemur"), which were counted simply as one action so as not to inflate engagement scores for some activities relative to others. Coders agreed on their tabulation of distinct utterances and actions 93% of the time, and all disagreements were resolved through discussion.

Measures of accuracy. Participants' responses to each activity were assigned a score that ranged from -1 to +2. Responses that revealed a positive misconception about the material at hand received a score of -1. Responses that were vague, ambiguous, or equivocal (including "don't know" responses) received a score of 0. Responses that revealed a partial understanding of the material at hand received a score of +1. And responses that revealed a full understanding of the material at hand received a score of +2. What constituted a correct response or an incorrect response is discussed below in relation to the corresponding activity. It should be noted that each dyad received a single score per activity, rather than separate scores for each dyad member, as almost all dyads offered a single response by the conclusion of almost all activities. Two coders independently assigned scores to all 245 responses (5 responses per dyad for 49 dyads). Overall agreement was 82% (Cohen's κ = .76), and all disagreements were resolved through discussion.

Results

How well do parent-child dyads reason about evolutionary phenomena?

Below we describe each activity used to elicit participants' evolutionary reasoning and the nature of their responses, followed by an analysis of participants' engagement with each activity and how that engagement was related to response accuracy per activity (summed across dyads). We address the question of whether, and how, engagement influenced response accuracy per dyad (summed across activities) in the following section.

Activity 1: Ordering. The introductory screen of the interactive cladogram featured all 19 orders of mammals, arranged in a semicircle with primates in the center (see Figure 2B). The first activity was designed to elicit participants' beliefs about the necessity, and potential flexibility, of this particular ordering. As mentioned previously, the ordering of the taxa in a cladogram is, to a large extent, arbitrary. While taxa that share a most recent common ancestor must be adjacent (e.g., chimpanzees and humans), their ordering relative to one another is arbitrary (i.e., chimpanzees can be on the left and humans on the right or humans can be on the left and chimpanzees on the right). The representation of nested taxa is governed by the same constraint, meaning that entire groups of taxa can be swapped with one another so long as the underlying branching relations are preserved. Thus, any one taxon could appear at any point in the row of terminal nodes, and any cluster of taxa could be reordered in many different ways.

We attempted to elicit participants' understanding of the ordinal relations among taxa by asking them to locate three particular taxa – the monkey, the tree shrew, and the flying lemur – and reflect on the ordering of those taxa. Specifically, we asked, "Does it matter that the flying lemur is on the left, the tree shrew is in the middle, and the monkey is on the right? Or could they be reordered so that the monkey is on the left, the flying lemur is in the middle, and the tree shrew is on the right? Why or why not?"

Because lemurs and shrews share a common ancestor with each other more recently than either shares with monkeys (as depicted by the relevant branching relations), the only constraint on ordering was that the lemur and the shrew had to be adjacent, which was true of the hypothetical ordering we asked participants to consider. Nevertheless, 17 dyads claimed that the taxa could not be reordered, justifying their judgment with an affirmation that ordering matters (scored -1). Most dyads (n = 23) were unsure whether or not the taxa could be reordered (scored 0), and only a few dyads (n = 9) claimed the taxa could be reordered but were unable to provide an adequate justification for their judgment (scored +1). None of the dyads provided both a correct judgment (that the taxa could be reordered) and a correct justification (that only the branching relations matter), and thus none received a score of +2. On the contrary, dyads that received a score of 0 or 1 seemed genuinely unsure of whether, and how, the ordinal properties of the display reflected information about the species' evolutionary origins.

Activity 2: Branching. Participants' understanding of the branching relations in a cladogram was elicited more directly in the second activity. Participants were asked to locate two non-placental mammals – the kangaroo and the platypus – and to read about their features in a pop-up window that appeared upon touching each. We chose non-placental mammals as our target taxa for this activity because their divergence from the other mammals occurred earliest (around 170 million years ago for egg-laying mammals and 130 million years ago for marsupials) and was thus highly salient. After participants had read about the features of kangaroos and platypuses, they were asked, "How are these two mammals different from the other mammals in the tree? Is this difference reflected in the tree itself somehow?"

Most dyads were able to identify a genuine morphological or geographic difference between the non-placental mammals and the other mammals (e.g., "only kangaroos have pouches and only platypuses lay eggs," "they're the only ones that live in Australia"), but very few were able to identify how that difference was reflected in the tree. Six dyads claimed the difference was *not* reflected in the tree (scored -1); 15 dyads claimed they were unsure whether or not the difference was reflected in the tree (scored 0); and 23 dyads claimed the difference was reflected in the tree but did not refer to the branching relations, e.g., "they [the labels] are different colors" or "they [the animals] just look different" (scored +1). Only 5 dyads correctly identified the branching relations as the relevant form of representation, e.g., "this one's branching out, completely separate from these" (scored +2).

Activity 3: Speciation. One of the unique features of the interactive cladogram was a slider at the bottom of the screen for manipulating the timeline, allowing users to scroll between the beginning of the divergence of the 19 orders of mammals (65 million years ago) and the present day. Moving back in time "shrank" the cladogram such that branching events that occurred after that time no longer appeared on the screen. We used this feature of the display to elicit participants' beliefs about the origin of species. Specifically, we asked participants to move the slider to 40 million years, which caused the taxon representing an extinct, hippo-like creature – the paleoparadoxiid – to disappear from the screen. We then asked, "Did you see that the paleoparadoxiid disappeared from the tree? Why do you think that happened? What might have occurred between 40 million years ago and 30 million years ago that led to the appearance of paleoparadoxiid?"

Of interest was whether participants could identify a biologically plausible cause of divergence – i.e., geographic isolation, reproductive isolation, or unique selection pressures. Five dyads did, in fact, cite such a factor, e.g., "maybe the climate changed" (+2). Fourteen dyads noted that the paleoparadoxiid must have evolved during the time period of interest but were unclear on what factors may have driven its divergence from its closest relative (scored +1). The majority of dyads (n = 21) were unsure of what the appearance and disappearance of the paleoparadoxiid icon was supposed to represent biologically (scored 0), and the remaining dyads (n = 9) acknowledged that the appearance and disappearance of the paleoparadoxiid icon was intended to represent a speciation event but cited biological implausible means of speciation, e.g., "the sea cow and the hyrax had a baby" or "it came from the ground" (scored -2).

Activity 4: Relatedness. Another unique feature of the interactive cladogram was that users could "launch the story" of a particular taxon, which opened a window containing detailed information about the taxon's habitat, diet, and morphology. Also contained in these windows were additional cladograms, depicting the relations among other species within the taxa not explicitly depicted in the main display. We used this feature as a vehicle for eliciting participants' understanding of the relatedness of species whose morphological features seemingly belie their evolutionary origins: camels, pigs, and whales. Participants were asked to "launch the story" for the camels, which then brought up a cladogram depicting pigs on the left, camels in the middle, and whales on the right. The branching relations among these taxa indicated that camels are more closely related to whales than to pigs; the outward behavior and morphology of these animals, however, would suggest that camels are more closely related to pigs. Faced with this tension, participants were asked to determine whether camels were more closely related to pigs or to whales and to explain how they were able to discern that fact from the tree. The story itself, we should note, did not explicitly reference this tension, nor did it give participants any clues as to which of the three species were most closely related.

The vast majority of dyads (n = 39) claimed that the tree indicated that camels are more closely related to pigs than to whales (scored -1). Three dyads claimed not to be able to discern what the tree indicated about relatedness (scored 0); four claimed that the tree indicated that camels are more closely related to whales but were unable to explain how it indicated that (scored +1); and three claimed the tree indicated that camels are more closely related to whales and were also able to justify that inference by reference to the branching relations (scored +2).

Activity 5: Extinction. The interactive cladogram contained another feature not represented in standard cladograms: it allowed participants to select images of the specimens on view next to the display and learn about those specimens in the context of the phylogenic relations and historical timeline contained within. We used this feature to elicit participants' reasoning about the relation between extinct species and extant species. Participants were asked to navigate to the "Mammals on Display" tab, select the "entelodont" (an extinct, pig-like creature), and decide (a) whether it still exists and (b) where it might fit in the cladogram in the main display. The entelodont's status as an extinct species could be discerned from the sliding timeline at the bottom of the screen, which made the entelodont non-selectable if moved to a point past its estimated date of extinction. It could also be discerned from the information contained in the pop-up window describing the species. Thus, the question of whether or not the entelodont is extinct was of less interest than the question of where it might fit in the cladogram, which depicted 19 extant species but only one extinct species (the paleoparadoxiid).

As expected, virtually all dyads were able to discern that the entelodont is extinct. Only 19 dyads, however, were able to discern that it would be located among the ungulates (scored +2). Twenty-two dyads indicated that the entelodont would be located in some other part of the display, typically the bottom, near the cladogram's root (scored +1); 6 dyads were unsure of where the entelodont would be located (scored 0); and 2 dyads claimed that the entelodont would *not* fit into the cladogram because it was extinct and extinct species cannot be represented in cladograms (scored -1).

Patterns of engagement. Participants' engagement with each activity was operationalized as the number of distinct utterances and actions produced during that activity. Engagement scores were calculated separately for each dyad member and each activity. To determine whether engagement scores varied by activity or by dyad composition, we ran a repeated-measures analysis of variance (ANOVA) of the engagement scores in which activity (ordering, branching, speciation, relatedness, extinction) and dyad member (parent, child) were treated as within-participants variables and parent's gender (male, female), child's gender (male, female), and child's age (4-8, 9-12) were treated as between-participants variables. This analysis revealed significant effects of activity ($F(4,164) = 5.08$, $p < .01$) and dyad member ($F(1,41) = 30.74$, $p < .001$), but no significant effects of parent's gender, child's gender, or child's age and no significant interactions either.

The effects of activity and dyad member on engagement scores are illustrated in Figure 3A. While dyads produced more utterances and actions for some activities (e.g., branching) than for others (e.g., extinction), this effect was small in comparison to the effect of dyad member. Across activities, the mean engagement score for children was 4.6 (SD = 1.4, range = 1.2 to 8.8), whereas the mean engagement score for parents was only 1.8 (SD = 2.5, range = 0 to 11.6). In fact, 27 of the 49 parents produced fewer than one utterance or action per activity, despite the fact that the interview was framed, in both verbal and written communications, as "a study of how parents and children converse about complex biological topics, like evolution and common descent." The remaining 22 parents, however, tended to produce as many utterances and actions as their children, if not more. This variance, while unexpected, ultimately proved fruitful in determining how engagement scores related to response accuracy at the level of the dyad (discussed subsequently).

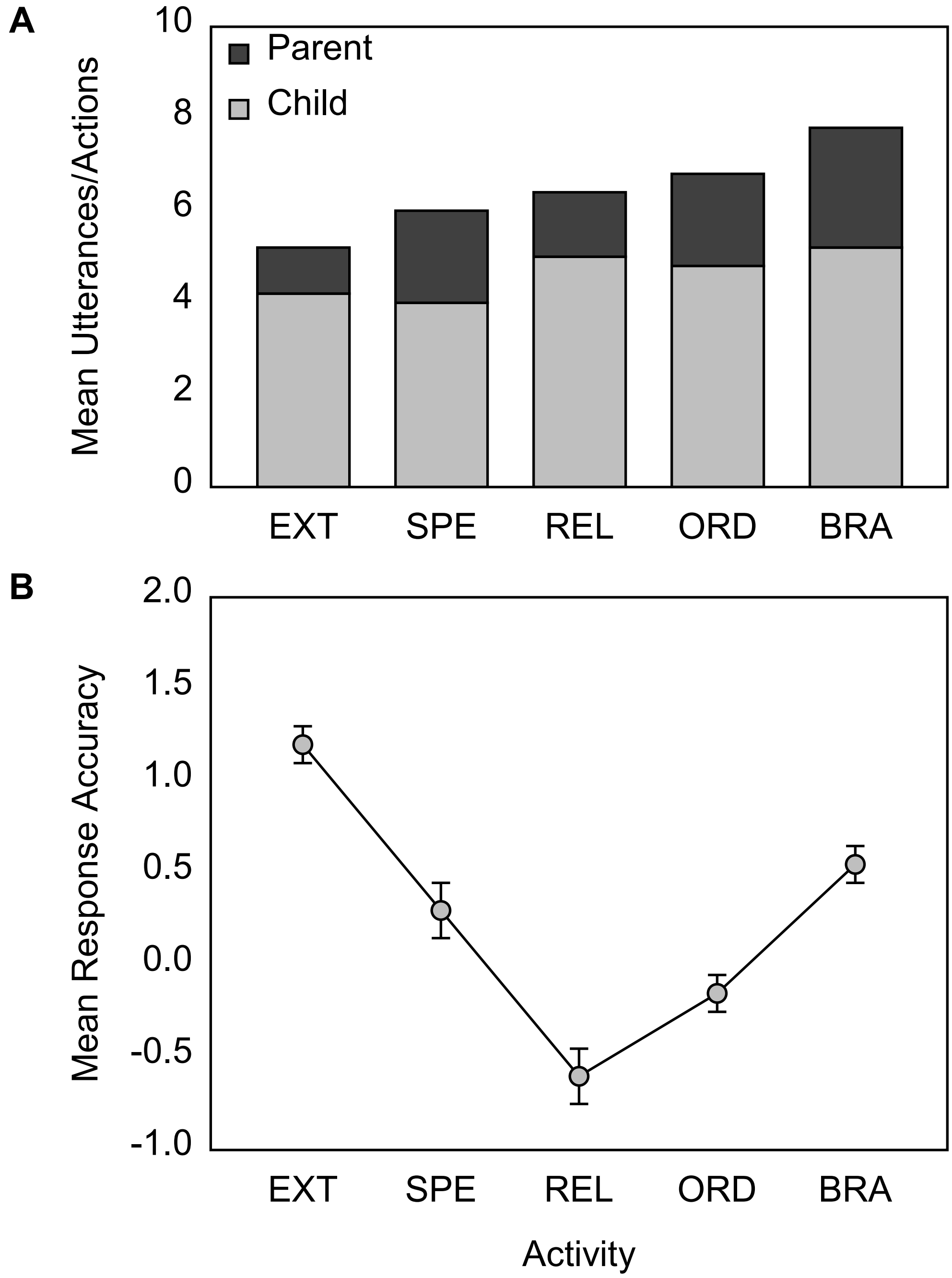


Figure 3. (A) Mean number of utterances/actions per activity per dyad member; (B) Mean accuracy of dyad responses per activity (+ *SE*). EXT = Extinction, SPE = Speciation, REL = Relatedness, ORD = Ordering, BRA = Branching.

Patterns of reasoning. Participants' mean accuracy on each activity is displayed in Figure 3B, ordered by the mean level of engagement with those activities. A repeated-measures

ANOVA revealed not only that accuracy scores differed by activity ($F(4,192) = 34.43$, $p < .01$) but also that accuracy scores were quadratically related to mean engagement scores ($F(1,48) = 96.41$, $p < .001$, for the quadratic contrast). The source of this quadratic relation is not entirely clear, though one possibility is that the association between accuracy scores and engagement scores may have been moderated by a third variable – task difficulty – such that engagement was negatively correlated with response accuracy for easy tasks but positively correlated with response accuracy for harder tasks. Put differently, increased engagement may have been productive for the more difficult tasks but counterproductive for the easier ones.

We did not, however, collect any independent measures of task difficulty, so this speculation requires further verification. Nevertheless, the two activities that differed most in engagement – activity 5 (extinction) and activity 2 (branching) – yielded opposite correlations between accuracy scores and engagement scores at the level of the dyad. Dyads who exhibited more engagement with the branching activity tended to produce more accurate responses ($r = .27$), whereas dyads who exhibited more engagement with the extinction activity tended to produce less accurate responses ($r = -.20$). These correlations were not significantly different from zero (given the small sample size), but they were still significantly different from one another ($z = 3.25$, $p < .01$), suggesting that the relation between engagement and accuracy may differ depending on the conceptual demands of the task.

What factors influenced accuracy of reasoning?

Dyads differed substantially in the accuracy of their reasoning, from providing responses scored -1 on four of the five activities to providing responses scored +2 on four of the five activities. To determine which factors, if any, were associated with accuracy, we regressed each dyad's accuracy score against five dyad-specific measures: the child's age (in years), the child's gender (0 = female, 1 = male), the parent's gender (0 = female, 1 = male), the parent's mean engagement score, and the child's mean engagement score. Only two measures emerged as significant predictors of response accuracy: the child's age ($\beta = .32$, $t = 2.39$, $p < .05$) and the parent's mean engagement score ($\beta = .36$, $t = 2.61$, $p < .05$). These effects are displayed in Figure 4, with child's age dichotomized as "younger" (4-8) and "older" (9-12) and parent's engagement scores dichotomized as "low" ($M < 1$ utterance or action per activity) and "high" ($M > 1$ utterance or action per activity). Dyads with older children produced more accurate responses than dyads with younger children, and dyads with high parental engagement scores produced more accurate responses than dyads with low parental engagement scores. A univariate ANOVA revealed no interaction between child's age and parental engagement ($F(1,45) < 1$, *ns*), though it did reveal that the effect of parental engagement ($\eta^2 = .13$) was nearly three times as large as the effect of child's age ($\eta^2 = .05$).

To explore the effect of parental engagement further, we computed an additional measure of engagement: the number of alternations between the child's contribution to the conversation (utterance or action) and the parent's contribution. While this measure was strongly correlated with our initial measure of parental engagement ($r = .74$), it was not entirely redundant with that measure. Figure 5 illustrates how the two measures diverged. The transcript shown in 5A contains multiple alternations between parent and child, with parent A actively eliciting information from the child, either from the display or from memory. The transcript shown in 5B, on the other hand, contains far fewer alternations, with parent B generally dominating the conversation. While parents in both dyads produced a

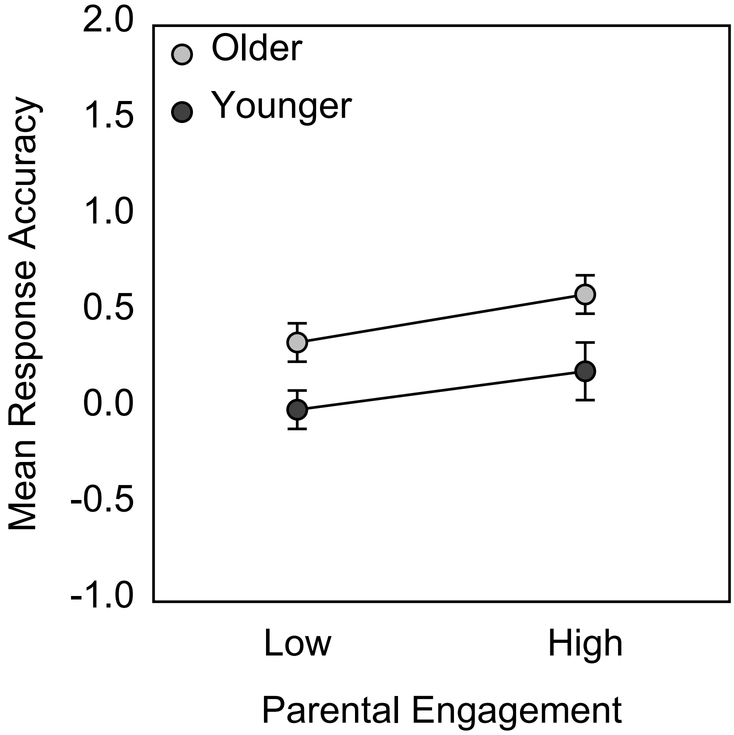


Figure 4. Mean accuracy of dyad responses (+ *SE*) as a function of parental engagement (low vs. high) and child's age (younger vs. older).

similar number of utterances or actions per activity (*M* for parent A = 5.0, *M* for parent B = 4.8), the two dyads received very different alteration scores (*M* for dyad A = 4.8, *M* for dyad B = 1.8). It should be noted that dyads A and B were two of only 33 dyads for which alternation scores could be computed at all. The remaining 16 dyads included too little parental engagement, typically because the parents in these dyads resigned themselves to simply watching their children complete the activities. It should also be noted that the overall paucity of parental involvement rendered the use of more detailed, content-specific coding schemes (like those used by Crowley et al., 2001, or Evans et al., 2010) impractical.

Among the 33 dyads for which alternations could be computed, the mean number of alternations was 14.2 ($SD = 10.7$, range = 2 to 44), and alternations were strongly correlated with response accuracy ($r = .55$). Alternations actually proved to be a stronger predictor of response accuracy than parental engagement in general ($r = .35$). But were alternations a *unique* predictor of response accuracy? We addressed this question by submitting dyads' mean accuracy scores to a hierarchical regression in which alternation scores were entered subsequent to parental engagement scores. Whereas parental engagement scores explained 14% of the variance in response accuracy in the initial model ($F(1,31) = 5.00$, $p < .05$), alternation scores explained an additional 17% of variance in the subsequent model (*F-change*$(2,30) = 6.69$, $p < .01$). In fact, the partial correlation between alternation scores and response accuracy, controlling for parental engagement scores, was nearly as large as the zero-order correlation between these two variables ($r = .44$, $p < .01$), whereas the partial correlation between parental engagement scores and response accuracy, controlling for alternation scores, was no longer significant ($r = .06$, *ns*). The *nature* of parental engagement thus appeared to be more important than the mere act of engagement.

A

E: Select the kangaroo and the platypus and read about their features.

P: Can you find the kangaroo?
C: [Hits the kangaroo]
P: [Reads it to her] OK, now the platypus. You keep looking over it.
C: [Hits the kangaroo]
P: [Reads it to her]

E: How are these two mammals different from the other mammals in the tree?

P: Where do their babies grow?
C: In their pouch.
P: Right, what about the platypus?
C: I don't know.
P: Remember it lays eggs, now where do they come from?
C: Australia and New Guinea.

E: Is this difference reflected in the tree itself somehow?
P: By the colors.

B

E: Select the kangaroo and the platypus and read about their features.

P: Kangaroo and platypus.
C: [Leans on display looks around]
P: OK, here's the kangaroo and the platypus.
C: [Reads them]
P: OK, now this one, most other mammals don't lay eggs, but the platypus is a mammal that lays eggs. That's a big deal, will you remember that?

E: How are these two mammals different from the other mammals in the tree?

P: Well, the platypus lays eggs and live in the water, and the kangaroo has a pouch and keeps their young in the pouch.

E: Is this difference reflected in the tree itself somehow?

P: Well, they are all mammals, but [points to tree, sweeping motion]. Hmm. It's obvious about the platypus...

Figure 5: Examples of high parental engagement involving (a) frequent alternations and (b) infrequent alternations between parent and child contributions. E = Experimenter, P = Parent, and C = Child. Actions are denoted in brackets.

Conclusions

Evolutionary concepts like *speciation*, *extinction*, and *common descent* are notoriously difficult to understand (Gregory, 2008; Shtulman, 2006), and canonical representations of those concepts are notoriously difficult to interpret (Catley et al., 2010; Novick et al., 2010). The present study sought to determine how parents and children discuss such concepts and interpret such representations in the context of an interactive museum display. We found that parents and children, like the college-aged students tested in previous research, exhibited significant difficulty in interpreting the core feature of a cladogram – namely, the branching relations among its taxa. Parents and children also revealed the same kinds of evolutionary misconceptions documented in previous research, including (a) that the ordering of a species in a cladogram carries biologically relevant information, (b) that the morphological overlap between species is a reliable indicator of shared ancestry, and (c) that speciation occurs through a kind of instantaneous transformation or emergence. Nevertheless, the frequency of those misconceptions was negatively correlated with the degree to which parents were involved in generating a response, particularly the degree of turn-taking between parents and children. The more often the two collaborated (by this measure), the more often they generated accurate interpretations and explanations of the phenomena at hand.

These findings have implications for both the study of evolutionary reasoning and the design of informal learning environments. With respect to evolutionary reasoning, they suggest that collaboration may be an effective means of reducing, or even eliminating, evolutionary misconceptions. Previous research by Asterhan and Schwartz (2007, 2008) found that undergraduates who answered evolution-based problems on their own learned less about evolution, in both the short-term and the long-term, than those who worked in pairs. They also found that the style of a dyad's interaction influenced learning such that dyads who engaged in argumentation exhibited greater learning gains than those who merely shared information. Assuming that turn-taking between parents and children served as a proxy for argumentation, the present study not only replicates Asterhan and Schwartz's findings but also extends those findings to (a) populations of different ages and (b) conversations elicited in more naturalistic contexts.

That said, the studies by Asterhan and Schwartz employed controls absent from the present study, including the provision of instruction on how to conduct a critical discussion and the administration of pre- and post-intervention measures of evolution understanding. Controls like these are difficult to implement in a museum setting, where participants' time and attention are limited resources. It may thus be beneficial to test parent-child dyads in a more controlled (laboratory) setting, particularly for the purpose of assessing *learning*. While we did, in fact, document a strong relation between the level of interaction between dyad members and the accuracy of their responses, it is unclear whether greater dyadic interaction was a cause, or merely a symptom, of more accurate responding. Dyads who entered the museum with better evolution understanding may have engaged in more discussion simply because they were more familiar, or more comfortable, with the topics at hand.

At least two considerations militate against this interpretation, however. First, not all forms of engagement were associated with higher response accuracy. Children's engagement scores, for instance, were uncorrelated with response accuracy, as were parental engagement scores after controlling for parent-child alternations. Second, engagement was positively correlated with response accuracy for some activities but negatively correlated with others. Still, future research should explore the effects of parent-child interaction on learning more directly by including both pretests and posttests and by administering them (separately) to both parents and children.

With respect to informal learning environments, our findings suggest that environments that encourage dyadic interaction – particularly parent-child interaction – may be more effective at fostering learning than those that encourage more monadic forms of exploration. These findings echo many other findings in the science education literature. Van Schijndel, Franse, and Raijmakers (2010), for instance, found that children who were actively coached by their parents during a museum visit engaged more effectively with the exhibits than children who were not coached in this way. Tenenbaum, Prior, Dowling, and Frost's (2010) found that parent-child dyads who explored a museum with a booklet of activities spent more time exploring and discussing the exhibits than dyads who explored the museum in a less structured way. And Tare et al. (2011) found that parents' explanatory behavior at a museum exhibit was positively correlated with their children's explanatory behavior. Two aspects of our findings, however, stand in contrast to these earlier findings.

First, the relation between dyads' level of engagement and accuracy of reasoning was not entirely straightforward. Dyads reasoned most accurately for activities in which they were either *most* engaged or *least* engaged; intermediate levels of engagement were, on the other hand, associated with the poorest performance (see Figure 3). Whatever the cause of this effect may be, the effect itself implies that increased engagement is not always productive or beneficial and that learners may need additional guidance in how to allocate their attention to the displays at hand (see, e.g., Allen & Gutwill, 2004).

Second, we did not explicitly instruct parents on how to engage with their children during the activities, leaving open the option that parents might chose not to engage at all, and, to our surprise, approximately half adopted that option. Anecdotal evidence suggests that what led to such high levels of parental attrition was not confusion about the purpose of the study but the nature of the touch-screen display, which was clearly designed for a single user. If two users touched the screen at the same time, then either the cursor on the display would toggle between the two points of contact or the display would reset itself altogether, closing all pop-up windows and restoring all default settings (e.g., restoring the sliding timescale to "present day"). Dyads quickly became aware of this contingency and delegated the task of manipulating the display to a single individual, typically the child. As a result, many parents seemed to become increasingly disengaged over the course of the interview, an observation backed by a significant drop in parental engagement scores from the first activity to the last (M = 1.9, 2.6, 2.0, 1.4, 1.0; linear contrast: $F(1,48) = 12.33$, $p < .01$). This unexpected finding highlights a paradox in the use of touch-screen displays increasingly populating the halls of science museums today: such displays are typically designed for a single user, yet, in our study, dyads in which a single member was actively engaged with the display profited less from the display than those in which both members were actively engaged.

The conceptual benefits of collaborative activity over solitary activity have been documented in a variety of studies (e.g., Craig, Chi, & VanLehn, 2009; Okada & Simon, 1997, Schwarz, Neuman, & Biezuner, 2000). The mechanism behind this effect appears to be the addition of a "social impetus" to explain and justify one's reasoning. Museum displays that can capture or create this kind of impetus would thus seem to be more efficacious than those that cannot. That said, the design of such displays is constrained by additional, pragmatic factors that might effectively limit interactivity, including how long visitors can be expected to use the display, how well visitors can discern the purpose of the display's affordances, and how constructively visitors can intervene on the phenomena of interest (Allen & Gutwill, 2004). Which considerations to privilege over others is a question that likely merits a different answer for every display. Still, our research suggests that, at a minimum, touch-screen displays should be tolerant of multiple points of contact so that parents and children may jointly interact with the display without accidentally erasing one another's paths of exploration.

In conclusion, visitors to a science museum are no more immune to evolutionary misconceptions than other populations. Science museums, however, provide unique opportunities for collaboratively discussing evolutionary phenomena, and such discussions appear to help attenuate evolutionary misconceptions. While further research needs to be done on how parent-child conversations foster accurate evolutionary reasoning and whether such conversations lead to long-term learning, the current findings point to a promising new method for increasing evolution understanding among the general public.

• • •

Acknowledgements

This research was supported by National Science Foundation grant DRL-0953384 awarded to Andrew Shtulman. We would like to thank Emi Yoshimura and the staff of the Los Angeles Natural History Museum for collaborating with us on the design and implementation of the testing procedure, as well as Katherine Abelson, Elahhe Afkhamnejad, Catherine Kiang, and Madalyn Long for their assistance with data collection and data analysis.

Andrew SHTULMAN. Received a Ph.D. in Psychology from Harvard University in 2006. He has been a member of the Department of Psychology at Occidental College since 2007. His research focuses on conceptual development and conceptual change, particularly as they relate to science education.

Isabel CHECA. Is an undergraduate at Occidental College, anticipated to graduate in 2013. She has been a member of Dr. Shtulman's lab since 2010 and is interested in early childhood education.

References

Allen, S., & Gutwill, J. P. (2004). Designing with multiple interactives: Five common pitfalls. Curator, 47, 199-212.

Anggoro, F. K., Waxman, S. R., & Medin, D. L. (2008). Naming practices and the acquisition of key biological concepts: Evidence from English and Indonesian. Psychological Science 19, 314-319.

Asterhan, C. S. C., & Schwartz, B. B. (2007). The effects of monological and dialogical argumentation on concept learning in evolutionary theory. Journal of Educational Psychology, 99, 626-639.

Asterhan, C. S. C., & Schwartz, B. B. (2009). Argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialog. Cognitive Science, 33, 374-400.

Baum, D. A., Smith, D. S., & Donovan, S. S. S. (2005). The tree-thinking challenge. Science, 310, 979-980.

Berti, A. E., Toneatti, L., & Rosati, V. (2010). Children's conceptions about the origin of species: A study of Italian children's conceptions with and without instruction. Journal of the Learning Sciences, 19, 506-538.

Bishop, B. & Anderson, C.A. (1990). Student conceptions of natural selection and its role in evolution. Journal of Research in Science Teaching, 27, 415-427.

Brumby, M. N. (1984). Misconceptions about the concept of natural selection by medical biology students. Science Education, 68, 493-503.

Catley, K. M., & Novick, L. R. (2008). Seeing the wood for the trees: An analysis of evolutionary diagrams in biology textbooks. BioScience, 58, 976-987.

Catley, K. M., Novick, L. R., & Shade, C. K. (2010). Interpreting evolutionary diagrams: When topology and process conflict. Journal of Research in Science Teaching, 47, 861–882.

Craig, S. D., Chi, M. T. H., & VanLehn, K. (2009). Improving classroom learning by collaboratively observing human tutoring videos while problem solving. Journal of Educational Psychology, 101, 779-789.

Crowley, K., Callanan, M., Jipson, J., Galco, J., Topping, K., & Shrager, J. (2001). Shared scientific thinking in everyday parent-child activity. Science Education, 85, 712-732.

Crowley, K., Callanan, M.A., Tenenbaum, H. R., & Allen, E. (2001). Parents explain more often to boys than to girls during shared scientific thinking. Psychological Science, 12, 258-261.

Diamond, J. (1994). Sex differences in science museums: A review. Curator, 37, 17-24.

Jipson, J. L., & Callanan, M. (2003). Mother-child conversation and children's understanding of biological and non-biological changes in size. Child Development, 74, 629-644.

Evans, E. M., Spiegel, A., Gram, W., Frazier, B. F., Tare, M., Thompson, S., & Diamond, J. (2010). A conceptual guide to natural history museum visitors' understanding of evolution. Journal of Research in Science Teaching, 47, 326-353.

Gelman, S. A. (2003). The essential child. Oxford, UK: Oxford University Press.

Gelman, S. A., & Coley, J. D. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. Developmental Psychology, 26, 796-804.

Gleason, M. E., & Schauble, L. (2000). Parents' assistance of their children's scientific reasoning. Cognition and Instruction, 17, 343-378.

Gunderson, E. A., & Levine, S. C. (2011). Some types of parent number talk count more than others: Relations between parents' input and children's cardinal-number knowledge. Developmental Science, 14, 1021-1032.

Gregory, T. R. (2008). Understanding evolutionary trees. Evolution: Education and Outreach, 1, 121-137.

Gregory, T. R. (2009). Understanding natural selection: Essential concepts and common misconceptions. Evolution: Education and Outreach, 2, 156-175.

Griffith, J. A., & Brem, S. K. (2004). Teaching evolutionary biology: pressures, stress, and coping. Journal of Research in Science Teaching, 41, 791-809.

Hatano, G., & Inagaki, K. (1994). Young children's naïve theory of biology. Cognition, 50, 171-188.

Legare, C. H., Lane, J., & Evans, E. M. (in press). Anthropomorphizing science: How does it affect the development of evolutionary concepts? Merrill-Palmer Quarterly.

Mayr, E. (1982). The growth of biological thought: Diversity, evolution, and inheritance. Cambridge: Harvard University Press.

Meir, E., Perry, J, Herron, J. C., & Kingsolver, J. (2007). College students' misconceptions about evolutionary trees. American Biology Teacher, 69, 71-76.

Miller, J. D., Scott, E. C., & Okamoto, S. (2006). Public acceptance of evolution. Science, 313, 765-766.

National Research Council (1996). National science education standards. Washington, DC: National Academy Press.

National Research Council (1998). Teaching about evolution and the nature of science. Washington, DC: National Academy Press.

Nettle, D. (2010). Understanding of evolution may be improved by thinking about people. Evolutionary Psychology, 8, 205-228.

Newport, F. (2010). Four in 10 Americans believe in strict creationism. Gallup Organization.

Novick, L. R., & Catley, K. M. (2007). Understanding phylogenies in biology: The influence of a Gestalt perceptual principle. Journal of Experimental Psychology: Applied, 13, 197-223.

Novick, L. R., Shade, C. K., & Catley, K. M. (2010). Linear versus branching depictions of evolutionary history: Implications for design. Topics in Cognitive Science, 3, 536-559.

Okada, T., & Simon, H. (1997). Collaborative discovery in a scientific domain. Cognitive Science, 21, 109-146.

Rigney, J. C., & Callanan, M. A. (2011). Patterns in parent-child conversations about animals at a marine science center. Cognitive Development, 26, 155-171.

Samarapungavan, A. & Wiers, R. W. (1997). Children's thoughts on the origin of species: A study of explanatory coherence. Cognitive Science, 21, 147-177.

Schwarz, B. B., Neuman, Y., & Biezuner, S. (2000). Two wrongs may make a right...if they argue together! Cognition & Instruction, 18, 461-494.

Shtulman, A. (2006). Qualitative differences between naive and scientific theories of evolution. Cognitive Psychology, 52, 170-194.

Shtulman, A., & Calabi, P. (2012). Cognitive constraints on the understanding and acceptance of evolution. In K. S. Rosengren, S. Brem, E. M. Evans, & G. Sinatra (Eds.), Evolution Challenges: Integrating Research and Practice in Teaching and Learning about Evolution (pp. 47-65). Cambridge, UK: Oxford University Press.

Shtulman, A., & Calabi, P. (in press). Tuition vs. intuition: Effects of instruction on naive theories of evolution. Merrill-Palmer Quarterly.

Shtulman, A., & Schulz, L. (2008). The relation between essentialist beliefs and evolutionary reasoning. Cognitive Science, 32, 1049-1062.

Sinatra, G. M., Brem, S. K., & Evans, E. M. (2008). Changing minds? Implications of conceptual change for teaching and learning about biological evolution. Evolution: Education and Outreach, 1, 189-195.

Solomon, G. E. A., & Zaitchik, D. (2012). Folkbiology. WIREs Cognitive Science, 3, 105-115.

Southerland, S. A., Abrams, E., Cummins, C. L., & Anzelmo, J. (2001). Understanding students' explanations of biological phenomena: Conceptual frameworks or p-prims? Science Education, 85, 328-348.

Spiegel, A., Evans, E. M., Frazier, B. F., Hazel, A., Tare, M., Gram, W., & Diamond, J. (2012). Changing museum visitors' concepts of evolution. Evolution: Education and Outreach, 5, 43-61.

Tare, M., French, J., Frazier, B., Diamond, J., & Evans, E. M. (2011). Explanatory parent-child conversation predominates at an evolution exhibit. Science Education, 95, 720-744.

Tenenbaum, H. R., Prior, J., Dowling, C. L., & Frost, R. E. (2010) Supporting parent-child conversations in a history museum. British Journal of Educational Psychology, 80, 241-254.

Torrens, E., & Barahona, A. (2012). Why are some evolutionary trees in natural history museums prone to being misinterpreted? Evolution: Education and Outreach, 5, 76-100.

van Schijndel, T. J. P., Franse, R. K., & Raijmakers, M. E. J. (2010). The exploratory behavior scale: assessing young visitors hands-on behavior in science museums. Science Education, 94, 794-809.

Waxman, S., Medin, D., & Ross, N. (2007). Folkbiological reasoning from a cross-cultural developmental perspective: Early essentialist notions are shaped by cultural beliefs. Developmental Psychology, 43, 294-308.

International Electronic Journal of Elementary Education, 2012, 5(1), 47-62.

Elementary Children's Retrodictive Reasoning about Earth Science

Julie C. LIBARKIN*
Michigan State University, East Lansing MI, USA

Matthew H. SCHNEPS
Harvard-Smithsonian Center for Astrophysics, Cambridge MA, USA

Received: 2 September 2012 / Revised: 10 November 2012 / Accepted: 10 November 2012

Abstract

We report on interviews conducted with twenty-one elementary school children (grades 1-5) about a number of Earth science concepts. These interviews were undertaken as part of a teacher training video series designed specifically to assist elementary teachers in learning essential ideas in Earth science. As such, children were interviewed about a wide array of earth science concepts, from rock formation to the Earth's interior. We analyzed interview data primarily to determine whether or not young children are capable of inferring understanding of the past based on present-day observation (retrodictive reasoning) in the context of Earth science. This work provides a basis from which curricula for teaching earth and environmental sciences can emerge, and suggests that new studies into the retrodictive reasoning abilities of young children are needed, including curricula that encourage inference of the past from modern observations.

Keywords: Earth Science, Reasoning, Retrodiction.

Introduction

This paper discusses the nature of children's reasoning about earth phenomena and processes, and specifically the extent to which retrodictive reasoning is evident in their discourse. We utilize a set of twenty-one interviews with elementary-aged children as the data set from which evidence of retrodictive reasoning emerged. We also documented the presence of alternative conceptions about the earth and considered the extent to which these alternative conceptions interfered with reasoning.

Retrodictive Reasoning in Earth Science

Retrodiction, the interpretation of present-day evidence to infer ancient processes, lies at the heart of much of earth science (e.g., Ault, 1998). While prediction has a role in earth science (e.g., for forecasting natural hazards or extrapolating the impact of human actions on natural

* ✉ Center for Integrative Studies in General Sciences & Dept. of Geological Sciences, Michigan State University, 288 Farm Lane, 206 Natural Science, East Lansing, MI USA 48824, Phone: 011+1+517-355-8369, E-mail: libarkin@msu.edu

ISSN:1307-9298

www.iejee.com

systems), retrodiction lies at the heart of all fields associated with paleo-processes, including geology, evolutionary biology, and cosmology. Interestingly, the concept of retrodiction is not commonly found in discourse about scientific reasoning that emerges from the science education community (Sibley, 2009), perhaps because predictive domains of science dominate the field. At the same time, retrodictive reasoning is of vital importance because of the role it plays in public debate about topics such as evolution, the creation of the universe, and the age of the Earth.

What are the unique characteristics of retrodictive reasoning? Retrodictive reasoning requires the understanding that patterns present in the modern world are the imprints of processes that have already occurred. As a consequence of this recognition, retrodictive reasoners must be able to extrapolate possible causes for these patterns, balance the probability of one specific cause against the likelihood of another, and rationalize a preference for one particular event resulting in an observed pattern. In this way, retrodictive reasoners recognize the role of causation in the production of patterns; interestingly, not all people are able to link processes and patterns together (Libarkin & Kurdziel, 2006). Within this ontology also lies the need for using narrative to explain phenomena (Norris et al., 2005) as well as reasoning about time.

Retrodictive reasoning about earth systems is inherently connected to systems thinking (Kali et al., 2003; Lawton, 2001). Systems thinkers must recognize that processes, the events that result in observable patterns, often interact to produce surprising results. Recognizing not only singular events but also the confluence of events is the hallmark of an effective systems thinker. Systems thinking also requires an understanding that processes, particularly within complex system like the Earth, do not always interact in linear ways. Non-linear processes, including negative and positive feedback loops, are important components of Earth systems thinking for both modern and ancient Earth. Retrodictive reasoning is inherently different from predictive reasoning. Both a process and its result are observable when they are occurring in real time, thus allowing a prediction to be tested. A process that has already occurred is not observable; one can only engage in experiments, in the lab or through observable natural systems, which replicate the process and look for results consistent with the original observation. In retrodiction, one can never actually observe the original process in action. This results in interesting problems from a scientific perspective; one can never completely disprove a hypothesis about a process that has long since occurred. One can only engage in a “more likely than not”, or vice versa, standard.

Curricula that explicitly address retrodiction, even in its simplest forms, are surprisingly uncommon in the earth sciences especially for young children, although inquiry in Earth science education requires attention to retrodiction (Pyle, 2008). Existing studies demonstrate that middle and high school students can engage in retrodictive reasoning about possible evolutionary pathways through inquiry with hominid skulls and radiometric data (Thomson & Chapman Beall, 2008). Similarly, geology majors in a capstone course specifically oriented towards retrodiction in global systems reported better understanding of Earth’s spheres after engaging in the course as well as more confidence in their ability to retrodict patterns based on observable modern processes (Sunderlin, 2009).

Alternative Conceptions in Earth Science

Although alternative conceptions are not the focus of this study, the students in our interviews present a number of non-scientific ideas that warrant discussion of alternative conceptions here. A growing body of literature has documented the alternative conceptions about Earth’s systems held by elementary, secondary, and advanced students (see reviews of Cheek, 2010; Dove, 1998; King, 2008). These conceptions provide a window into the reasoning that might be occurring as students interact with Earth science concepts and phenomena, as well as insight into the potential difficulties students may face in the classroom. An understanding of alternative conceptions is vital for teachers interested in aligning curriculum with student needs, and exposure to the ideas of others can provide students themselves with a gateway into learning complex material.

Alternative conceptions about Earth science have been documented across the Earth system and across age groups. For the purposes of this paper, we focus on alternative conceptions related to the topics covered in our study, including geologic time, particularly as it relates to the timing and rate of Earth processes, rock and soil formation, and deep Earth processes related to, for example, plate tectonics and magma formation. Studies of student ideas about Earth science are much rarer than in other disciplines. Where possible, we report on studies of young children, and include studies of older students (high school, college) where relevant studies of young children are not available.

Student conceptions about Earth's surface processes are often related to their personal observations of the natural world. For example, alternative conceptions about rock formation mechanisms can be understood in the context of observable surficial processes. College students in two studies (Kortz & Murray, 2009; Kusnick, 2002) articulated the idea that rocks form when water dries up or when water deposits material into piles; the simple acts of drying and depositing generate aggregate rocks. Younger students also describe rocks as growing from smaller objects or pebbles (Ault, 1984; Blake, 2005; Dal, 2007), an idea that may also be present in older students (Kusnick, 2002). The relationship between rocks and soils is also sometimes misunderstood, with some teachers believing that soils are deposited as rock layers (Gosselin & Macklem-Hurst, 2002). Happs (1984) noted particularly the importance of geologic time in understanding soil formation, and many of the aforementioned studies note difficulty students have in conceptualizing deep time.

Conceptions about deep Earth processes may more often be driven by instruction, rather than personal experience. Phenomena that are not directly tangible but are rather recognized by their effects, such as plate tectonics, geomagnetism, and gravity, can be particularly difficult for students in Earth science courses to understand (Libarkin & Kurdziel, 2006). This can result in confusion at basic levels, such as with simple terminology used to explain deep Earth processes (e.g., Libarkin et al., 2005), and at more conceptual levels. For example, students may believe that earthquakes push tectonic plates (Barrow & Haskins, 1996; Ross & Shuell, 1993), that mountains simply grow (e.g., Muthukrishna et al., 1993; Trend et al., 2000), or that volcanic magma originates at the Earth's core (Nelson et al., 1992). Students also draw a surprising array of models of the Earth's interior when asked to imagine cutting the Earth in half, including an Earth containing flat or no layers (e.g., Blake, 2005; DeLaughte et al., 1998; Libarkin et al., 2005; Lillo, 1994). The physical location of tectonic plates and the physical state of Earth's interior are also areas of significant confusion.

Student ideas about geologic time have focused on understanding of the order of events as well as the long timescales ('deep time') inherent to many Earth processes. Ault (1982) recognized that young children are able to reason about relative time, an ability that is found across age ranges (e.g., Dahl, Anderson, & Libarkin, 2005; Libarkin et al., 2005; Trend, 1998, 2000, 2001). A number of alternative conceptions about the relative ordering of events have been identified, including that man and classical dinosaurs co-existed (Schoon, 1995), life and supercontinents existed when Earth first formed (Libarkin et al., 2005), and similar ideas related to the misunderstanding of the order and scale of geologic events. The idea that temporal reasoning is unique from other abilities (Montangero, 1996) has also been applied to studies of high school students, with the conclusion that difficulties in reasoning about deep time may be related to deficiencies in diachronic thinking (Dodick & Orion, 2003). Similarly, Trend (1998) suggests that mathematical difficulties associated with understanding large numbers may inhibit understanding of deep time.

Aim of the Study

The Earth sciences have generally not received as much attention within science education as biology, chemistry, and physics have, although a growing recognition of the importance of Earth science in schooling and global discourse is increasing the attention paid to it and related disciplines by science educators (Lewis & Baker, 2009). Our goal in undertaking this study is to further the understanding that teachers and researchers have about the ways in which children reason about Earth processes. To further this goal, we analyzed student ability to engage in simple retrodictive reasoning, as reflected in their discourse about Earth processes, and utilized this opportunity to also document the presence of alternative conceptions.

Based on existing work that clearly shows predictive reasoning ability among children (i.e., Zimmerman, 2000), we hypothesized that elementary students would be capable of engaging in the aspects of simple retrodictive reasoning needed to reason in Earth science. We present evidence of retrodictive reasoning as suggested by student discourse, and considered the role that age might play in student reasoning ability. Secondarily, we hypothesized that, given the limited research into children's alternative conceptions about Earth science, confirmation of existing studies and new alternative conceptions about Earth processes and phenomena would be evident in student responses.

Methods

Context

Interviews with twenty-one elementary students were analyzed for this study. These data originated from a series of interviews conducted during the creation of a video series created to support teacher-education, published in 2004 (Argow, Reilly, & Schneps, 2004). This video series, containing edited components of the interviews analyzed here, is accessible online (http://www.learner.org/resources/series195.html) and was designed to help elementary-grade teachers develop deep understanding of the science concepts needed to effectively address standards in Earth and Space Science. Interviews with children were used to engage teachers about the prevalence of student misconceptions, and were coupled with in situ interviews with real geologists, explanatory simulations, and online activities. Students were interviewed singly and in pairs, and some students were interviewed more than once.

Participants (and guardians) provided written informed consent for this study. Nevertheless, we delayed publication of these data expressly because some of the children discussed here are shown in the video series – we wanted enough time to pass between publication of the series and analysis of these data to ensure that neither we nor readers would be able to identify specific participants. That is, given the nine-year delay between these interviews and this analysis, we are now unable to link any de-identified interview transcripts with specific children. Finally, while we present general demographic information below, we have explicitly limited the information provided to ensure interviewee anonymity.

Procedures

Recruitment. Participants were twenty-one elementary-aged children recruited from classrooms in a large, North American city. Given that the interviews were conducted to provide source material for teacher education, participants were selected on the basis of their interests in science and their willingness to appear on camera, as gauged by their teachers. Ethnicities represented a broad diversity as expected for a large city; we do not provide specific ethnicity information. Children were enrolled in first through fifth grade and were 62% (n=13) female (Table 1).

Table 1. *Grade and Gender Distribution of Interviewees*

Grade	*Gender**	*General Topics Covered*
1st	1 F, 1 M	Soil, earth's interior, air
2nd	2 F, 1 M	Rock formation, soil, volcanoes, earth's interior, plate tectonics
3rd	5 F, 2 M	Rock formation, soil, volcanoes, earth's interior, plate tectonics
4th	3 F, 3 M	Rock formation, volcanoes, earth's interior, erosion
5th	2 F, 1 M	Rock formation, soil, volcanoes, earth's interior

* F=Female, M=Male.

Semi-Structured Interviews. All interviews were conducted in the same room. Interviews were audio- and video-taped, with up to three video technicians present to manage multiple cameras and the audio recording. The first author conducted all but three of the interviews, and the second author observed all of the interviews. During most of the interviews, the second author was able to speak to the interviewer through an earpiece; this allowed the second author to make suggestions for the interviews without interfering with the interview process directly. Care was taken to ensure the interviewer did not to introduce verbal or non-verbal cues that would direct or lead the participant's response. In particular, a set of predetermined interview prompts was used with all subjects and any interview probes were derived from the interviewees' own language during the interviews.

Each interviewee was given a unique identifying code prior to analysis; the second letter and the number indicate gender and grade level, respectively, while the first letter makes each code unique. Most of the children were interviewed alone; eight children were interviewed in pairs and two participants were interviewed more than once. Interviews started with general topics related to rocks, soil, mountains, and water, although interviewees ultimately discussed a variety of other topics with interviewees (Table 1). Each interview started with a visual or drawing prompt. At the start of each interview, children were asked to draw pictures and/or were shown rocks (e.g., Fig. 1), a bucket of soil, a rain stick, or photos to prompt their thinking. These probes were used as needed throughout the interview, and interviewees were encouraged to draw out their ideas. The interview protocol was semi-structured. The interviewer began each interview with a few scripted questions, and these questions were used throughout the interview to redirect the discussion, as well. As the interview progressed, probes were generated in response to interviewee discourse and drawings. This resulted in a wide variety of topics being covered across the entire sample.

Coding. Transcript analysis focused on retrodictive reasoning patterns, with a secondary purpose of documenting alternative conceptions. First, we analyzed transcripts for the presence or absence of basic retrodictive reasoning. This coding scheme does not represent retrodictive reasoning in its entirety, but rather models the most simple aspects of retrodictive reasoning that are necessary for effective reasoning about the Earth:

1) Links are made between observations and processes as evidenced by an understanding that patterns present in the modern world are the imprints of processes that have already occurred.
2) Multiple working hypotheses are raised (multiple possible causes for these patterns are held simultaneously).
3) Preference for one hypothesis over others are rationally explained.
4) Reasoning references time beyond human timescales.

Reliability. Inter-rater reliability for codes was established through concurrent coding of ~10% of the responses by the first author and a colleague (a geologist and former high school teacher). The raters calculated the single measures intraclass correlation across all retrodictive codes evaluated by both authors. The average measures intraclass correlation was 0.92 (min.=0.82 and max.=0.96). An intraclass correlation of 1.0 implies perfect reliability; a correlation of 0.92 indicates that both raters were in strong agreement. Disagreement related to the similarity in two codes, one that related to interacting events and the other associated with feedback; this disagreement was clarified through recognition that neither was fundamental to retrodictive reasoning and both were removed. This high reliability for the coded subsample allowed one author to complete the coding with a reasonable assumption of reliability for all analyses.

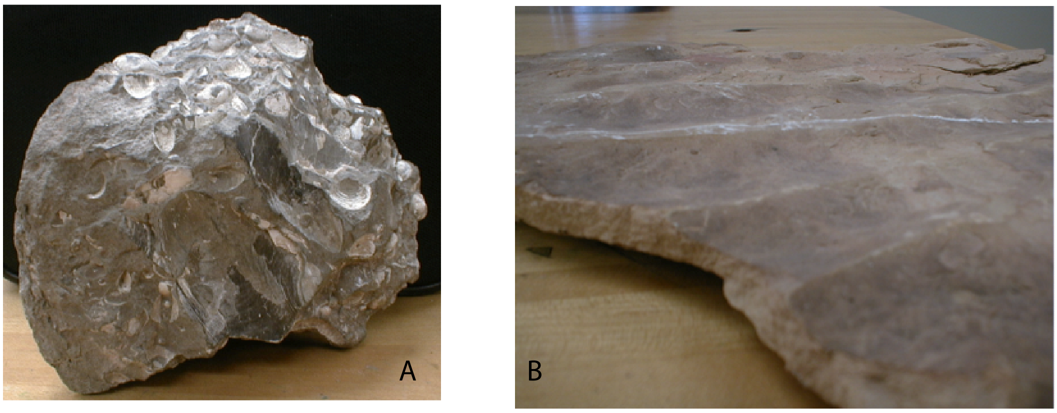


Figure 1. Examples of rocks used to engage students during interviews. A) Fossil-rich limestone. B) Fine-grained sandstone with ripple marks.

Results and Discussion

The structure of students' retrodictive reasoning patterns is intertwined with the presence of alternative conceptions about Earth phenomena. As such, we discuss retrodictive reasoning patterns in significant detail, and note where alternative conceptions are also evident in the student discourse. We particularly focus on evidence that highlights the presence of retrodictive reasoning in students.

The elementary students interviewed in this study exhibited varying degrees of retrodictive reasoning. Nearly every interviewee recognized that Earth materials and patterns present today are the result of prior events. This reasoning was evident most commonly when students were presented with tangible materials, such as dirt or rocks, or when students were drawing pictures of structures they had seen or learned about, such as volcanoes or the Earth's interior. Most students were also able to articulate a possible process that could have created a specific observed pattern, although not all students were able to articulate non-anthropogenic processes. That is, only a subset of students recognized that processes occur on Earth that are independent of human action. In addition, a few students articulated more than one possible process, demonstrating that young children are able to generate multiple working hypotheses about Earth processes.

Links Between Observations and Processes

All interviewees (n=21) recognized that processes leave traces, and that these traces can be used to reason about processes themselves. This is the most fundamental component of retrodictive reasoning and was evident in student discourse about both simple and complex phenomena. In addition, students were generally able to articulate reasoning for why a process would create a specific outcome. To demonstrate this, we provide examples related to students' ideas about rock formation as well as more complex ideas about why some rocks are found in seemingly unlikely places. Students most often called upon processes that were familiar from everyday life, and less commonly called upon processes that had clearly not been experienced directly.

Many of the children interviewed about Earth's interior held the previously documented idea that the Earth contains a layer of magma, often with that layer at the Earth's center, and that this layer is the source of lava that is extruded by volcanoes. Interviewees who discussed the Earth's interior all expressed an idea that the center of the Earth is very hot. One fourth grader (JF4) explained that, in addition to other reasons, the inside of the Earth is hot because neither water nor air could get inside the Earth to cool it off. She compared a hot, stuffy room to the Earth's interior, explaining that:

> *JF4: If no air can be in [a] room, it gets horribly hot.*
>
> *Q: Why does it get really, really hot in that room?*
>
> *JF4: Because no fresh air from the outside can get in here and make us feel nice and cool. So that's why.*

This student is clearly calling upon familiar experiences with fire and stuffy rooms in building her explanations about Earth phenomena, and may related to the alternative conception that insulation by itself is a source of heat (e.g., Wiser & Amin, 2001).

Another example of everyday explanations is found in a student's model for the formation of dirt and rocks included in a model in which the rotation of the Earth resulted in the mixing of materials that form dirt. For example, a 1st grade female indicated that dirt forms because "*the earth is spinning so fast, that kind of mixes everything together*" (EF1). This student used the mixing that occurs when she makes cookies as an explanation for why this mixing would result in a solid rock. Although not based on everyday experience, a fourth grade female presented a disconnected set of ideas for how shells might form a rock (Fig. 1A) that included Earth's rotation: "*the weather or the sun or you know how the earth sort of rotates around, maybe they got hardened and that's how this thing was made*" (ZF4). Experience with heat and the Sun's impact on materials at Earth's surface likely also influenced the processes students called upon to explain rock formation. One third-grade student believed that rocks would harden because "*the sun probably was looking at it for a long time, like a lot of heat was on it for a lot of years and then it finally just hardened*" (EF3). A fourth grader (MF4) expressed the partially correct idea about sedimentary rock formation, stating that, "*sand rocks, they're made from sand crushed under the ground for many years*"; it is unclear in this case whether 'crushing' refers simply to pressure pushing sand together (a scientific idea) or whether it indicates the formation of sand from rocks underground (an alternative conception).

Beyond simple explanations of rock formation, students were asked to reason about seemingly contradictory information. A rock containing shells (Fig. 1A) was presented to some students as having been discovered beneath the ground within the interior of the USA. The students were generally adamant that the rock formed near a beach, and also recognized that the interior USA is far from beaches. This cognitive dissonance provided an opportunity for students to reason retrodictively. For example, a third-grade student (OF3) provided an explanation for the contradiction of a shell-containing rock far from a beach that relied upon her prior knowledge:

> *Q: ...How could this [rock in Fig. 1A], that you said came from a beach, end up in the middle of Kansas?*

OF3: Maybe a long time ago…maybe that was formed when the dinosaurs, where the United States wasn't all together…it was probably next to the ocean once, but then it got in while the states formed together to make United States of America. Probably just stayed there until you guys dug it up.

With prompting, OF3 provided a drawing of the process that would result in the "*states form[ing] together to make United States of America*", with each state representing a unique tectonic plate (Fig. 2). She clearly articulated that the rock needed to form near the ocean, and recognized that while an ocean did not currently exist in the region an ocean may have existed in that location in the past. Several researchers have noted that students often mistake continental boundaries for plate tectonic boundaries (e.g., (Marques & Thompson, 1997). The above-mentioned student as well as others in this study similarly suggested that tectonic plates are delineated by non-related external boundaries. In this case, OF3 confounded geopolitical boundaries associated with the borders of states within the USA with physical boundaries separating tectonic plates. Despite this confounding, OF3's hypothesis of tectonic plate movement is a good explanation for the presence of shells at great heights such as in the Alps, although a rock in the USA's interior likely resulted from a different mechanism (an intracratonic ocean).

Figure 2. Drawing of the "State Tectonics" process made by a third-grade student (OF3).

Each landmass, outlined in brown, was drawn as a different state (e.g., Kansas, Minnesota). The brown dot represents the location in the interior of the USA where the interviewer indicated the rock in Fig. 1A was discovered. Blue represents ocean locations during the time of the rock's formation.

A second third-grader (TF3) provided an alternative, human-based cause for the rock's location in the continent's interior. She also recognized the necessity that the ocean must have existed in the rock's place of origin in the past. Her model appears to be based on prior exposure to beach replenishment strategies undertaken by municipalities to combat beach erosion.

Q: …How could this [rock in Fig. 1A], that you said came from a beach, end up in the middle of Kansas?

TF3: …I've heard that they took dirt or sand or stuff and dumped it into the ocean, so the ocean got smaller and smaller and smaller, like somewhere over there, and then it became land. So maybe…they put dirt and sand in the water to make land, like extra land.

Q: So who did that?

TF3: People that were there…not necessarily people today, because they would be long gone by now, but like people that existed a long time ago.

In this exchange, the student still recognizes that the process must have been an ancient one. This requires her to call upon peoples who are "*long gone by now*", rather than requiring the process to be modern. Both of these exchanges, that of OF3 and TF3, suggest that these students are using retrodictive reasoning to combine disparate pieces of information (i.e., state boundaries, ancient peoples, beach replenishment) into a single explanatory model that still allows for earth processes to have occurred in the past.

Presence of Multiple Working Hypotheses

In most cases, interviewees provided only one explanation for their observations or ideas; that is, most of the interviewed students did not offer more than one process that would result in an observed material or pattern. However, a subset of students (n=8) did exhibit the ability to consider more than one explanation for their observations. Most often, these ideas were presented in response to interviewer prompts over time, rather than as a set of possible mechanisms presented in tandem. This is an important distinction in that we cannot always know if these children abandoned one idea before presenting a new idea, or if ideas were truly held as multiple working hypotheses simultaneously.

Returning to the idea of rock formation, a fifth grade girl (OF5) provided two possible mechanisms for the formation of a shell-filled rock (Fig. 1A). Her first explanation revolved around lava mixing with seashells:

Q: You said that the seashells were in the rock; how did the seashells get in the rock?

OF5: Well, maybe if a volcano erupted or something, maybe the lava mixed in the sea shells and became a rock.

The interviewer prompted the student to provide other explanations in an attempt to identify multiple hypotheses. Although many students articulated just one mechanism, OF5 provided a second explanation similar to ideas other students held about the Sun baking rocks:

OF5: Maybe it's clay and seashells hardened.

Q: How would that happen?

OF5: Maybe it was just left out in the sun with sea shells and it just became really hard and you couldn't get the sea shells out of it anymore.

Other students provided multiple explanations without significant prompting. A third-grade boy (JM3) presented a suite of possible mechanisms for the formation of mountains that included meteor impacts, earthquakes, pressure from underground, volcanoes, and wind deposition. This represents many of the common alternative conceptions about mountain formation documented in the literature (e.g., Muthukrishna et al., 1993). In the case of JM3, he articulates quite clearly that different mountains are formed in different ways, clearly demonstrating his ability to work with multiple hypotheses at once. In the following exchange, JM3 is explaining how mountains can form from meteor impacts, volcanoes pushing up from underground, or from the remnants of old volcanoes:

> *Q: You said earlier you think Mount Everest formed when a crater hit the earth; can you tell me more about that?*
>
> *JM3: If it were formed by a crater, it must have been a whole bunch of craters… A crater is a giant hole in the ground that was formed by a meteorite …[the mountains are] in between most of the craters.*
>
> …
>
> *Q: Do you think all mountains form that way?*
>
> *A: No, I don't.*
>
> *Q: How can you form other mountains?*
>
> *A: …something happening underground, so it pushes all the layers of land up so it forms a mountain.*
>
> *Q: What's happening to push that land up underground?*
>
> *A: Maybe it was when a volcano was erupting, or many at the same time. Or maybe it was just an old volcano that was made in the sea and then after years, it wasn't a volcano anymore, because the rocks change over time, so it becomes a mountain instead of a volcano.*

The student did not change his mind when subsequently probed, but rather stuck with the idea that multiple processes can produce the same effect. This is indicative of reasoning with multiple hypotheses.

Preference for One Hypothesis Over Others

As explained above, all students possessed the ability to identify processes that must have occurred to produce modern observations, and some students were able to generate and hold at once multiple explanatory processes. However, students generally did not reason about which possible cause was the most likely cause, an important aspect of retrodictive reasoning, and perhaps the most difficult. Students were comfortable providing explanations for their own ideas, and were willing to argue with others. However, we were unable to identify any discourse in which these children rationalized a preference for one idea over another. This may suggest that elementary-aged children see no need to choose one hypothesis over another or may simply be an artifact of the type of interviews conducted here.

Reasoning References Time Beyond Human Timescales

Almost every student (n=19) explicitly discussed the importance of time in their explanations. Students generally recognized that geologic processes take a long time, although the exact nature of "long" was unclear. Because elementary children are unlikely to understand the meaning of specific large numbers, we chose to focus on relative temporal descriptions (i.e., long, short) rather than on absolute ages. In essence, we believe that the absolute numbers stated by interviewees were generally meaningless and should not be over interpreted. As an example, we consider the discussion of how long students thought it would take for dirt to form. This question was posed after a discussion of the formation of dirt from solid rock, initiated either by the interviewee or as an interviewer prompt. Not surprisingly, students' perspective about how long this process might take ranged from a few days to thousands of years. Interestingly, students' ideas were generally well aligned with the actual amount of time it takes for soils to form. We suggest that this does not indicate strong conceptual understanding of the temporal nature of soil formation, but rather that soil happens to form over timespans that align quite well with the numbers the interviewed students happened to know.

Although humans can impact the Earth system, as evidenced by modern climate change, most Earth processes occur without human intervention and certainly the vast majority of Earth history passed before humans evolved. As noted above and by other researchers (e.g., Blake, 2005), some students were unable to recognize that humans play almost no role in most Earth processes. A particularly good example of this is reflected in a discussion between the interviewer and a first-grade student (LM1). In this exchange, the student has explained that a rock looks like it has been "*knocked around*":

> *Q: How would a stone like this get knocked around?*
>
> *LM1: Humans kicking it, maybe. It came out of a volcano, or something, and it'll hit and crash apart because it came down with such force.*
>
> *Q: Is there any other way it can get knocked around?*
>
> *LM1: Yeah, by humans kicking it.*

This student is clearly most comfortable with a human cause for the rock's movement, although a geologic event (volcano) is mentioned in passing. This is similar to TF3 explaining shells found in the interior USA through humans filling up the ocean with sand.

The idea that humans cause some geologic events is not limited to the youngest children. For example, a fifth grade student (TF5) is discussing dirt as an unchanging material:

> *Q: Is the dirt in the dinosaur period the same dirt that we find today?*
>
> *TF5: Yes, I think so. It's possible for dirt to live that long, because…Like people, they will get killed and stuff, and they might just die. And plants, anybody could smoosh it, or a tree, they could cut down. But dirt's so small, nobody would really want to do anything to it. And if they do anything, nothing would really happen. You can't kill dirt. It's just there forever.*

This student views at least some aspects of the Earth as being static and unchanging. This discourse also reflects a notion that humans must be involved in changing or destroying dirt. Generally, those students who called upon human activities to explain changes, or lack of changes, to the Earth were least likely to recognize the importance and scale of geologic time. However, most of the young children interviewed here recognized the importance of time in creating Earth phenomena, as evidenced by the data presented in preceding sections, suggesting that deep time in an abstract sense is not outside the reach of young children.

Conclusion

The children interviewed here showed themselves to be remarkably capable of some aspects of retrodictive reasoning, despite the presence of a number of interesting alternative conceptions about Earth processes. All interviewed students present evidence of an understanding that patterns present in the modern world are the imprints of processes that have already occurred, and most also explicitly reasoned about geologic time. This reasoning was more nuanced than the simple idea that processes take time to occur. Rather, these young students recognized that geologic processes resulting in modern features generally occur in the past and often before humans, or at least modern humans, were living in the same areas as modern features. This understanding of the relationship between modern observations and past events is the first prerequisite for retrodictive reasoning, and its presence indicates that young children are capable of making this type of inference about the past from modern evidence.

Children in this study were often able to describe specific events that might result in modern features and demonstrated the ability to use a variety of knowledge types in the context of their retrodictive reasoning. In some cases, children were drawing analogies from their own observations. This is particularly well exemplified by the young student who reasoned that the hot center of the Earth might result from a lack of air, as might occur in a stuffy room on a hot day. Similarly, an analogy of ingredients mixed together with a beater to form cookie dough was used to support the idea that the Earth's rotation results in materials mixing together to form rocks. These analogies clearly result from everyday experience. Everyday experiences often aligned with domain-general types of information, such as the idea that rocks will harden in the Sun; we suggest that this notion comes from physical experience with materials, such as mud, drying and hardening, rather than from instruction.

In other cases, such as with the model of states as tectonic plates or people infilling the ocean, the child was clearly pulling from an idea they had been taught in school or heard from the media, an authority, or friend. Most of these ideas were quite domain-specific. For example, students had very specific ideas about the ways in which mountains form that were unlikely to derive from everyday observations, such as the idea that meteors generate craters that are mountains. While children might experiment in the classroom with craters in simple experiments using for example flour and golfballs, meteor impacts are not likely a phenomenon directly observed by these students. The role of everyday experience and learned ideas in influencing a child's ability to reason retrodictively warrants further research. Most importantly, the ways in which everyday experience could be used to encourage retrodictive reasoning should be explored.

The presence of alternative conceptions about Earth systems, of which there were many documented in even this relatively small number of interviews, did not seem to interfere with the ability to reason retrodictively. Certainly, alternative conceptions will interfere with the ability to ascertain an accurate reason for Earth features and processes. However, the ability to reason despite significant alternative conceptions suggests that retrodictive reasoning in young students can be fostered even when children are cognitively unable to grasp some of the more complicated principles underlying strong scientific literacy. Gaining deep understanding of the fundamental laws and principles that govern the Earth system is vital for the reasoning, both retrodictive and predictive, that is needed for decision-making about human impacts on Earth.

• • •

Acknowledgements

We thank the entire Science Media Group team (including, notably, Alex Griswold) as well as all teachers and students who participated in this study. We especially thank our interviewees for sharing their ideas and passion for learning. An anonymous reviewer also provided helpful suggestions that significantly improved our discussion. This work was supported in part by the Annenberg Foundation through Annenberg Media.

Julie LIBARKIN is an Associate Professor of Geological Sciences at Michigan State University and Director of the Geocognition Research Lab. She specializes in assessment of student ideas and communication of science, with a specific focus on Earth Science.

Matthew SCHNEPS is the Director of the Laboratory for Visual Learning at the Harvard-Smithsonian Center for Astrophysics. His interests include science education, student ideas, educational media, learning differences, and cognitive psychology, with a focus on non-verbal learning.

References

Argow, B., Reilly, J., & Schneps, M. (2004). *Essential science for teachers: Earth and space science*. Harvard-Smithsonian Center for Astrophysics. Science Media Group.

Ault, C. R. (1982). Time in geological explanations as perceived by elementary-school students. *Journal of Geological Education*, *30*(5), 304–09.

Ault, C. R. (1984). The everyday perspective and exceedingly unobvious meaning. *Journal of Geological Education*, *32*(2), 89–91.

Ault, C. R. (1998). Criteria of excellence for geological inquiry: The necessity of ambiguity. *Journal of Research in Science Teaching*, *35*(2), 189–212.

Barrow, L., & Haskins, S. (1996). Earthquake knowledge and experiences of introductory geology students. *Journal of College Science Teaching*, *26*(2), 143–46.

Blake, A. (2005). Do young children's ideas about the Earth's structure and processes reveal underlying patterns of descriptive and causal understanding in earth science? *Research in Science & Technological Education*, *23*(1), 59–74. 0

Cheek, K. A. (2010). Commentary: A summary and analysis of twenty-seven years of geoscience conceptions research. *Journal of Geoscience Education*, *58*(3), 122–134.

Dahl, J., Anderson, S. W., & Libarkin, J. C. (2005). Digging into Earth Science: Alternative conceptions held by K-12 teachers. *Journal of Science Education*, *12*, 65–68.

Dal, B. (2007). How do we help students build beliefs that allow them to avoid critical learning barriers and develop a deep understanding of geology? *Eurasia Journal of Mathematics, Science and Technology Education*, *3*(4), 251–269.

DeLaughter, J. E., Stein, S., Stein, C. A., & Bain, K. R. (1998). Preconceptions abound among students in an introductory earth science course. *Eos, Transactions American Geophysical Union*, *79*(36), 429–432.

Dodick, J., & Orion, N. (2003). Cognitive factors affecting student understanding of geologic time. *Journal of Research in Science Teaching*, *40*(4), 415–442.

Dove, J. E. (1998). Students' alternative conceptions in Earth science: a review of research and implications for teaching and learning. *Research Papers in Education*, *13*(2), 183–201.

Gosselin, D. C., & Macklem-Hurst, J. L. (2002). Pre-/post-knowledge assessment of an earth science course for elementary/middle school education majors. *Journal of Geoscience Education*, Papers in Natural Resources, *50*(2), 169–175.

Happs, J. C. (1984). Soil genesis and development: Views held by New Zealand students. *Journal of Geography*, *83*(4), 177–180.

Kali, Y., Orion, N., & Eylon, B.-S. (2003). Effect of knowledge integration activities on students' perception of the earth's crust as a cyclic system. *Journal of Research in Science Teaching*, *40*(6), 545–565.

King, C. (2008). Geoscience education: an overview. *Studies in Science Education*, *44*(2), 187–222.

Kortz, K. M., & Murray, D. P. (2009). Barriers to college students learning how rocks form. *Journal of Geoscience Education*, *57*(4), 300–315.

Kusnick, J. (2002). Growing pebbles and conceptual prisms - Understanding the source of student misconceptions about rock formation. *Journal of Geoscience Education*, *50*(1), 31–39.

Lawton, J. (2001). Earth System Science. *Science*, *292*(5524), 1965–1965. doi:10.1126/science.292.5524.1965

Lewis, E. B., & Baker, D. R. (2009). A call for a new geoscience education research agenda. *Journal of Research in Science Teaching*, *47*(2), 121-129.

Libarkin, J., Anderson, S., Dahl, J., Beilfuss, M., Boone, W., & Kurdziel, J. (2005). College students' ideas about geologic time, Earth's interior, and Earth's crust. *Journal of Geoscience Education*, *53*(1), 17–26.

Libarkin, J. C., & Kurdziel, J. P. (2006). Ontology and the teaching of earth system science. *Journal of Geoscience Education*, *54*(3), 408.

Lillo, J. (1994). An analysis of the annotated drawings of the internal structure of the Earth made by students aged 10–15 from primary and secondary schools in Spain. *Teaching Earth Science*, *19*, 83–87.

Marques, L., & Thompson, D. (1997). Misconceptions and conceptual changes concerning Continental Drift and Plate Tectonics among Portuguese students aged 16-17. *Research in Science & Technological Education*, *15*(2), 195–222.

Montangero, J. (1996). *Understanding Changes In Time: The Development Of Diachronic Thinking In 7-12 Year Old Children*. Taylor & Francis.

Muthukrishna, N., Carnine, D., Grossen, B., & Miller, S. (1993). Children's alternative frameworks: Should they be directly-addressed in science instruction? *Journal of Research in Science Teaching*, *30*(3), 233–248.

Nelson, B. D., Aron, R. H., & Francek, M. A. (1992). Clarification of selected misconceptions in physical geography. *Journal of Geography*, *91*(2), 76–80.

Norris, S. P., Guilbert, S. M., Smith, M. L., Hakimelahi, S., & Phillips, L. M. (2005). A theoretical framework for narrative explanation in science. *Science Education*, *89*(4), 535–563.

Pyle, E. J. (2008). A model of inquiry for teaching Earth Science. *Electronic Journal of Science Education*, *12*(2). Retrieved from http://ejse.southwestern.edu/article/view/7770

Ross, K. E. K., & Shuell, T. J. (1993). Children's beliefs about earthquakes. *Science Education*, *77*(2), 191–205.

Schoon, K. (1995). The origin and extent of alternative conceptions in the earth and space sciences: A survey of pre-service elementary teachers. *Journal of Elementary Science Education*, *7*(2), 27–46.

Sunderlin, D. (2009). Integrative mapping of global-scale processes and patterns on "imaginary Earth " continental geometries: A teaching tool in an Earth History course. *Journal of Geoscience Education*, *57*(1), 73–81.

Thomson, N., & Chapman Beall, S. (2008). An Inquiry Safari: What can we learn from skulls? *Evolution: Education and Outreach*, *1*(2), 196–203.

Trend, R. (1998). An investigation into understanding of geological time among 10- and 11-year-old children. *International Journal of Science Education*, *20*(8), 973–988.

Trend, R. (2000). Conceptions of geological time among primary teacher trainees, with reference to their engagement with geoscience, history, and science. *International Journal of Science Education*, *22*(5), 539–555.

Trend, R. D. (2001). Deep time framework: A preliminary study of U.K. primary teachers' conceptions of geological time and perceptions of geoscience. *Journal of Research in Science Teaching*, *38*(2), 191–221.

Trend, R., Everett, L., & Dove, J. (2000). Interpreting primary children's representations of mountains and mountainous landscapes and environments. *Research in Science & Technological Education*, *18*(1), 85–112.

Wiser, M., & Amin, T. (2001). "Is heat hot?" Inducing conceptual change by integrating everyday and scientific perspectives on thermal phenomena. *Learning and Instruction*, *11*(4–5), 331–355.

Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review*, *20*(1), March 2000, 99-149.

International Electronic Journal of Elementary Education, 2012, 5(1), 63-78.

The Role of CLEAR Thinking in Learning Science from Multiple-Document Inquiry Tasks

Thomas D. GRIFFIN*
University of Illinois at Chicago, Chicago, IL, USA

Jennifer WILEY
University of Illinois at Chicago, Chicago, IL, USA

M. Anne BRITT
Northern Illinois University, Dekalb, IL, USA

Carlos R. SALAS
University of Illinois at Chicago, Chicago, IL, USA

Received: 21 August 2012 / Revised: 6 November 2012 / Accepted: 6 November 2012

Abstract

The main goal for the current study was to investigate whether individual differences in domain-general thinking dispositions might affect learning from multiple-document inquiry tasks in science. Middle school students were given a set of documents and were tasked with understanding how and why recent patterns in global temperature might be different from what has been observed in the past from those documents. Understanding was assessed with two measures: an essay task and an inference verification task. Domain-general thinking dispositions were assessed with a Commitment to Logic, Evidence, and Reasoning (CLEAR) thinking scale. The measures of understanding were uniquely predicted by both reading skills and CLEAR thinking scores, and these effects were not attributable to prior knowledge or interest. The results suggest independent roles for thinking dispositions and reading ability when students read to learn from multiple-document inquiry tasks in science.

Keywords: Thinking Dispositions, Learning From Text, Climate Change, Earth Science, Multiple-Document Inquiry Tasks.

Introduction

The internet has become a primary means by which people search for information to answer many science-related questions. Adults read internet sources to help them understand phenomena in the world around them. They read to learn about the development of new

* Thomas D. Griffin, Department of Psychology, University of Illinois at Chicago, 1007 W. Harrison St (M/C 285), Chicago, IL, 60607 USA, Phone: (312) 996-8190, E-mail: tgriffin@uic.edu.

ISSN:1307-9298

technologies. They read to form beliefs on policy issues as well as to inform health-related decision-making. To form an understanding, they need to integrate information that is spread out across numerous documents and sources. Hence, the ability to learn about scientific phenomena from multiple documents is a critical skill for life-long learning. Yet, instruction in this area is not well represented in science classrooms, and research identifying potential sources of variance in how well students can engage in multiple-document inquiry learning on science topics has been limited. The main goal for the current study was to investigate whether individual differences in either reading skills or domain-general thinking dispositions might affect learning from multiple-document inquiry tasks in science.

What Processes are needed to Learn from Multiple-Document Inquiry Tasks?

One reason why learning from multiple documents is so complex is because it requires all of the processes necessary for comprehending individual informational texts, plus an additional set of processes that become particularly important when readers are confronted with information from more than one text. According to theories of text comprehension (Kintsch, 1998; Kintsch & van Dijk, 1983), understanding even a single informational text requires the construction of several levels of representation. At the lowest level, a reader creates a surface representation, which generally consists of a fleeting episodic trace capturing the exact words and format of the text. At the next level of processing, the reader attempts to develop the text-based representation. This is essentially a propositional representation of the ideas presented in each clause or sentence. Basic word and sentence-level reading processes contribute to the construction of this text-based representation. In addition, to learn from informational text, the reader must attempt to develop yet another level of representation, referred to as the situation model by Kintsch. On this level, the reader attempts to connect ideas between the sentences and with prior knowledge to develop a coherent understanding of the content that is being described. When the goal for reading informational science texts is to develop an understanding of how or why a phenomenon occurs, then the situation model can be thought of as a causal chain or mental model of the phenomena being described (Kintsch, 1994; Trabasso & van den Broek, 1985; Wiley, Griffin & Thiede, 2005).

Yet, in many learning situations, readers are presented with more than a single text from which to obtain information (Britt, Perfetti, Sandak, & Rouet, 1999; Perfetti, Rouet, & Britt, 1999). One framework for describing the cognitive processes involved in multiple-document comprehension is the MD-TRACE model (Multiple Documents - Task-based Relevance Assessment and Content Extraction) proposed by Rouet and Britt (2011). According to the MD-TRACE model, students begin multiple-document reading by creating an interpretation of the task (called the Task Model). This Task Model includes the goals and subgoals for reading (e.g., why is the text being read? what is the question to be answered? what does developing an argument or explanation entail?) and plans to reach those goals (e.g., find evidence or causes). In other words, the task model includes the goals for reading and the basic steps that should be taken to achieve the desired outcome. Depending on the reading context, including the reader's interpretation and the instructions that are given, task models will range from cursory to clearly delineated. Although reading goals may affect learning in single text contexts, they become even more critical to consider in multiple-document contexts.

Learning from multiple documents instead of a single document also requires another level of representation (a Documents Model) that captures the relation of information across the document set, and information about each document, in addition to the representations of the content of individual texts. The Documents Model (Britt, Perfetti, Sandak, & Rouet, 1999;

Perfetti, Rouet, & Britt, 1999) has been proposed to capture these two needs. One part of a Documents Model, the Intertext Model, includes information about the sources of the various documents (e.g., who wrote it), and notes relations among the documents (e.g. the presence of corroborating or conflicting statements). The Documents Model also contains the Integrated Model, which serves as the representation of the situation or phenomena described across documents. According to the MD-TRACE model, the extent to which readers develop Integrated Models or Intertext Models from multiple-document contexts will be partially determined by the goals readers have in their Task Model.

With a single text, comprehension can be driven by a text's intended purpose, structure, or argument. With multiple texts, the reader must impose selection and organization in order to form a model that integrates the information from different texts (rather than simply constructing distinct models of each text). A reader's goals guide the process of reading and evaluating the individual texts, selecting relevant information, and reassembling what is selected into a new coherent model. Thus, the interpretation of the task and the goals a reader sets for reading are a critical determinant of multiple-document comprehension.

What Leads to Successful Learning from Multiple-Document Inquiry Tasks in Science?

A burgeoning area of investigation at the intersection of literatures on subject-matter learning and learning from informational texts has been exploring what conditions facilitate student understanding from multiple-document inquiry activities in science (Braten, Britt, Stromso,& Rouet, 2011; Cerdan & Vidal-Abarca, 2008; Goldman, Braasch, Wiley, Graesser, & Brodowinska, in press; Mason, Boldrin, & Ariasi, 2010; Sanchez, Wiley & Goldman, 2006; Wiley, Ash, Sanchez & Jaeger, 2011; Wiley, Goldman, Graesser, Sanchez, Ash, & Hemmerich, 2009). There are numerous factors that are likely to impact learning from multiple documents. These factors can include features of the set of sources that are provided, as well as the nature of the inquiry task that is given. One general approach within this literature has been to provide students with a set of informational texts as reading material, often through the guise of the results of an internet search. As a goal for the processing of the informational texts, students are generally tasked with learning about how or why a phenomenon occurred such as "What caused the eruption of Mt. St. Helens?" or "How do bacteria resist the effects of antibiotics and which biological mechanisms explain this phenomenon and its transmission to other bacteria?" or "What caused the extinction of dinosaurs?" Similar to studies on learning from multiple sources in history (i.e. Wiley & Voss, 1996; 1999), when students are prompted to use the text sets to generate a causal argument or explanatory model of a phenomenon, it results in better learning from the activity. For example, Cerdán and Vidal-Abarca (2008) found that prompting students to read texts in order to explain how resistance to antibiotics develops resulted in a deeper and more integrated understanding than did asking students very specific questions that could be answered by searching for, finding, memorizing, and reproducing isolated bits of information within the text set. These results are generally consistent with the idea that students' understanding may benefit from multiple-document activities to the extent that students engage in constructive processing that builds connections across ideas in order to form a coherent, integrated model of the phenomena (Britt & Rouet, 2012; Wiley & Voss, 1996; 1999).

Yet, further research has demonstrated that not all learners take advantage of this opportunity, especially when the reading material requires selective use of information (Wiley, Ash, Sanchez & Jaeger, 2011). For example, in Wiley et al (2009) which provided students with texts from both reliable and unreliable sources about volcanic eruptions, the ability to evaluate the sources was seen as a gatekeeper to the development of an accurate mental model of the phenomena. In the first study, better learning was related to the ability to differentiate reliable from unreliable sources. In the second study, the presence of a pre-inquiry instructional module on source evaluation was manipulated (see also Sanchez, Wiley, & Goldman, 2006). In this module, called "SEEK", students were taught to evaluate reliability of each document by considering not only the source of each document, but also whether any evidence was presented that could be related to an explanation of the phenomenon prompted by the inquiry question, and also how information in the document related to other knowledge about the phenomena. Participants completed the SEEK module on an unrelated topic prior to the inquiry task. Thus, this instructional manipulation stressed the need to consider the reliability of Source information as well as the importance of thinking about Explanations, Evidence, and integration with both prior Knowledge and the information in other documents. The main result of this study was that students given SEEK instruction demonstrated better learning from the subsequent multiple-document inquiry task on volcanic eruptions. Together these two studies demonstrated that multiple-documents learning is predicted by individual differences in evaluating sources and content of the documents, and that learning is improved by instruction that targets skills related to evaluation based on source, evidence, and coherence with other information.

Additional lines of investigation have further explored what the better learners were doing in the first Wiley et al (2009) study by using eyetracking and think-aloud methodologies. Wiley, et al. (2011) found that when students were asked to write an argument explaining the causes of volcanic eruptions, the best learners showed more selective reading behaviors. These participants were more likely to skim pages, but would often go back and thoroughly re-read a page if it contained conceptually relevant information. Eyetracking data showed that these better learners were also more likely to spend a greater proportion of their time on the specific sentences of a relevant page that (i.e., regions of interest, ROIs) were most critical for forming an explanation. Lastly, while all readers looked at the illustrations that accompanied the texts, the better learners tended to look at the conceptual images more so than decorative images. These findings suggest that the students that were more goal-directed, strategic, and selective in their reading and use of available information learned more and created better causal explanations. These better learners seemed to be more engaged in the process of creating an integrated mental model, as they showed better use of conceptual illustrations, and more integrated reading patterns. Importantly, the best learners were not simply spending more overall time reading—but instead they were more selective with their reading efforts. For the best learners, it appears that they responded to an argument-writing task by directing their attention to the most relevant information for the construction of an accurate mental model or explanation. An additional finding in this study is that when students were given an instruction to write a report rather than an argument, they were generally less selective in which information they read and included in their essay.

The findings of the think-aloud study (Goldman, et al. in press) where readers were asked to simply think aloud about what they were doing while viewing the documents, also suggested that better learners were more selective in what they read and how they utilized the information. Better learners made more comments related to their evaluation of source credibility and reliability, especially in relation to why they went to certain pages and not others or why they were leaving a page before they read all of its content. Their comments revealed a more strategic approach to reading in which they referenced their inquiry task, their current understanding, and what they still needed to accomplish. Consequently, they incompletely read pages that they judged would not further their understanding and finished pages that they judged would further their understanding. In addition, the better learners were more selective in how they used the information. They engaged in more sensemaking, self-explanation and comprehension-monitoring processes while they were reading reliable sites than they did on unreliable websites. These behaviors are usually related to incorporating information into mental models. Self-explanation helps in both the construction of models (Chi, 2000) and as a source of cues for evaluating the quality and completeness of one's own mental model (Griffin, Wiley, & Thiede, 2008). Thus, these findings suggest that better readers were more selective and integrated the more reliable information into their mental models.

Interestingly, the think-aloud evidence available from this study does not suggest that the better learners had better a priori knowledge of which sites might be considered more reliable and useful. Rather, the think-aloud comments suggested that readers assessed whether their understanding was increasing and what additional information was needed to achieve the inquiry goal. This type of evaluation of the sites in terms of goal-relevant content resulted in learners being more strategic in their reading and spending a greater proportion of their time on the more reliable sites, hence resulting in better learning from the inquiry task.

Other work on learning from internet-inquiry tasks also suggests that differences in how students approach the evaluation of information quality can affect their learning. Following an internet inquiry task on dinosaur extinction, Mason, Boldin, and Ariasi (2010) asked learners how they decided which information they found on the internet was true. They found that students who were more likely to appeal to comparisons of information from multiple sources and to scientific evidence as a basis for evaluation were those who were more likely to learn the scientifically accepted conception of extinction from the activity. This study also administered the Conley, Pintrich, Vekiri and Harrison (2004) self-report measure of epistemic beliefs about science. They found that scores on the Justification subscale were related to whether students learned the scientifically accepted view of extinction. This subscale assesses beliefs about the nature of science and the importance science places on experiments, replication, and the source of scientific ideas. Similarly, other work from Braten and Stromso (2010) has also shown that some features of epistemological beliefs (specifically about the to-be-learned topic such as climate change) can predict who might learn most effectively from multiple-document inquiry tasks, and that readers who engage in more source evaluation behaviors develop better understanding from multiple documents (Braten, Stromso & Britt, 2009).

Together, these results suggest that multiple-document inquiry tasks provide the opportunity for readers to engage in more integrative processing and model construction. Yet, the extent to which readers are able to take advantage of this opportunity depends on whether readers selectively process and integrate the most reliable, central, and relevant information. Several lines of research suggest that individuals differ in how they approach inquiry tasks and whether they engage in evaluation, selective reading and integration. The reviewed research suggests that beliefs about science matter, as do instructions to use the information in order to form causal arguments and instructions to evaluate evidence, relevance, and source reliability. Individual differences in general thinking dispositions may be a promising source of influence on the likelihood that a person will engage in successful learning from multiple documents.

Delineation of two classes of individual differences: Capacities and Dispositions.

Expanding on the work of reasoning theorists such as Baron (1985), Stanovich and West (1997) delineated two classes of individual differences, capacities and dispositions that may be useful to consider in this context. What and how a person does on any cognitive task is determined by a combination of what they are capable of doing and what they are disposed towards doing. Capacity constraints have been a primary focus within Cognitive Psychology. These range from more basic processing abilities and constraints that are not likely teachable, such as working memory capacity, to more teachable factors including reading skills such as decoding and word knowledge. Individual differences in capacity constraints would be expected to play a role in learning from multiple documents just as they would in learning from single passages. Although rarely included in studies on learning from multiple documents, reading ability has been shown to be a significant predictor of multiple-document comprehension in at least one study (Mason, Boldrin, & Ariasi, 2010).

In contrast, dispositional individual differences are those that relate to a person's goals, their orientation towards the task, and their willingness towards applying whatever relevant skills and capacities they have to the processes required for effective thinking, learning, and task performance. Griffin and Ohlsson (2001) showed that people vary in whether they report forming their belief on a topic in terms of considering relevant evidence versus deferring to their affective preferences, and this in turn predicts people's willingness to revise their belief in light of new evidence. More recently, Griffin (2008) reported findings which suggest that people have a general disposition towards whether they consider evidence or affect when forming their beliefs. People reported the extent to which they based their beliefs on either considering evidence or relying upon faith. The beliefs varied across eight different topics that were both religious and non-religious. The topics were largely unrelated in content as indicated by the fact that *what* a person believed on each topic did not predict what they believed on the other topics. However, the degree to which people relied upon faith versus evidence to arrive at whatever belief they held on a topic was correlated with their reliance on faith versus evidence for all other topics. This pattern of consistency across distinct topics implicates a general thinking disposition relating to intellectual values.

A reader's goals when attempting to learn from multiple documents might be affected by many factors, including their interest in the topic, as well as by general intellectual values. In a multiple-documents inquiry context, effective learning requires engaging in evidence-based reasoning in the service of argument construction. A general disposition of valuing evidence-based thinking in the evaluation of beliefs and claims would seem to orient one toward attempting to construct the kind of coherent argument that such an inquiry task requires. Thus, having an evidence-based disposition may play a unique role above and beyond learners' capacities and skills in determining whether they engage in the extra-textual processing required for developing an integrated model of phenomena across texts as required by multiple-document inquiry tasks.

In previous studies on domain-general thinking dispositions, Stanovich and colleagues have shown that a general disposition towards actively open-minded thinking (AOT) predicts cognitive performance on higher-order thinking tasks over and above measures of cognitive capacity, such as the SAT, Raven's Matrices, and Nelson-Denny Reading Comprehension Test (e.g., Stanovich & West, 1998; for a recent review see Stanovich, 2012). This research has employed various versions of a 41-item AOT scale comprised of several subscales. Some items tap moral authoritarianism and openness to others' values (e.g., "I believe we should look to our religious authorities for decisions on moral issues."; "There are a number of people I have come to hate because of the things they stand for."). However, other items focus upon a general openness to intellectual inquiry, evidence, and belief revision (e.g., "People should always take into consideration evidence that goes against their beliefs."; "One should disregard evidence that conflicts with your established beliefs."). The present research is interested in these latter items assessing a disposition that is more directly related to the kind of evidence-based thinking that should impact a multiple-documents inquiry task in science. The AOT scale has typically been employed as a composite with a single score that is used to predict performance on higher-order cognitive tasks assessing logical reasoning, rational judgment, normative decision making, and informal reasoning processes such as syllogistic reasoning, probabilistic reasoning, statistical reasoning, covariation detection, and argument evaluation (Stanovich & West, 1997; 1998). Individuals with high AOT scores evaluate objective argument quality more accurately than those with lower AOT scores, and their evaluations are less biased by consistency with prior beliefs, even when controlling for cognitive ability (Stanovich & West, 1997; West, Stanovich, & Toplak, 2008). Sá, Kelley, Ho, and Stanovich (2005) found that people low in AOT were more likely to generate arguments that simply reiterated their personal theory rather than providing supporting evidence. Identifying and incorporating relevant information across multiple texts is likely to be impacted by some of the same factors that impact the kind of argument evaluation and construction tasks employed in these studies.

Although most of the studies on AOT have involved adults, one study has found that thinking dispositions can predict performance on several standard reasoning tasks with children (Kokis et al., 2002). However, none of the prior studies have used the AOT measure to examine the effects of thinking dispositions in a classroom learning task where the goal is to acquire knowledge and understanding in a content area. Prior research shows that this thinking disposition impacts performance on tests of one's reasoning proficiency, but has not examined the impact on the content learning and knowledge acquisition that partially depend on such skills. Thus, it is informative to examine individual differences in this thinking disposition in a real classroom context where the goal of the task is science learning via selective integration of information across multiple information sources and inclusion of that information in the form of an explanatory argument.

The Present Study

The purpose of the current study was to explore the effects of individual differences in both capacity and dispositional constraints on learning from a multiple-document inquiry task in science. Middle school students were given a set of documents about the global temperature system and were asked to write an essay explaining how and why recent patterns in global temperature are different from what has been observed in the past. Understanding of the science topic was assessed both by considering the quality of the essays that were written as well as by performance on an inference verification task. The main question for the current study was whether individual differences in domain-general thinking dispositions might have unique effects from reading skill on the understanding that results from a multiple-document inquiry task in science.

Method

Participants

Participants in this study were 59 seventh grade students from 3 science classes in an urban public middle school in the United States. The average age was 13.31 years (*SD* = .64). The sample was 57% female. Self-reported ethnicity was 22% Hispanic, 27% African American, 10% Asian, 59% White, 30% Native American/Pacific Islander and 25% Other. (Students were able to select multiple ethnicities and 5 students did not select any.)

Materials and Measures

Global Temperature Document Set. All participants were given a set of 7 documents containing information related to the causes of global temperature change, based on material that has been used in previous studies with older students about the causes of Ice Ages (Sanchez & Wiley, 2006; Sanchez & Wiley, 2009). Five text-based documents covered several main topics including Ice Ages, the Carbon Cycle, The Greenhouse Effect, Solar Radiation, and Energy from Fossil Fuels. The document set also included a graph of CO^2 Concentrations over the last 400,000 years, presented as its own document. In addition, students were provided with seventh document, titled "Changes in Global Temperatures", which provided textual background on the methods used to assess global temperatures. This document also included a graph of average global temperatures over the last 400,000 years, and a second graph showing the increases in average global temperatures from 1870 to 2010.

The texts were excerpted from several online sources from the United States Geological Survey, the Public Broadcasting Service, the NASA earth observatory, the Environmental Protection Agency, as well as an extension module from an earth science textbook series (Bennington, 2009). To adapt the texts for younger grade levels, vocabulary and sentence structures were simplified. The final text-based documents were on average 326 words long (range: 208-475), with an average in Flesch Reading Ease of 62.36, and an average Flesch-Kincaid grade level of 7.9. The documents were presented to students on pieces of paper contained in a pocket folder, with each of the 7 documents printed on a separate page.

Inquiry Task Essay Prompt and Essay Coding. One main source of information about student understanding was the essays that students wrote in response to the inquiry prompt. Students were asked to "use this set of documents to write an essay explaining how and why recent patterns in global temperature are different from what has been observed in the past." Student responses to this essay prompt were evaluated for the presence of 5 critical target concepts that directly relate to recent changes in global temperature, and thus address the inquiry question students were asked. These concepts were:

1. We are in an unusually long warming period.
2. CO^2 levels in the atmosphere are at their highest in at least 400,000 years.
3. Fossil fuel burning releases CO^2.
4. CO^2 is a greenhouse gas.
5. Greenhouse gases in the atmosphere cause warming.

All essays were evaluated for the presence of the target concepts by two independent coders, who produced a high level of interrater reliability (Krippendorf's $\alpha = .90$, $p < .05$). Any differences were resolved through discussion.

Inference Verification Task. As another measure of student understanding, a sentence judgment task was created in which students were asked to indicate which of a list of statements seemed true based on the texts they had just read. This test (based on Sanchez & Wiley, 2006; Wiley & Voss, 1999) consisted of 18 statements that represented potential connections or inferences that could or could not be made based on the information in the document set. Some example items are "In the past 100 years, both fossil fuel use and CO^2 levels have increased" and "Increases in fossil fuel use increase the amount of heat that escapes into space." The first is an example of a conclusion that is supported by the documents but requires connections across documents. The second is an example of a statement that is false based upon connecting multiple ideas across documents, namely ideas 3, 4, and 5 that were coded in the essays. The items represented 8 correct and 8 incorrect inferences. For every correct and incorrect inference appropriately identified, the students received a single point. An overall proportion score was computed for the task, and higher levels of performance indicated better understanding of the inferences that could be made from the documents.

CLEAR Thinking Scale. Students' CLEAR thinking refers to their Commitment to Logic, Evidence, and Reasoning. The 5-item scale assesses the extent to which students place value and importance on reasoning about evidence when forming and revising beliefs. The construct is measured at the most domain general level. The scale incorporates items from the flexible thinking scale (Stanovich & West, 1997) and the belief identification scale (Sa', West, & Stanovich, 1999), which were revised by Kokis et. al.,(2002) to be used with children. The items were selected based upon the criteria that they directly ask about belief revision in the face of new evidence or information.

The items used for the CLEAR Thinking Scale were:

1. I never change what I believe in - even when someone shows me that my beliefs are wrong.
2. People should always consider evidence that goes against their beliefs.
3. It's important to change what you believe after you learn new information.
4. People shouldn't pay attention to evidence that contradicts their strongly held beliefs.
5. To decide what is true, you often have to ignore your emotions and stick just to the evidence.

Students were asked to respond to these items using a 1-6 scale with 1 meaning Strongly Disagree and 6 meaning Strongly Agree. The scale was scored by subtracting the average rating for the negatively worded items (1 and 4) from the average of the positively worded items. A difference score between the weighted averages means that the combined influence of the positively worded items on the total score is the same as the combined influence of the negatively worded items. This avoids the problem of giving more weight to positively worded items as a group, which creates a response bias effect such that people who simply anchored all of their ratings at a higher value would receive a higher score.

Descriptive Student and Teacher Surveys. Due to the Family Educational Rights and Privacy Act (FERPA), official student standardized test scores could not be obtained. To obtain measures of reading skill in lieu of test scores, a teacher survey asked teachers to indicate each student's level of reading skill relative to their grade level as low, medium or high. A student self-report survey was created to collect basic descriptive information including gender, date of birth, and ethnicity. Students were also asked to rate on a 1-to-5 scale their level of interest in science, interest in the topic, and prior knowledge about the topic.

Procedure

Students participated in the inquiry activity as part of their normal science classes. All materials for the inquiry activity were distributed to students in folders, including the inquiry task essay prompt, blank writing pages and the document set. Students were asked to read along as the inquiry task essay prompt instruction was read out loud.

The full instructions for the Reading and Writing task were:

The primary purpose of reading in science is to understand the causes of scientific phenomena. This means your goal for reading is to understand how and why things happen. To reach an understanding of a new topic in everyday life, we often need to gather information from multiple sources. In today's task your goal is to learn about the causes of global temperature changes from several documents. You will have to piece together important information across the documents to construct a good understanding. No one text will provide the answer. This task is interesting because you are the one making the connections across documents and coming up with an explanation. No author has already done the work for you. It is also important that you use information from the documents to support your explanation of the causes.

Your task is to use this set of documents to write an essay explaining how and why recent patterns in global temperature are different from what has been observed in the past. Be sure to use specific information from the documents to support your conclusions and ideas.

Students had access to the documents as they wrote the essays. Then, the essays and document sets were collected and students completed the Inference Verification Task, without access to the documents. These were collected and students completed a final booklet including the CLEAR Thinking Scale and the self-report descriptive student surveys. Teachers were asked to fill out the teacher survey while students worked on the inquiry task. The activity was done over two 50 minute periods.

Results

Descriptive Statistics

The descriptive statistics for all variables in the study are displayed in Table 1. All variables showed normal distributions and high variance covering the range of possible values. The mean CLEAR Thinking score was greater than 0, reflecting that most students had at least slight agreement with an evidence-based disposition. However, there was high variability and many students had negative scores and disagreed with an evidence-based disposition.

Table 1. *Means, Standard Deviations, and Ranges for all Measures*

Measure	Mean	SD	Observed Range	Possible range
CLEAR Thinking	1.19	1.93	-2.67 - 4.67	-5.00 - 5.00
Reading Skill	2.32	0.78	1.00 - 3.00	1.00 - 3.00
Prior Knowledge	3.46	1.18	1.00 - 5.00	1.00 - 5.00
Science Interest	2.98	1.42	1.00 - 5.00	1.00 - 5.00
Topic Interest	2.88	1.25	1.00 - 5.00	1.00 - 5.00
Essay Concepts	1.83	1.66	0.00 - 5.00	0.00 - 5.00
Inference Test	0.70	0.14	0.44 - 1.00	0.00 - 1.00

Correlations among Measures of Understanding and Individual Differences

As shown in Table 2, there was a significant positive correlation between the two outcome measures of understanding. Students with greater conceptual coverage in their essays (Essay Concepts) also tended to have higher scores on the Inference Verification Task (Inference Test), despite the fact that the texts were only available during the essay writing. Table 2 also shows that both these measures of understanding were predicted by CLEAR Thinking dispositions and by reading skill, and that prior knowledge predicted inference test performance. The relationship between prior knowledge and essay concepts was trending in the same direction but weaker and non-significant ($p = .15$).

Table 2. *Pearson Correlations among CLEAR Thinking, Reading Skill, Prior Knowledge, Science Interest, Topic Interest, Number of Essay Concepts, and Inference Test Scores*

Measure	1	2	3	4	5	6
1. CLEAR Thinking	-					
2. Reading Skill	.26*	-				
3. Prior Knowledge	.27*	.27*	-			
4. Science Interest	.11	.01	.08	-		
5. Topic Interest	.13	.05	.01	.73**	-	
6. Essay Concepts	.36**	.46**	.19	.13	.07	-
7. Inference Test	.39**	.42**	.30*	.13	.14	.45**

Note. $N = 59$. $^{*}p < .05$. $^{**}p < .01$.

Students' interest in both science and the specific topic failed to predict performance on the essays and the inference test, and were also unrelated to CLEAR thinking and the other predictors. However, the two interest measures were highly correlated with each other. In addition, interest levels differed for male and female students. Consistent with prior findings (for a meta-analysis, see Weinburgh, 1995) males had significantly higher interest in both science and the topic (*M*s = 3.32 and 3.42) than females (*M*s = 2.58 and 2.66), *t*s(57) = 2.33 and 2.02, *p*s < .05. Gender did not relate to any of the other predictors or to either outcome measure.

Unique Effects of Thinking Dispositions and Reading Skill on Understanding

The main question for the current study was whether individual differences in domain-general thinking dispositions might have unique effects from reading skill on the understanding that results from a multiple-document inquiry task in science. To examine this question, Reading Skill and CLEAR Thinking scores were entered simultaneously into a regression predicting the number of key explanatory concepts in the essays. As seen in the top half of Table 3, the regression resulted in a significant model accounting for 27% of the variance in Essay Concepts. The beta tests showed that both CLEAR Thinking and Reading Skill each accounted for significant unique variance. The inclusion of key explanatory concepts increased with Reading Skill. In addition, regardless of Reading Skill, students with a stronger general disposition towards evidence-based thinking were more likely to incorporate the key explanatory concepts into their essays.

Table 3. *Regression Analyses Predicting Inference Test Scores and Number of Essay Concepts from CLEAR Thinking and Reading Skill Scores*

Predictor	R^2	*F* Value	B	SEM	*B*	*t* Value
(DV) Essay Concepts						
Model	.27	10.29*				
CLEAR Thinking			.22	.10	.25*	2.13
Reading Skill			.84	.25	.39*	3.31
(DV) Inference Test						
Model	.26	9.81*				
CLEAR Thinking			.02	.01	.30*	2.51
Reading Skill			.06	.02	.34*	2.87

Note. $N = 59$. $^{*}p < .05$.

Another regression was conducted in which Reading Skill and CLEAR Thinking scores were entered simultaneously to predict Inference Test performance. The results reported in bottom half of Table 3 were very similar to the Essay Concepts results. The overall model was significant and accounted for 26% of the variance in test performance. The beta tests show that both CLEAR thinking and Reading Skill each accounted for significant unique variance. Inference Test performance was better for students with more reading skill. More importantly, regardless of reading skill, students with a stronger general disposition towards evidence-based thinking were more likely to correctly identify statements that could and could not be inferred by integrating the information from the multiple documents. Since prior knowledge of the topic was related to inference performance and CLEAR Thinking (see Table 2), this analysis was rerun adding prior knowledge as a control predictor. The results did not change, except for a slight increase in the total variance explained from 26% to 28%.

Conclusions

Across two measures of student understanding, the results of the present study demonstrate the influence of both reading skill and a domain-general thinking disposition on learning science from multiple-document inquiry tasks. These influences were independent from each other and from self-reported ratings of prior topic knowledge, interest in the topic and interest in science. Of these individual differences, only the interest ratings were not related at all to understanding. Although reading skill and prior topic knowledge were not assessed with standardized measures, the measures that were used did predict understanding as expected and were correlated with each other, suggesting they are capturing variance in their respective constructs.

The two measures of understanding (Essay Concepts and Inference Tests) similarly correlated with reading skill and with CLEAR Thinking, but correlated only modestly with each other. In addition, the inference test but not the essay concepts were significantly related to prior knowledge. The lack of relation between essays and prior knowledge makes sense given that the documents were available during writing, so students did not need to rely upon retrieval from long term memory in order to construct a more complete argument. Thus, the IVT and essay measures reflect somewhat different aspects of multiple-documents comprehension. Yet, a motivating disposition towards considering evidence (CLEAR Thinking) related to both of these different aspects of comprehension independently from reading skill, prior knowledge, and topic and domain interest. This is consistent with Stanovich's (2012) distinction between individual differences in what a person might be capable of (e.g., reading skill) versus what a person might be disposed to do. This study demonstrates that both are required for successful learning from multiple-document inquiry tasks in science.

Although the current study does not directly test the MD-Trace model, the finding that CLEAR thinking scores predicted middle school students' learning from multiple documents is consistent with the importance of the task model. Bråten et al (2011) hypothesized that epistemic beliefs contribute to the creation of a task model and it is likely that thinking dispositions function in much the same way. It is expected that readers who are disposed to using evidence and reasoning to form and update their beliefs will have a very different task model from those who do not. As a result, they will create different subgoals to guide reading. For example, students with an evidence-based disposition will be expected to seek coherence across explanatory elements and look for evidence to support claims. These subgoals will lead to the integration of more of the key causal concepts from the document set into their mental model of climate change. The present results also highlight that a task model may be more than what a learner thinks is expected and required for learning, but may also include their personal goals related to their desire to learn and update their views versus to protect and maintain their existing views. An interesting direction for future research would be to investigate the manner in which thinking dispositions influence the development of a task model.

Limitations and Future Directions

The working assumption behind the present findings is that students' dispositions towards evidence-based thinking impacts how they approach a multiple-documents inquiry task, Such dispositions make students more likely to engage in the kind of integrative, coherence-building, argumentation processes that have been shown to improve learning in these contexts (e.g., Wiley et al., 2011). The previously reviewed literature shows there is much variance in reading behaviors and strategies when readers are faced with multiple documents. Thinking dispositions may be a generalized individual difference that contributes to this variance. However, we note that the current study only measured the learning outcomes that were presumed to result from these different behaviors, but did not include any on-line measures of processing and reading strategies to verify actual differences in processing. Griffin and Ohlsson (2001) speculated that people who had previously formed a belief on a topic via evidence-based reasoning rather than relying on affective preferences may be better able to represent new belief-relevant concepts. Thus, an alternative to differences in how readers are actively engaging in the task is differences in how their past reasoning on the topic impacts their ability to represent the concepts. Future research is needed to provide evidence that readers vary in their processing during reading in ways that might mediate the observed learning outcomes.

Implications for Instruction

The fact that such a general thinking disposition was able to show relations to learning on a specific topic within science is pedagogically useful. The trend in research on the related construct of epistemology has been toward measuring more domain-specific rather than more general thinking dispositions (e.g., Hofer, 2006). In fact, Braten and Stromso (2010) have even argued for using topic specific epistemology, and have demonstrated that it can be used as a successful predictor of science learning for multiple documents on that topic. Predicting learning does seem to benefit from measuring epistemology in more specific ways (for a review, see Muis, Bendixen, & Haerle, 2006). However, from a pedagogical perspective, identifying general dispositions that could improve learning may be more pragmatically useful as targets for instruction. Domain-specific and topic-specific dispositions imply that separate pedagogies would be needed to target the development of thinking dispositions within each domain or on each specific topic. Any benefit of such interventions would be limited to that domain or topic. The present results suggest that there are more general thinking dispositions regarding the value of evidence that can also have a substantial impact on learning.

There has been little work on interventions targeting an evidence-based disposition. The fact that the disposition itself has some domain generality does not imply that topic-specific learning activities would be ineffective in fostering a dispositional change. Topic specificity may be necessary in order to expose students to examples of such thinking and to have them engage in tasks that require it. Long-term impact may prove difficult if this disposition reflects core values related to commitment to evidence versus the perceived value of sticking to one's beliefs. Such values would seem to be shaped by the social reinforcement students receive in many areas of life outside of school, from their home, religious upbringing, media, and popular culture. On the other hand, there appears to be little existing effort in schools to directly and explicitly foster an evidence-based disposition. Thus, even minor interventions could notably increase students' exposure to the importance and utility of adopting such a disposition. This is another direction for future research. If such a general disposition can be effectively encouraged, developed or leveraged through instruction, it has the potential to impact learning across topics and domains more generally.

• • •

Acknowledgement

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education in part by Grant R305B07460 Improving Metacomprehension and Self-Regulated Learning from Scientific Texts from the CASL program to Thomas D. Griffin and Jennifer Wiley and in part by Grant R305F100007 Reading for Understanding Across Grades 6 through 12: Evidence-based Argumentation for Disciplinary Learning from the Reading for Understanding Research Initiative to Anne Britt, Thomas Griffin and Jennifer Wiley. The authors thank Karyn Higgs, Kris Kopp, Brent Steffens, Andrew Taylor and the other members of Project READi for their assistance and contributions. The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences or the U.S. Department of Education.

Thomas D. GRIFFIN. Research Professor, Department of Psychology, University of Illinois at Chicago. His research interests include contexts and individual differences that impact learning in science and history, the impact of affect-based beliefs on the comprehension and acceptance of scientific theories, and the factors that determine metacomprehension monitoring accuracy for explanatory texts.

Jennifer WILEY. Professor, Department of Psychology, University of Illinois at Chicago. Her research interests include contexts that improve learning from informational texts in history and science, and the role of individual differences, especially in working memory capacity and spatial ability in comprehension processes.

M. Anne BRITT. Professor, Department of Psychology, Northern Illinois University. Her research interests include argument comprehension and production, and learning and reasoning from multiple documents.

Carlos R. SALAS. Graduate Student Researcher, Department of Psychology, University of Illinois at Chicago. His research interests include text comprehension, argumentation, and belief formation.

References

Baron, J. (1985). *Rationality and intelligence.* Cambridge, England: Cambridge University Press.

Bennington, J. B. (2009). *The Carbon Cycle and Climate Change*. Cengage Learning.

Bråten, I., Britt, M.A., Strømsø, H.I., & Rouet, J.F. (2011). The role of epistemic beliefs in the comprehension of multiple expository texts: Towards an integrated model. *Educational Psychologist, 46*, 48-70.

Bråten, I., & Strømsø, H.I. (2010). Effects of task instruction and personal epistemology on the understanding of multiple texts about climate change. *Discourse Processes, 47*, 1-31.

Bråten, I., Strømsø, H.I., & Britt, M.A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly*, *44*, 6–28.

Britt, M.A., Perfetti, C.A., Sandak, R., & Rouet, J-F. (1999). Content integration and source separation in learning from multiple texts. In S. R. Goldman, A. C. Graesser, & P. van den Broek (Eds.). *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso (pp. 209-233).* Mahwah, NJ: Lawrence Erlbaum Associates.

Britt, M.A., & Rouet, J-F. (2012). Learning with multiple documents: Component skills and their acquisition. In M.J. Lawson & J.R. Kirby (Eds.), *The Quality of Learning: Dispositions, Instruction, and Mental Structures*. Cambridge University Press.

Cerdán, R., Vidal-Abarca, E. (2008). The effects of tasks on integrating information from multiple documents. *Journal of Educational Psychology, 100*, 209-222.

Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 5, pp. 161-238). Mahwah, NJ: Erlbaum.

Conley, A. M., Pintrich, P. R., Vekiri, J., & Harrison, D. (2004). Changes in epistemological beliefs in elementary science students. *Contemporary Educational Psychology, 29*, 186–204.

Goldman, S .R., Braasch, J. L .G., Wiley, J., Graesser, A. C., & Brodowinska, K. (in press). Comprehending and learning from Internet sources: Processing patterns of better and poorer learners. *Reading Research Quarterly*.

Griffin, T. D. (2008). Faith: serving emotional epistemic goals rather than evidence coherence. In V. Sloutsky, B. Love, and K. McRae (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, pp. 2059-2064. Austin, TX: Cognitive Science Society, Inc.

Griffin, T. D., & Ohlsson, S. (2001). Beliefs versus knowledge: A necessary distinction for explaining, predicting and assessing conceptual change 364-369 In J. D. Moore and K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, pp. 364-369. Austin, TX: Cognitive Science Society, Inc.

Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition, 36,* 93-103.

Hofer, B. (2006). Domain specificity of personal epistemology: Resolved questions, persistent issues, new models. *International Journal of Educational Research. 45,* 85-95.

Kintsch, W. (1994). Learning from text. *American Psychologist, 49,* 294–303.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363-394.

Kokis, J., Macpherson, R., Toplak, M., West, R. F., & Stanovich, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology, 83*, 26-52.

Mason, L., Boldrin, A., & Ariasi, N. (*2010*). Epistemic metacognition in context: Evaluating and learning online Information. *Metacognition and Learning, 5*, 67–90.

Muis, K. R., Bendixen, L. D., & Haerle, F. C. (2006). Domain-generality and domainspecificity in personal epistemology research: Philosophical and empirical questions in the development of a theoretical model. *Educational Psychology Review, 18*, 3-54.

Perfetti, C.A., Rouet, J-F., & Britt, M.A. (1999). Towards a theory of documents representation. In H. van Oostendorp & S.R. Goldman (Eds.) *The Construction of Mental Representations during Reading (pp. 99-122)*. Mahwah, NJ: Lawrence Erlbaum Associates.

Rouet, J-F., & Britt, M.A. (2011). Relevance processes in multiple document comprehension. In M.T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text (pp.* 19 - 52). Greenwich, CT: Information Age Publishing.

Sá , W., Kelley, C., Ho, C., & Stanovich, K. E. (2005). Thinking about personal theories: individual differences in the coordination of theory and evidence. *Personality and Individual Differences, 38*, 1149-1161.

Sá, W. C., West, R. F., & Stanovich, K. E. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology, 91*, 497-510.

Sanchez, C., & Wiley, J. (2006) Effects of working memory capacity on learning from illustrated text. *Memory & Cognition*, *34*, 344-355

Sanchez, C., & Wiley, J. (2009). To scroll or not to scroll: Interactions of text presentation and working memory capacity. *Human Factors, 51*, 730-738.

Sanchez, C. A., Wiley, J., & Goldman, S. R. (2006). Teaching students to evaluate source reliability during internet research tasks. *Proceedings of the 7th International Conference of the Learning Sciences*, Bloomington, IN.

Stanovich, K. E. (2012). On the distinction between rationality and intelligence: Implications for understanding individual differences in reasoning. In K. Holyoak & R. Morrison (Eds.) (pp. 343-365), *The Oxford handbook of thinking and reasoning*. New York: Oxford University Press.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology, 89*, 342-357.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*, 161-188.

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language, 24,* 595–611.

Weinburgh, M. (1995). Gender differences in student attitudes toward science: A meta-analysis of the literature from 1970 to 1991. *Journal of Research in Science Teaching,* 32(4), 387-398.

West, R. F. Stanovich, K. E., & Toplak, M. E. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions, *Journal of Educational Psychology, 100*, 930-941.

Wiley, J., Ash, I.K., Sanchez, C.A., & Jaeger, A. (2011). Clarifying readers' goals for learning from expository science texts. In M. McCrudden, J. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 353-374). Greenwich, CT: Information Age Publishing.

Wiley, J., Goldman, S., Graesser, A., Sanchez. C., Ash, I., & Hemmerich, J. (2009). Source evaluation, comprehension, and learning in internet science inquiry tasks. *American Educational Research Journal, 46*, 1060-1106.

Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the comprehension in meta-comprehension. *Journal of General Psychology*, *132*, 408-428.

Wiley, J., & Voss, J. F. (1996). The effects of "playing" historian on learning in history. *Applied Cognitive Psychology*, *10*, 63-72.

Wiley, J., & Voss, J. F. (1999) Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91,* 301-311.

International Electronic Journal of Elementary Education, 2012, 5(1), 79-92.

How do Openers Contribute to Student Learning?*

Amber ZERTUCHE**
University of California, Berkeley, United States of America

Libby GERARD
University of California, Berkeley, United States of America

Marcia C. LINN
University of California, Berkeley, United States of America

Received: 20 August 2012 / Revised: 24 October 2012 / Accepted: 24 October 2012

Abstract

Openers, or brief activities that initiate a class, routinely take up classroom time each day yet little is known about how to design these activities so they contribute to student learning. This study uses technology-enhanced learning environments to explore new opportunities to transform Openers from potentially busy work to knowledge generating activities. This study compares the impact of teacher-designed Openers, Opener designs based on recent research emphasizing knowledge integration, and no Opener for an 8th grade technology-enhanced inquiry science investigation. Results suggest that students who participate in a researcher-designed Opener are more likely to revisit and refine their work, and to make significant learning gains, than students who do not participate in an Opener. Students make the greatest gains when they revisit key evidence in the technology-enhanced curriculum unit prior to revision. Engaging students in processes that promote knowledge integration during the Opener motivate students to revise their ideas. The results suggest design principles for Openers in technology-enhanced instruction.

Keywords: Technology, Science Education, Teaching Practices, Classroom Assessment, Formative Assessment

* This article is based on the first author's unpublished MA thesis entitled "How Teacher-led Openers Contribute to Student Learning in a Web-based Inquiry Science Environment." This material is based upon work supported, in part, by the National Science Foundation under Grant DRL-0733299 [Logging Opportunities in Online Programs for Science (LOOPS)]. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

** ✉ Amber Zertuche, Graduate School of Education, Tolman Hall, University of California, Berkeley, U.S.A. Phone: (415) 656-5840. E-mail: anzertuche@berkeley.edu

ISSN:1307-9298

Introduction

Openers, or brief activities that initiate a class, routinely take up classroom time each day yet little is known about how to design these activities so they contribute to student learning. Teachers often give mini-lectures to remind students of what they studied in the previous class. Or, teachers may assign a short writing assignment. Technology-enhanced learning environments provide teachers with new tools to transform Openers from busy work to valuable learning activities. Openers can engage students in knowledge integration activities such as making predictions, critiquing a claim, assessing peer essays, or reflecting on progress—all activities that have been shown to improve student learning (Chiu and Linn, 2011; Linn and Eylon, 2011; White and Frederiksen, 1998). This study compares the impact of teacher- and researcher-designed Openers to no Opener in an 8th grade technology-enhanced inquiry science investigation. Questions include: How effective are Openers? How does the design of the Opener contribute to student learning? And, what are effective post-Opener revision processes?

Teacher-led Openers can play an essential role in the success of technology-enhanced inquiry science instruction. Several studies show that the quality of teacher implementation of technology-enhanced instruction predicts the impact of the technology on learning (Tamin, Bernard, Borokhovski, Abrami, and Schmid, 2011). In a second-order meta-analysis conducted on technology-enhanced instruction over the past 40 years, the authors conclude that the teachers' ability to monitor student understanding and help students sort out their ideas is essential to successful technology implementation (Williams, Linn, M, Ammon, and Gearhart, 2004). This meta-analysis found that as teachers asked more knowledge oriented questions during the course of a technology-enhanced science investigation as opposed to procedural questions, learning gains improved significantly. Further, to realize the potential of technology-enhanced instruction, teachers need to help students critically analyze visualizations relative to the conceptual learning goal as students often overestimate their understanding of dynamic computer visualizations, particularly in Chemistry (Chiu and Linn, 2012).

Technology-enhanced learning environments provide teachers with unique tools to structure effective Openers. The computer stores and organizes a record of each student's work including their multiple revisions, and provides ways to make examples of the student's work public to the whole class. This means teachers can purposefully select examples of the student's work for class discussion. Effective selection of examples could prompt students (and their teachers) to monitor understanding, sort out ideas about difficult concepts, and revisit and refine their reasoning (Izsak, 2012; Linn and Eylon, 2011).

Using Student Work to Design Openers

Student work is collected in technology enhanced learning environments and can be used to design openers. Black and Wiliam (1998) reviewed hundreds of studies and came to the conclusion that the use of student work for formative assessment and refinement of instruction is one of the most powerful ways to increase students' learning gains, particularly among low achieving students. Similarly, Shute (2008) reviewed work on formative feedback, defined as information communicated to the learner that is intended to modify his or her thinking or behavior to improve learning, and found that effective feedback is non-evaluative, supportive, timely, and specific. Gerard, Spitulnik, and Linn (2011) showed that when teachers use student work to refine their teaching, students' learning benefits. These studies distinguish between activities providing verification, where students choose whether an answer is correct or not, and elaboration, where teachers provide cues to guide learners toward a correct answer. Gerard et al found that use of student work to design instruction is most effective when the resulting instruction provides advice on what to do to improve performance, rather than on comparisons to other students or on accuracy of responses.

One example of a successful use of student work is for self- and peer-assessment. Research suggests that students can benefit from self and peer assessment activities when they recognize the desired learning goal, have adequate evidence to determine where the work stands relative to the learning goal, and have an understanding of a way to close the gap between the two (Black and Wiliam, 1998). Under these conditions, students can distinguish between their own ideas and the goal of instruction and strengthen their understanding. Taken together, these results suggest that effective Openers should elicit student work and provide hints or guidance.

What are the challenges of formative assessment?

Black and Wiliam's review of formative assessment studies point out that, although formative assessment has proven to be useful to student learning, there are concerns about how to enact successful formative assessment (1998). One issue is that in order to implement formative assessment in a useful way, teachers must have easy access to evidence of students' ideas and efficient routines to elicit students' ideas during class so they can help students to distinguish among these ideas. This calls for assessment or discussion questions that prompt students to make their reasoning explicit and provide multiple entry points for students with various levels of understanding. Students must also be actively involved in the feedback they are receiving in order for it to have an effect on their learning. This requires significant changes in a traditional secondary science classroom. Traditional science classroom routines often center on teacher-directed lectures and demonstrations, leaving little space for students to reflect on their understanding and sort out the variety of ideas they hold about the topic gathered from everyday experiences, peers, and school curricula. Ruiz-Primo and Furtak (2007), for instance found that teachers' formative assessment routines focused on procedural elements of inquiry learning, learning such as checking the students' knowledge of the correct procedure and asking them to apply a procedure to a new situation, rather than knowledge generation.

How to Design an Effective Opener in Chemistry?

This study investigates the design of knowledge integration Openers. Typical classroom Openers usually reflect the absorption model of instruction and use the question, response, evaluation (QRE) approach. In this model, the teacher asks a question with one answer in mind and prompts until a student gives the answer or fills in the response if none is elicited. This theory of learning assumes that the task of the learner is to acquire the body of connections that an expert analysis of the subject matter reveals (Greeno, Collins, and Resnick, 1996).

In contrast, the knowledge integration perspective on learning resonates with the Black and Wiliam's findings (2011) and guides the design of the Openers, curriculum and assessments in this study. The knowledge integration perspective draws on findings from learning sciences research. Specifically, learners hold multiple conflicting ideas about scientific phenomenon as has been documented in numerous studies of student intuitions about science topics (diSessa, 2000). In addition, learners, often in collaboration with others, can deliberately sort out, link, and critique their ideas when making sense of new scientific phenomena and benefit from encouragement to engage in this process (Linn, Lee, Tinker, Husic, and Chiu, 2006; Linn and Hsi, 2000; Novak and Gowin, 1984; Slotta, Chi, and Joram, 1995). This means that providing opportunities for students to compare alternative ideas to their own, develop criteria for sorting-out and distinguishing among ideas, and reflect on their ideas can help them form coherent hypotheses or explanations.

The knowledge integration Openers in this study were designed to help students make connections among their ideas about chemical reactions. In chemistry, one of the most difficult things for students to learn is how chemicals react. Students often have difficulties translating between symbolic representations, molecular representations, and observable phenomena (Ardac and Akaygun, 2004). Particularly, students struggle to make sense of chemical phenomena at the molecular level (Johnstone, 1993; Krajcik, 1991). For example, many students think of chemical reactions as an instantaneous process without bond breaking and formation, while others think all the molecules break into atoms. In addition, prior studies demonstrate that students often isolate molecular visualizations rather than linking them to existing knowledge or everyday experiences and have difficulty interpreting stand-alone dynamic visualizations (Tversky, Morrison, and Betrancourt, 2002; Zhang and Linn, 2011). This study addresses this gap in learning chemical reactions with Openers that ask students to reflect upon and critique peers' visual molecular representations of hydrogen and oxygen combustion.

Methods

Research Design

This study investigates how a researcher-designed Opener, teacher-designed Opener, and a control condition (no Opener) contribute to students' revision of work and understanding of chemical reactions. Three central questions guide this research:

1. Do Openers contribute to student understanding of chemical reactions?
2. How does the Opener design influence students' learning outcomes?
3. What are effective post-Opener revision processes?

Curriculum and Assessments

The Web-based Inquiry Science Environment (WISE) is an open-source on-line learning environment that includes multiple standards-aligned science inquiry curriculum units. To engage students in knowledge integration processes, WISE projects guide students in collaborative activities with visualizations of scientific phenomena that are difficult to observe, such as molecular views of chemical reactions (Figure 1). Students investigate hypotheses, design solutions to problems, critique scientific claims, and build scientific models, scaffolded by guidance based on knowledge integration principles.

Students in this study worked on the WISE Hydrogen Fuel Cell Car unit. This is a one week unit designed to teach students about chemical reactions, alternative fuels, and energy (http://wise.berkeley.edu/webapp/vle/preview.html?projectId=911). The project begins by asking students if they would rather buy a hydrogen or gasoline powered car. It also elicits their ideas about energy and adds ideas about conservation of energy. Then, gasoline combustion in cars is explored including the relationship between carbon dioxide, a product of gasoline combustion, and temperature changes over the last 200 years. Students then create their own energy story about cars. This story includes where the energy came from to power the car and any chemical reactions that are involved in their story. Hydrogen combustion is then explored using a dynamic visualization of hydrogen combustion (Figure 1).

Then students are taught about the difference between exothermic and endothermic reactions and finally, students investigate a visualization of a hydrogen fuel cell to learn how this technology works. Students are then asked which kind of technology they would prefer when buying a car.

Assessments are embedded throughout the WISE projects to help students and teachers monitor student understanding and progress as students interact with visualizations (Figure 2). The embedded assessments ask students to make predictions about the visualizations, sort out evidence, and link ideas together to explain their thinking. Students can also get hints to help them complete the tasks.

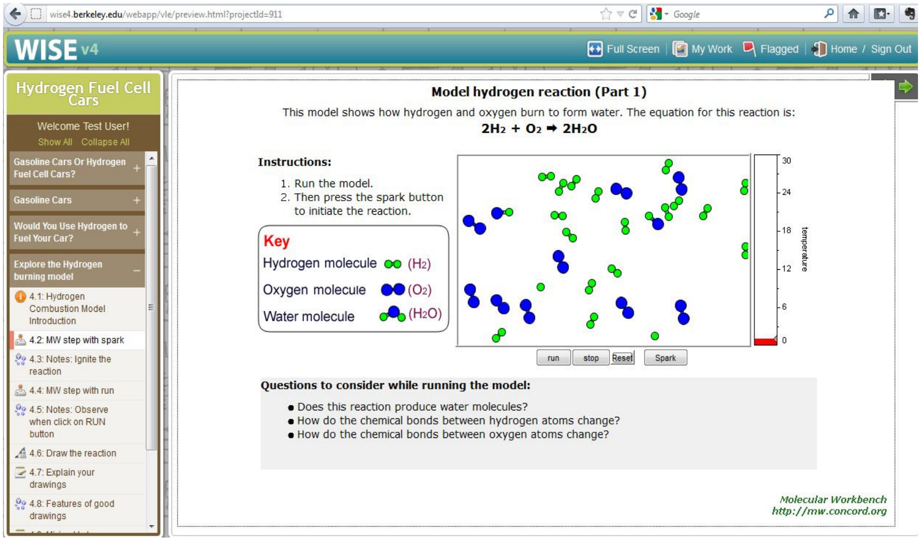


Figure 1. One of the visualizations in WISE's Hydrogen Fuel Cell Cars project models the hydrogen combustion reaction. Students can run, stop, reset, and 'spark' to start the reaction and explore the nature of chemical reactions.

In the Hydrogen Fuel Cell Cars unit, an embedded assessment immediately follows the visualization of hydrogen combustion asking students to draw four frames of hydrogen combustion (Figure 2). This is meant to help students make sense of the dynamic visualization, recognizing features such as conservation of mass, bonds breaking, and the progression of the reaction. The Openers in this study focused on student work from this embedded assessment since in previous years this particular task was particularly challenging to students and yet still, understanding how to draw basic hydrogen combustion is critical to student understanding of chemical reactions.

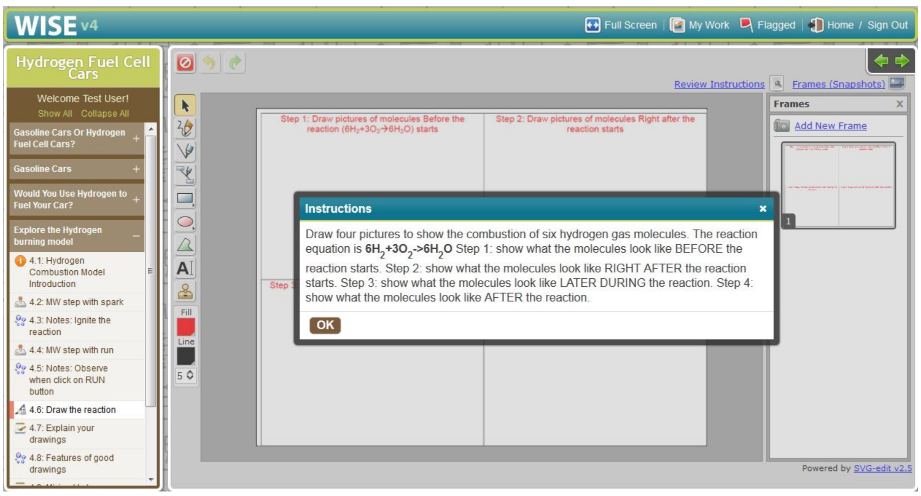


Figure 2. The WISE project screen has an inquiry map on the left, a navigation bar on top, and an embedded assessment in the center of the screen.

Teacher Assessment Tools

WISE provides teachers convenient access to a record of student work as students progress through a WISE unit (Figure 3). The WISE grading tools allow teachers to view embedded assessment data by step, and the flag tool allows teachers to select key student work examples for display. Students can view flagged work at any time by clicking on a tab at the top of their WISE screen. This allows students to actively see their peers' work and be a part of the feedback process to improve their understanding. Students can revise their work based on teachers' and peers' comments. All revisions are logged in the grading tool so the teacher can measure the impact of their Opener, or comments, by viewing the change in students' ideas from their original to revised work.

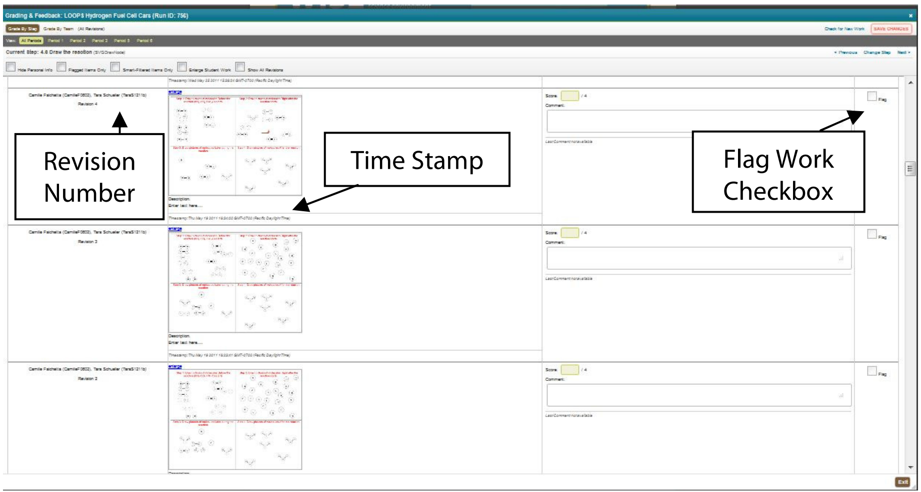


Figure 3. WISE assessment tool with revision number on the left, including time stamp, and checkbox to flag work on the right.

Study Participants

Two 8th grade teachers and their 236 students from 1 public school participated in this study. The school has medium diversity (22% receiving free lunch, 5% ELL, and 27% non-white). Both teachers had over five years experience teaching with WISE and frequently used Openers in their regular and WISE instruction.

Students were randomly assigned by class period to one of the three conditions. There were 78 students in the teacher-designed Opener, 128 students in the researcher-designed Opener, and 30 students in the control group condition. The uneven sample sizes were due to the uneven number of class periods that each teacher taught Physical Science that year.

Opener Design

The researcher-designed Opener engaged students in Knowledge Integration processes, as shown in Table 1. Activities included a small group discussion, voting, a whole group discussion, and then a closing summary by the teacher. The examples were selected to illustrate the range of conceptual errors in student representations of the chemical reaction. These related to (a) conservation of mass, (b) breaking of bonds, and (c) progression of the reaction.

The teacher-designed Opener alternatively focused on picking out what is good and bad about the student work and a lecture to try to re-teach chemical reactions from a different angle. Examples of student work were selected to illustrate responses that could be described as, "the good, the bad and the ugly".

Table 1. *Description of Openers used in Teacher 1 and Teacher 2's classroom using the Knowledge Integration Framework.*

	Teacher-Designed	Researcher-Designed
Teacher 1	**Elicit Ideas**: Question on the board about how a balanced equation obeys the law of conservation of mass. Students fill out a chart on how much mass (amu) exists before and after reaction. **Add Ideas**: Students use physical model of molecules to break bonds and put back together. Teacher shows multiple types of reactions. **Distinguish Ideas**: Teacher shows four examples of student work and asks students if the example is good or bad. **Integrate Ideas**: ____ (~12 min)	Students open WISE project to view four examples of student work in WISE project by clicking on the "Flagged Work" button. Each example has one unique link missing (Figure 4) **Elicit Ideas**: Teacher asks students to write down which drawing best represents the visualization of hydrogen combustion and use evidence to explain why. Students vote and teacher tallies votes. **Add Ideas**: Students discuss their choices in groups of 4 and revisit evidence in the visualization. **Distinguish Ideas**: Students reconsider their initial choice in light of their discussion with peers and revisit the evidence. Make a new vote. **Integrate Ideas**: Teacher tallies new votes and asks students to justify their choice. Teacher synthesizes criteria used to evaluate drawings, and instructs students to revise their own drawing. (~20 minutes)
Teacher 2	**Elicit Ideas**: ____ **Add Ideas**: Used embedded assessment problem and a tree to house analogy to take students through the chemical reaction steps. Tree must break into parts, then recombine parts and build a house. Uses physical models of hydrogen and oxygen molecules to show students progression necessary to make water. **Distinguish Ideas**: ____ **Integrate Ideas**: ____ (~8 min)	

Link missing: conservation of mass	Link missing: bonds breaking

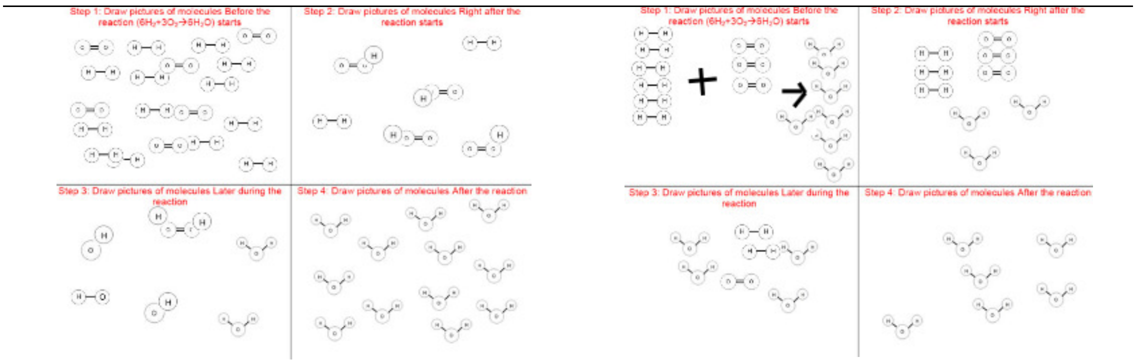


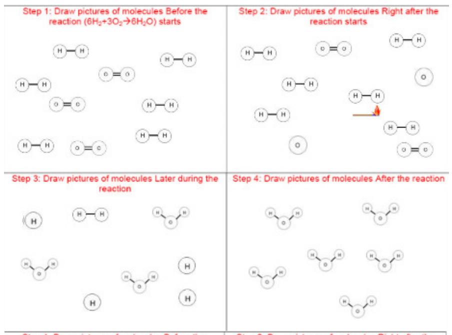
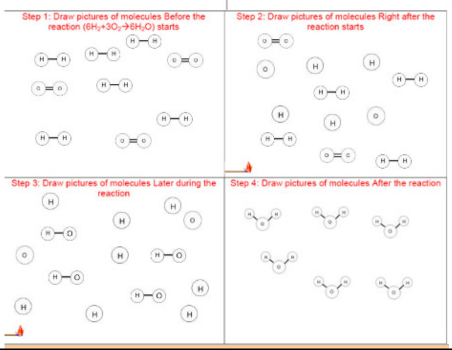
Figure 4. Examples of student work shown during the researcher-designed Opener.

Data Collection and Analysis

Data sources include: the original student work on the embedded assessment, the revised student work on the embedded assessment after the Opener, WISE log files illustrating student navigation through WISE immediately after the Opener, pre and post tests administered immediately before and a 1-14 days after the WISE project, classroom video of teacher implementation of the researcher-designed and teacher-designed Openers, and teacher interviews.

Knowledge integration rubrics were used to score the embedded and pre/post assessments. The embedded assessment rubric is shown in Table 2.

Table 2. *KI rubric for embedded assessment where students make step-by-step drawings of hydrogen combusting with oxygen*

KI Level	Score	*Characteristics*	*Example*
Invalid	*1*	Blank or I don't know	
No links	*2*	No normative ideas conveyed but work has been done	
Simple, 1 link	*3*	Represents conservation of mass OR bonds breaking OR progression*	
Advanced, 2 links	*4*	Conservation of mass AND progression* OR Conservation of mass AND bonds breaking OR Progression* AND bonds breaking	
Complex, 3 links	*5*	Conservation of mass AND bonds breaking AND progression* of reaction	

* Progression must be apparent on all 3 transitions

Pre and post questions were also scored using a KI rubric. The main ideas that students should understand in the pre- and post- test are that there is a reaction process that occurs and that H_2 and Cl_2 should start breaking bonds before forming new bonds.

Results

We examine the effects of Openers on student learning outcomes, and then explicate the contributing factors including Opener design and students' learning practices. We focus on embedded assessments and pretest-posttest performance. We consider the actual implementation of the conditions and student performance as reflected in the log files.

Embedded assessments

To investigate the impact of the Opener versus no Opener on student understanding we compared students' original (before Opener) and revised (after Opener) responses to the embedded assessments. Responses were scored using the knowledge integration rubric. Both teacher and researcher-designed Openers (n=105 pairs) were compared to the no Opener condition (n=30 pairs). Pairs who did not complete the embedded assessment before the Opener were excluded from the analysis (n=26 pairs). Although the aim was to facilitate the Opener after 75% of students completed the embedded assessment, this number was based on an automated progress screen that only monitored whether or not students submitted work at least once for this assessment. Once the researchers looked at the student's work, it was obvious that, although work was submitted, many students did not finish their drawings and therefore these were not included in the data analysis.

Responses were scored using the KI rubric. Time stamps from the WISE log files were used to identify students' final work immediately before the Opener and their revised work on the day of the Opener.

The analysis suggests that students who had an Opener, either teacher- or researcher-designed, made substantially greater learning gains than students who did not have an Opener. As shown in Table 4, students who had an Opener (M = .29, SD = .78) doubled the mean gain score of those students who did not have an Opener (M = .13, SD = .51) on the embedded assessment. There was no significant difference between conditions in students' pre-Opener scores.

Students who had an Opener (M = .68, SD = .48) were significantly more likely to revisit and revise their work than students who did not have an Opener (M = .33, SD = .48), $t(131) = -3.5$, $p < 0.001$. In the Opener condition, revision was twice as likely as in the no Opener condition.

Table 3. *Pre to Post Gain Scores with and without Opener.*

	N Pairs Who Revised Their Work	Mean Gain (SD)	Rate of Revision (N Pairs who Revised/Total Pairs)
Opener	105	.29(.78)	66%
No Opener	30	.13(.51)	33%

The Openers were particularly effective for students who demonstrated at least a basic understanding on the embedded assessment prior to the Opener (Table 7). Having an Opener had a significant effect on students who began with at least a partial understanding, or level 3 on the knowledge integration scoring rubric (M = .71, SD = .69), t(26) = -2.07, p = .05. Students who began with partial understanding developed their ideas into a basic understanding (level 4) after the Opener. In contrast, students who began with partial understanding and had no Opener continued to demonstrate only partial understanding after revising their work (M = .18, SD = .60). The large effect of the Opener on level 3 students is partially due to the examples of student work selected for critique in the researcher and teacher-designed Openers. The selected examples that were shown during the Opener illustrated characteristics of level 3 understanding, making them most accessible to this population of students. Since student work was chosen this way, this Opener was unwittingly designed to increase the learning gains of students with an already basic understanding of chemical reactions.

Students with non-normative ideas, or level 2 understanding, made modest gains with or without an Opener, as shown in Figure 5. Level 2 students may need examples aligned with their own ideas to improve to partial or high level understanding. For instance, showing student work with no conservation of mass or bond breaking would lead to a discussion where students point out that both are missing from the drawing. Alternatively, showing a common level 2 student work may show this population of students that their attempt is acknowledged and that they have direct feedback for improving their drawing.

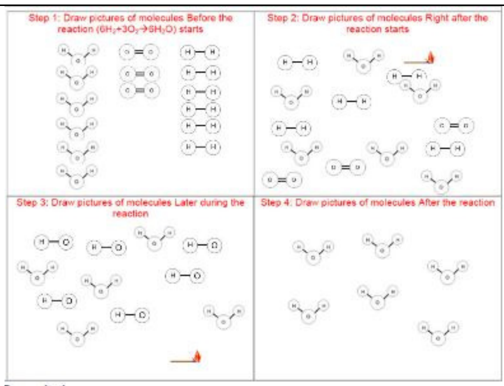
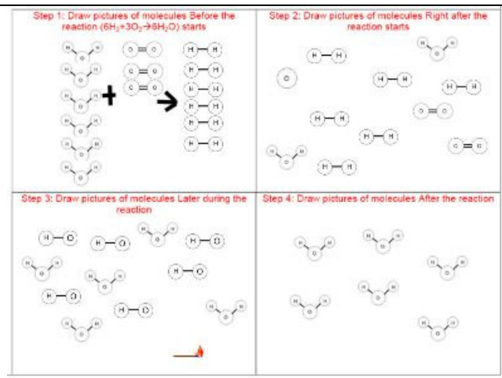
Pre Opener: No Links Present	Post Opener: No Links Present
	

Figure 5. Examples of students' work with a level 2 understanding before and after an Opener.

Table 6. *Pre to Post Opener Gain Scores Distributed by Pre-Opener Scores*

Pre-Opener Score	N Pairs Who Revised Their Work	Mean Gain With Opener(SD)	N Pairs who Revised	Mean Gain Without Opener (SD)
2	10	.4(.70)	5	.4(.89)
3	17	.71(.69)	11	.18(.60)*
4	42	.21(.42)	10	0(0)

*p=.05

Pre-test to Post-test effects. The pre- and post- tests were administered before students began the week long project and after they finished the project. In spite of the gains found for the specific item addressed by the Opener, there was no difference on the pre/post tests between students who had an Opener (M = 1.74, SD = 2.50) and those who did not (M = 1.55, SD = 2.16), $t(255) = -.511$, p = .61. This is not surprising since the Opener treatment was only 10-20 minutes out of the 5-7 hour long project. Perhaps several Openers over the project length would have been able to affect the pre test to post test results. Also, one of the teachers in the study did not give the post test until 2 weeks after the project was completed. Although the schedule was controlled, the teacher ultimately has the authority on when to give the post test. This delay may have masked any immediate effects of the Opener on the post test. Because of these time related anomalies, student performance on the embedded assessment is a better representation of the immediate effects of the Openers.

How does the Opener design influence students' learning outcomes?

The Opener designs, described in Table 2, differed primarily in their support for distinguishing and integrating ideas. In both Openers, students were presented with evidence regarding a chemical reaction and prompted to think about conceptual features of a chemical reaction. The researcher-designed Opener had additional components to support integration of ideas. After considering the new evidence, students were guided to reconsider their initial views in light of the evidence presented, and refine their criteria for conceptual features of a chemical reaction. These additional components made the researcher-designed Opener longer (approximately 20 min) than the teacher-designed Opener (approximately 10 min). Since the length of both Openers was difficult to predict ahead of time, the students in the teacher-designed Opener did not have an additional task to account for the 10 minute different time on task.

Students who participated in the researcher-designed Opener were significantly more likely to revise their work on the embedded assessment ($M = .75$, $SD = .44$) compared to students who participated in the teacher-designed Opener ($M = .56$, $SD = .50$), $t(101) = -1.98$, $p < 0.05$ (Table 8). There was no significant difference in learning gains between the two conditions (Table 9).

The researcher-designed Opener was not completely implemented as planned. The voting process in the researcher-designed Opener was a new activity for students and did not work as anticipated. In some classes, students directly copied the examples of student work that received the most votes even though none of the student work examples were in fact "correct". Although teachers reminded students that none of the student work examples were correct, this could have been emphasized and the teacher could have checked that students truly did understand that examples were not to be directly copied (have a student repeat it back, whole class response etc).

Table 7. *Student KI Score (Maximum Score 5) of Embedded Assessment Before and After the Opener Treatment*

	n Revised	Total Teams	Avg Rate of Revision	Pre Opener	Post Opener	KI Score Gain
Teacher Designed Opener	22	39	.56(50)	3.88	4.26	0.38
Researcher Designed Opener	48	64	.75(.44)	3.59	3.95	0.36

Table 8. *Student Pair KI Score (Out Of 5) for Pre and Post Test*

	n	Pre Test	Post Test	KI Score Gain
Teacher Designed Opener	78	2.69	3.54	0.85
Researcher Designed Opener	121	2.58	3.30	0.72

Log files. To analyze post-Opener learning practices we looked at log files. Two students from each period, one with the highest gain and one with the lowest gain, were selected for log file analysis in order to get an equal distribution of student learning practices from each condition. Students' learning practices may explain why some students made greater gains on the embedded assessment post Opener than others. The WISE log files show that the biggest contribution to learning gains was the time students spent revising their initial work, and revisiting relevant evidence in the WISE unit. The students who made the greatest improvement revisited an evidence page immediately after the Opener or spent more time revising their original work than other students (Table 10). An evidence page could be the dynamic visualization of hydrogen combustion or notes that students wrote about what a good drawing of hydrogen combustion should include. This suggests the value of providing students with access to evidence to sort out their ideas during an Opener.

Table 10. *Comparison of Students Who Had a Gain in Their KI Score By At Least 1 and Those Who Had No Gain. After The Opener, Students With a Gain Either Spent More Time on the Embedded Assessment or Revisited an Evidence Page After the Opener. Percentage Results Are Aggregated for Each Condition.*

Learning Practice	Gain (n=6)	No Gain (n=6)
Revisited evidence page	67%	33%
Revisited embedded assessment for more than 1 minute	100%	50%
Both	67%	33%

Discussion

This study illustrates how Openers can improve student learning compared to not using Openers. Both the teacher-designed and researcher-designed Openers used in this study encouraged students to review their ideas. The researcher-designed Openers encouraged students to distinguish ideas and reflect, and resulted in a greater propensity to revise answers than did the teacher-designed Openers.

These results for Openers resonate with research showing the benefit of giving students feedback based on their responses to assignments (Black and Wiliam, 1998; Shute, 2008). Openers supported students to reflect upon their initial ideas, reconsider evidence, and refine their views. Openers were particularly useful for students who began with a partial understanding of a concept. When students benefitted from Openers, they took advantage of the evidence presented in the Opener, revisited a dynamic visualization or another evidence source, and reflected on the new information.

The findings from this study suggest the following design principles for Openers:

1) Openers can reinforce normative conceptual ideas by drawing attention to the distinction between student ideas and normative views. Getting evidence from peers, visualization, teacher, or other classroom resource can help students understand complex ideas. This evidence helps students close the gap between what they know now and the normative view of the phenomena studied.

2) Openers help students when they occur soon after a topic is introduced and direct attention to specific ways to improve their ideas.

3) Openers succeed when they are non-evaluative and support students to explore evidence or views of their peers. For example, teachers can support students by giving them the opportunity to distinguish among the ideas held by the group of students in the class. Giving students a chance to appreciate conflicting views held by classmates and use evidence to sort them out helps students integrate their ideas.

4) Openers should be short. The Openers in this study were much longer and had more teacher involvement for most Openers. It may work better to have students review examples of student work as homework, then discuss with their partner at the very start of class. This would free the teacher up to take attendance and give more time for the peer and class discussion.

• • •

Amber ZERTUCHE. Received Masters degree and Single Subject Credential in Science and Math Education from University of California, Berkeley in 2012. She is a first year Physics teacher at Burton High School in San Francisco Unified School District since 2012. Before becoming an educator, she helped engineer the James Webb Space Telescope's primary mirrors.

Libby GERARD. Postdoctoral fellow with the Technology-Enhanced Learning in Science center in the Graduate School of Education at the University of California, Berkeley. She completed her PhD from Mills College in 2008. She studies the relationship between professional development, technology-enhanced instruction and student knowledge integration in inquiry science. Libby is particularly interested in how to use technology-enhanced assessment data to support teacher and principal decision-making to improve inquiry instruction and student learning. Libby leads the TELS professional development activities including summer institutes, sustained inquiry based communities, and teacher mentoring.

Marcia LINN. Professor, Graduate School of Education, University of California, Berkeley. Mbr: National Academy of Education. Fellow: American Association for the Advancement of Science, the American Psychological Association. Elected: Chair of the AAAS Education Section, President of the International Society of the Learning Sciences. Awards: National Association for Research in Science Teaching Award for Lifelong Distinguished Contributions to Science Education, Council of Scientific Society Presidents first award for Excellence in Educational Research.

References

Ardac, D. & Akaygun, S. (2004). Effectiveness of multimedia-based instruction that emphasizes representations on students' understanding of chemical change. Journal of Research in ScienceTeaching, 41(4), 317-337.

Berland, L.K. & Reiser, B.J. (2011). Classroom communities' adaptations of the practice of scientific argumentation. Science Education, 95(2), 191–216).

Black, P. & Wiliam, D. (1998). Inside the black box: raising standards through classroom assessment. Phi Delta Kappan, 80(2), 139-148.

Chiu, J. & Linn, M. (2012). The Role of Self-monitoring in Learning Chemistry with Dynamic Visualizations. In *Metacognition in Science Education* (pp. 133–163). Dordrecht: Springer.

diSessa, A. (2000). Changing minds: Computers, learning and literacy. Cambridge, MA: MIT Press.

Eylon, B. & Linn, M. (1988). Learning and instruction: An examination of four research perspectives in science education. *Review of Educational Research*, 58(3), 251–301.

Gerard, L.F., Spitulnik, M., & Linn, M.C. (2011). Teacher Use of Evidence to Customize Inquiry Science Instruction. *Journal of Research in Science Teaching*, 47(9), 1037-1063.

Greeno, J.G., Collins, A.M., & Resnick, L.B. (1996). Cognition and learning. In D.C. Berliner & R.C. Calfee (Eds.), Handbook of educational psychology (pp.15-46). New York: MacMillan.

Johnstone, A.H. (1993). The development of chemistry teaching: A changing response to changing demand. Journal of Chemical Education, 70(9), 701-704.

Krajcik, J. (1991). Developing students' understandings of chemical concepts. In S. Glynn, R. Yeany, & B. Britton (Eds.), *The psychology of learning science* (pp. 117-147). Hillsdale, NJ: Erlbaum.

Linn, M.C., Clark, D., & Slotta, J.D. (2003). WISE Design for Knowledge Integration. Sci Ed, 87, 517-538

Linn, M.C. & Eylon, B.S. (2011). *Science Learning and Instruction: Taking Advantage of Technology to Promote Knowledge Integration*. New York: Routledge.

Linn, M. C. & Hsi, S. (2000). *Computers, teachers, peers*. Hillsdale, NJ: Erlbaum.

Linn, M.C., Lee, H.S., Tinker, R., Husic, F., & Chiu, J.L. (2006). Teaching and Assessing Knowledge Integration in Science. *Science*, 313, 1049-1050.

Novak, J., & Gowin, D. (1984). *Learning how to learn*. New York: Cambridge Books.

Slotta, J. D., Chi, M. T. H., & Joram, E. (1995). Assessing the ontological nature of conceptual physics: A contrast of experts and novices. *Cognition and Instruction*, 13(3), 373–400.

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies*, 57, 247-262.

Tamin, R., Bernard, R., Borokhovski, E., Abrami, P., & Schmid, R. (2011). What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study. *Review of Educational Research, 81*(1), 4-28.

White, B.Y. & Frederiksen, J.R. (1998). Inquiry, Modeling, and Metacognition: Making Science Accessible to All Students. *Cognition and Instruction*, 16(1), 3-118.

Williams, M., Linn, M. C., Ammon, P., & Gearhart, M. (2004). Learning to Teach Inquiry Science in a Technology-Based Environment: A Case Study. *Journal of Science Education and Technology*, 13(2), 189-206.

Zhang, Z. & Linn, M. C. (2011). Can Generating Representations Enhance Learning with Dynamic Visualizations? *Journal of Research in Science Teaching, 48*(10), 1177-1198.

International Electronic Journal of Elementary Education, 2012, 5(1), 93-108.

Cognitive factors that influence children's learning from a multimedia science lesson

Florencia K. ANGGORO*
College of the Holy Cross, Worcester, MA, USA

Nancy L. STEIN
National Opinion Research Center and the University of Chicago, Chicago, IL, USA

Benjamin D. JEE
College of the Holy Cross, Worcester, MA, USA

Received: 7 September 2012 / Revised: 3 November 2012/ Accepted: 12 November 2012

Abstract

The present study examined the cognitive factors that influence children's physical science learning from a multimedia instruction. Using a causally coherent text and visual models, we taught 4^{th}- and 7^{th}- grade children about the observable and molecular properties of the three states of water. We manipulated whether the text was read by a tutor (which supports simultaneous encoding of the verbal and visual information, i.e., *temporal contiguity*) or whether children read the text on their own (which supports self pacing and interpretation of the information). Children in each condition received either static or dynamic graphics. Results showed that, regardless of the type of graphics, children demonstrated the greatest learning gains when the text was read to them by a tutor. This effect was more pronounced for the younger children. Thus, conditions that promote integration of verbal and visual information may provide the greatest support to children's learning from a causally coherent multimedia science lesson.

Keywords: Science Learning, Multimedia Instruction, Causal Coherence, Elementary Education.

Introduction

By the 4^{th} grade, national and state standards in the United States require children to learn about the water cycle and states of matter (National Research Council (U.S.), 1996; Project 2061 (American Association for the Advancement of Science), 1993). One central educational concern is that children are often presented with materials that are

* Florencia K. Anggoro, Department of Psychology, College of the Holy Cross. 1 College Street. Worcester, MA 01602. USA. Phone: 508-793-2748. E-mail: fanggoro@holycross.edu

ISSN:1307-9298

www.iejee.com

incomplete, inaccurate, and otherwise ineffective (Bar & Galili, 1994). Existing materials designed to teach physical science rarely contain all relevant concepts necessary for an accurate understanding of states of matter, and even when necessary concepts are included, the materials often lack clarity and coherence (Duschl, et al., 2007). This can be devastating to novice students, who are especially dependent on coherence and explicitness during learning (McNamara, Kintsch, Songer, & Kintsch, 1996; Shwartz, Weizman, Fortus, Krajcik, & Reiser, 2008; Stein, Hernandez, & Anggoro, 2010; Stein & Trabasso, 1982; Trabasso & Bouchard, 2000; Trabasso, Secco, & van den Broek, 1984). Another concern, besides the materials themselves, is that learning is greatly affected by a student's capacity to process the information that they receive. Even if the learning materials are coherent and contain all of the relevant concepts, learning may be inhibited if the demands on a student's cognitive resources are too great. The present study focused on this latter issue and examined the conditions that support and hinder children's early learning of physical science. Using materials developed from a theory of complex learning (Stein et al., 2010), we examined how different processing demands affected 4^{th}- and 7^{th}-grade children's learning about the observable and molecular properties of the three states of water.

A theory of complex learning

In previous research, Stein and colleagues used a theory of complex learning to create learning modules designed to teach elementary-school children about the observable and molecular properties of the three states of water (Stein et al., 2010; Stein, Hernandez, Anggoro, & Hedberg, under review). According to this theory, knowledge acquisition in the sciences requires three types of learning: concept learning, causal explanation-based learning, and argument learning.

In *concept learning* (Klausmeier, 1992; Mandler, 2008; Winston, 1986), explicitness is necessary, especially when learners are novices with little or no prior knowledge of the concepts to be learned. The learning materials must describe all dimensions of the concept as well as the dimensions that are not part of the concept (especially when similar concepts exist) (Klausmeier, 1992; Winston, 1986). An explicit compare/contrast procedure must be used to evaluate similar concepts on critical dimensions, especially when error analyses show a high rate of confusion among certain features in two concepts (Klausmeier, 1992). The reason for such explicitness is to achieve an accurate representation of the chosen concepts, and to avoid over- or under-inclusion of members of a category due to faulty knowledge.

Science learning also involves *causal explanation-based learning* (Stein & Levine, 1989; Stein & Trabasso, 1982; Trabasso & Stein, 1997). For example, understanding states of water requires knowledge of the mechanism that causes water to retain its shape and volume in a solid state. Learning about these causal mechanisms provides learners with transferable knowledge that allows them to explain related phenomena, such as how liquid water has a flexible shape but invariant volume. Failing to provide causal explanations, however, results in superficial understanding (Stein & Levine, 1989; Stein & Trabasso, 1982; Trabasso & Stein, 1997), especially in novices who have little or no knowledge of the domain.

Finally, knowledge acquisition in science involves *argument learning*, which emphasizes the correction of learners' misconceptions about newly learned concepts. Many misconceptions occur because of the presence of an incorrect causal explanatory structure that underlies the misconception (e.g., Slotta & Chi, 2006, Vosniadou & Brewer, 1992). Inaccurate beliefs in the domain must be changed and updated. This can be accomplished by providing evidence in favor of the correct concept and showing learners why their incorrect beliefs need to be changed. Thus, in correcting student errors, an entirely new explanatory structure often needs to be acquired (Stein & Miller, 1993; Thagard, 2000).

The benefits and demands of multimedia instruction

These elements of complex learning—concept learning, causal explanation-based learning, and argument learning—cannot be supported through text-based instruction alone. Explicit visual models are needed to teach learners about complex spatial and causal properties and processes, such as the organization, speed, and movement of molecules, that are difficult to convey in words (e.g., Larkin & Simon, 1987). When used properly, visual models have been found to accelerate learning for both children and adults (Gobert & Buckley, 2000; Goldberg & Bendall, 1995), and can benefit learners at different levels of expertise (Goldberg & Bendall, 1995; Jose & Williamson, 2005; Mayer, Hegarty, Mayer, & Campbell, 2005; Tversky, et al., 2008). Recent work on children's learning about states of water has found that the absence of visual models that exemplify core concepts and their relationships reduces comprehension by about 20%, even with a causally-coherent text (Stein et al., under review).

Forming a coherent conceptual representation from visual and verbal information places high demands on the learner's limited cognitive resources. As Mayer and Moreno (2003) discuss, the learner must organize the presented verbal information into a verbal model, the presented images into a visual model, and integrate these two representations into a coherent whole. There are several ways in which students can become overwhelmed during this process (see Mayer & Moreno, 2003, for an extended discussion). For one, each of the processing channels (verbal and visual) can become overloaded. Thus, understanding can be derailed early on by the complexity of processing novel scientific text and images. Another potential source of cognitive overload comes further downstream. If the student is unable to *simultaneously* hold the verbal and visual representations in working memory, then they will be unable to integrate them. Maintaining and combining representations in each channel is therefore critical as well.

Mayer and colleagues have investigated ways to improve multimedia learning by targeting the different sources of cognitive load. To reduce the burden on visual and verbal processing, Mayer and Chandler (2001) broke a science lesson into smaller units and gave the learner control over the pacing of the lesson. Compared to students who received the same information in one continuous stream, the students who could self-pace showed better learning and transfer of knowledge. Thus, self-pacing could reduce the burden on a student's limited cognitive resources and enhance their ability to form verbal and visual models from the lesson.

To reduce the burden on holding and integrating the visual and verbal models, Mayer and Moreno (2003) suggested that text and images should be presented simultaneously. Mayer and Anderson (1991), for example, found that students evidenced better transfer of learning when they received a lesson in which narration accompanied—as opposed to followed—an animation. When text and images are presented simultaneously, the student may be less likely to lose the visual or verbal representations that must be integrated to form a coherent understanding from a multimedia science lesson.

Self-pacing and temporal contiguity have been shown to reduce different sources of cognitive load in multimedia learning, yet common forms of instruction often involve a tradeoff between these two factors. Learning from an illustrated textbook, for example, is ubiquitous at all levels of education. This form of instruction may support self-pacing, since the student has control over how fast they read and progress. Yet, reading from a textbook is low in temporal contiguity, since the text and images are encountered separately. Another common form of instruction involves a tutor or instructor reading to the child. If a tutor takes the same learning materials (text and images) and reads aloud, then temporal contiguity is increased and it may be easier for the student to hold and integrate the visual and verbal components. Yet this could reduce or eliminate the benefits of self-pacing, since the tutor would hold some or all of the control over the pace of reading.

Purpose and overview of research

Given the potential tradeoffs inherent to different ways of presenting the same multimedia science lesson, we sought to test which task, self-reading vs. tutor-reading, provided the most benefit to learners at different grade levels, 4^{th} and 7^{th} grade. We adopted the causally-coherent text from Stein et al. (under review), which was developed using the principles of concept learning, causal coherence, and argumentation discussed earlier. We also used the same visual models as Stein et al., which served to visually illustrate characteristic molecular properties of the three states of water that were verbally described in the text. Because these learning materials are highly explicit and causally coherent, the burden of interpreting the text and images may be relatively low compared to a typical lesson on the same topics. Nevertheless, the content of the lesson may be novel and challenging for children. If interpreting the content of the text and images is the primary challenge that learners face, then a self-paced lesson may be more effective than tutor-paced lesson. However, if the main challenge of the lesson is holding and integrating information across verbal and visual modalities, then tutor-reading could be most effective. Indeed, Stein et al. (under review), which used tutor-read instruction exclusively, found evidence of impressive learning gains in this condition.

The effects of the different conditions could also depend on age. Older students may have greater metacognitive awareness (e.g., Flavell, 2000; Metcalfe & Shimamura, 1994; Schneider, 2008) in addition to greater reading skills and cognitive capacities. Thus, the 7^{th} graders may be more resilient to the demands of processing the visual and verbal components of the lesson and integrating them into a coherent representation. If so, then the effects of reading condition should be especially pronounced for the younger children, who may be most reliant on self-pacing or temporal contiguity.

In addition to the *Self-Read* condition and *Tutor-Read* conditions at each grade level, we also manipulated the nature of the visual models (static vs. dynamic) contained in the lesson as in Stein et al. (under review). It is possible that the effects of condition will be especially pronounced for one type of visual model, for example, the *Tutor-Read* condition may be especially effective when the visual models are dynamic, because the student is better able to attend to changes in the visual models over time. This may be less important in a static image. Thus, we had four experimental conditions: Tutor-Read/Static, Tutor-Read/Dynamic, Self-Read/Static, and Self-Read/Dynamic. Our control group included children who did not receive our instruction but instead received regular, "business-as-usual" classroom instruction.

Method

Participants

Participants were 158 fourth-grade children (*M* = 9 years, 11 months; range = 9 years, 0 months to 10 years, 8 months; 87 boys, 71 girls) and 172 seventh-grade children (*M* = 13 years, 2 months; range = 11 years, 9 months to 14 years, 8 months; 90 boys, 82 girls) recruited from four Chicago Public Schools. Participating schools were a classical magnet school, a math-science magnet school, an arts magnet school, and a neighborhood school. To enroll in magnet schools, children had to satisfy certain requirements specific to each school (e.g., standardized test scores in reading and math, or interest in an academic domain). To enroll in a neighborhood school, children qualified based on the geographical location of their parents' home address. The racial composition of the sample was 43% African-American, 21% Hispanic, 18% White, 10% Asian/Pacific Islander, and 8% Multi-Racial. This distribution roughly paralleled the overall distribution of ethnicity in the Chicago Public Elementary Schools, as we purposely intended.

Materials

We adopted two modules from the learning sequence developed by Stein et al. (under review). The first module introduced and defined matter, the three states of water, and the shape and volume of solid and liquid. This module focused on whether or not the observable properties (i.e., shape and volume) of solid and liquid water change when water is transferred from one container to another. It also explained that gas (i.e., water vapor) is invisible to the human eye, and that to learn about gas requires an understanding of molecules. The second module focused on the organization, speed, and movement of molecules that define each state of water, and then compared and contrasted these properties in each of the three states. Thus, the goal of the two modules was for children to understand that matter has properties that cannot seen by the human eye, that these properties can be visually modeled, and that the three states of water differ from one another in terms of the organization, speed, and movement of molecules. The shape and volume of water vapor were discussed *after* children learned about the molecular properties of the three states of water. Visual depictions of water vapor molecules allowed children to "see" how the invisible properties of molecules in a gaseous state enable gases to take on the shape or volume of any container.

Descriptions of the observable and molecular properties of each state of water were embedded in a causally-coherent sequence such that shape was discussed first, volume next, and the organization, speed, and movement of molecules third. During the presentation of the organization, speed, and movement of water vapor molecules, the changeability of shape and volume were discussed. Direct comparisons were then made between each of the three states, in terms of shape and volume, and the organization, speed, and movement of molecules. Descriptions for the shape and volume in each state are presented in Table 1. Descriptions for the molecular properties of the three states are presented in Table 2.

Table 1. *Observable Properties of Solid and Liquid Water*

	Shape	Volume
Solid	Constant	Constant
Liquid	Changeable	Constant

Table 2. *Molecular Properties of the Three States of Water*

	Organization	Speed	Movement
Solid	Locked in place	Vibrate and jiggle in place	Don't move out of lattice structure
Liquid	Close and "cling" to other water molecules	Moderate speed	Slip and slide around and over other molecules
Gas	Fill the entire container	Very fast speed	"Fly" around in all directions

The causal coherence of the text becomes important in describing and illustrating how heat energy regulates the speed and movement of molecules, which in turn determines the state of matter. The speed and movement of molecules increases in proportion to the amount of heat energy absorbed by the molecules. After each state was defined in terms of both observable and molecular properties, the three states were contrasted. A solid was presented first, with a description of the speed and movement of molecules. A liquid was presented next, with an explanation of how the speed and movement of molecules increase and why shape is flexible in liquids versus solids. The gaseous state was presented last, with a discussion of how an even bigger increase in energy leads to molecules breaking away from one another, moving rapidly in a random fashion, and taking up all of the volume of a closed container or escaping into the air if the container is opened.

We also adopted the visual models developed by Stein et al. (under review). The static graphics, in the form of JPEG files, presented either as single illustrations (e.g., the lattice structure of solid water ice molecules), or as a series of three snapshots, representing the beginning, middle, and end of an event (e.g., water as a gas being transferred from one container to another). Whenever possible, a series of three static pictures was used so that comparable content was presented in both the Static and Dynamic Graphics conditions. Thus, even though children in the Static Graphics condition never saw speed or movement conveyed dynamically, they did see three snapshots depicting the beginning, middle, and end points of each event sequence. The dynamic graphics, in the form of QuickTime movies, presented actual motion (e.g., liquid water molecules moving over and under one another, water vapor molecules rapidly moving in a container).

The learning modules and assessments were presented on individual MacBook Pro laptops. A data management program, "FileMaker Pro 8" was used to present the text and graphics, to collect pre and posttest assessment data, as well as to code all of the assessment responses. Each study session was audio-recorded on the laptop and on an iPod as a back-up recorder.

Design and Procedure

Children's receptive vocabulary and verbal ability were assessed using the Peabody Picture Vocabulary Test, third edition (PPVT-III). Children's performance on the PPVT was computed in terms of standard scores (4^{th} grade M = 103.54, SD = 14.77; 7^{th} grade M = 101.22, SD = 15.30) and percentile rank (4^{th} grade M = 58^{th}, SD = 29.20; 7^{th} grade M = 53^{rd}, SD = 30.13). These scores showed no difference between girls and boys in either grade. In each grade, children were assigned to one of the five conditions using a stratified randomization procedure. Stratification assignments were based on children's PPVT scores to ensure that vocabulary scores were normally distributed and equivalent across the five conditions at each grade level.

The four experimental conditions were (1) a *Tutor-Read/Static Graphics* condition, where the text was read aloud to the child in conjunction with the presentation of static graphics, (2) a *Tutor-Read/Dynamic Graphics* condition, where the text was read aloud to the child with the presentation of dynamic graphics, (3) a *Self-Read/Static Graphics* condition, where the child read the text aloud in conjunction with the presentation of static graphics, and (4) a *Self-Read/Dynamic Graphics* condition, where the child read the text aloud with the presentation of dynamic graphics. Children in the *Control* group received only pre and posttests, with the same period of time in between the tests as in the experimental conditions.

In each of the experimental conditions, children participated individually, with a trained tutor guiding each child through the learning modules and assessments. All text and accompanying graphics were presented on the computer screen, with the text on the left hand side of a computer screen and the accompanying graphics on the right hand side of the screen. In the Tutor-Read conditions, the experimenter read the text aloud to the child. The child was encouraged to read along silently, but was not required to do so. In the Self-Read conditions, the child read the text aloud to the experimenter. Children in all experimental conditions were also asked to attend to the embedded graphics.

Children in the four instructional conditions participated in five sessions over an eight- to ten-week period of time. In Session 1, the PPVT was administered and demographic data were collected. Session 2 consisted of a *Pretest* that assessed children's knowledge of the States of Water. Session 3 consisted of the presentation of the *First States of*

Water Module (on solids and liquids), immediately followed by an assessment of children's comprehension of the module. Session 4 consisted of the presentation of the *Second States of Water Module* (on gases and comparison across the three states), and a knowledge assessment immediately following the module. Session 5 consisted of the *Posttest* on States of Water knowledge. Post-testing occurred approximated three to four weeks after the completion of Session 4. Children in the Control condition completed all pretests (Sessions 1 and 2) and the posttest (Session 5). During the time between pre and posttests, children in the Control condition participated in their regular classroom instruction.

The pre and posttests included the same items, which were composed of: (1) true/false questions, (2) yes/no questions, (3) explanations for T/F and Y/N answers, (4) short answer questions, and (5) open-ended questions. We began by asking children to name the three states of matter. We then asked 10 questions for each state. For the purposes of our analyses, the relevant questions were the following (using solid water as an example):

1. Did the shape of the solid change when you transferred it from container 1 (short and skinny) to container 2 (tall and wide)? Why or why not?

2. Did the volume of the solid change when you transferred it from container 1 to container 2? Why or why not?

3. True/False: The solid changes shape as it is transferred from container 1 to container 2.

4. True/False: The solid changes volume as it is transferred from container 1 to container 2.

5. True/False: There are more solid molecules in container 2 than there were in container 1.

6. Do you know anything about the molecules that make up solid water ice?

7. What do you know about the molecules that make up solid water ice?

Scoring. FileMaker Pro automatically saved children's pre and posttest responses as the answers were typed into the computer. The computer program automatically scored responses to the T/F and Y/N questions. The remaining responses were scored manually (reliability among three coders was 96%). All questions concerning observable properties of water were T/F or Y/N, whereas all questions concerning molecular properties of water were open-ended, as described below.

For the *observable properties* (i.e., shape and volume) of solid and liquid water, we tabulated children's responses to the T/F and Y/N questions (questions 1-4 listed above). Thus, there were a total of 8 questions, all requiring dichotomous responses. Accuracy scores were computed as the proportion of correct responses out of 8. For the *molecular properties* of each state, children's answers to the three open-ended questions for each state (see question 7 above for solid) were scored with respect to ideal correct responses and "gist" responses, as described below.

Ideal correct responses were explicitly stated in the text. For each state of water, three components constituted a complete, ideal answer. The components focused on the organization, speed, and movement of molecules in each state (see Table 2). In addition to ideal responses, children also provided responses that were acceptable variations on the ideal correct responses (i.e., they maintain the "gist" of the components). For example, some children stated that solid water molecules are frozen in place rather than locked in place. These *gist correct responses* were coded as correct because they showed that children understood the conceptual content, even though they did not use the exact language provided in the text. All correct responses (i.e., ideal and gist), as well as examples of children's actual responses, are listed in Table 3.

Table 3. *Ideal and "Gist" Correct Responses for the Molecular Properties of the Three States of Water*

Solid Water	
Ideal correct responses	Examples
Molecules of solid water are locked in place	*The molecules of solid water [...] locked in place [...]*
Molecules of solid water vibrate; jiggle back and forth	*Molecules of solid water ice do not move, but they still vibrate*
Molecules of solid water do not move over and around one another	*The molecules just vibrate instead of moving around*
Gist correct responses	Examples
Molecules of solid water are frozen	*Um, the molecules [...] they're frozen in place*
Molecules of solid water move slower than molecules of liquid water	*The solid water ice molecules move slower than they would in water*
Molecules of solid water form a lattice structure	*The molecules are [...] in a lattice structure*
The molecular structure of solid water results in a fixed shape	*[...] they stay in the same shape they were put in before they were frozen.*
Liquid Water	
Ideal correct responses	Examples
Molecules of liquid water move at a moderate speed; faster than solid water molecules, but slower than water vapor	*The molecules of liquid water can move but not at a very fast speed [...]*
Molecules of liquid water move around, slip and slide over and under one another	*The molecules of liquid water [...] slide under and over each other [...]*
Molecules of liquid water cling to one another	*The molecules of liquid water stick together.*
Gist correct responses	Examples
Molecules of liquid water are loose, not locked in place	*The molecules of liquid water are not locked together-they are loosely packed*
Molecules of liquid water move around more than molecules of solid water; no mention of slipping and sliding	*The molecules of liquid water [...] move aroun because they aren't solid so they don't stay i place*
Molecules of liquid water do not move fast enough to break away from one another	*Move fast but not fast enough [...] to break far away from each other*

Molecules of liquid water are close together	*The molecules in liquid water […] stay next to each other.*
Water Vapor	
Ideal correct responses	Examples
Molecules of water vapor move very fast	*[…] moving really, really quickly*
Molecules of water vapor are able to break away from one another	*[…] they can break away from each other […]*
Molecules of water vapor fill any space in which they are placed (e.g., container, room)	*[…] spread out to make the same size as whatever it's in*

Table 3.(cont.) *Ideal and "Gist" Correct Responses for the Molecular Properties of the Three States of Water*

Gist correct responses	Examples
Molecules of water vapor move around freely in any direction, *without* reference to breaking away from one another	*Water vapor, the molecules can move in any direction they want […]*
Molecules of water vapor spread out all over, *without* reference to filling an entire space	*They spread out all over the place.*
Molecules of water vapor are loose, with a possible reference to other states	*They're really loose and not compact at all […]*
The lack of structure for water vapor molecules results in no fixed volume	*They have no fixed volume […]*
The lack of structure for water vapor molecules results in no fixed shape	*They have […] no fixed shape […]*

If a child generated at least one ideal correct response for a question, they were given one point. If they were unable to generate at least one ideal correct response for a question, they were given no points. This procedure was applied to each of the three molecular questions (i.e., for solid water, liquid water, and water vapor), and then the mean of the three scores was computed to obtain the proportion of ideal correct responses across all three states.

Results

Observable properties of solid and liquid water

We expected children to have some prior knowledge about the observable properties of solid and liquid water, especially the older children. Thus, pre-post gain scores on observable properties of solid and liquid should be relatively small compared to gains on learning about molecular properties of the three states. Nevertheless, we compared learning gains on observable properties across the grade levels and conditions.

The results were analyzed with a 2 (Grade: 4^{th} vs. 7^{th}) x 2 (Reading Condition: Tutor-Read vs. Self-Read) x 2 (Graphics Condition: Static vs. Dynamic) between-groups analysis of covariance (ANCOVA). The dependent variable was the pre-post gain score in the proportion of correct responses to Y/N and T/F questions about the shape and volume of solid and liquid water. Standardized scores on the PPVT were included as a covariate.

The analysis revealed a marginally significant effect of Grade, $F(1, 279) = 3.69$, $MSE = .07$, $p = .06$, $\eta_p^2 = .01$, with 4^{th} grade participants showing greater gain scores ($M = .23$, *SD* = .25) than 7th grade participants (*M* = .15, *SD* = .28). No other main effects or interactions approached significance, *F*s < 1.5, *p*s > .25.

In addition to the comparisons between the experimental conditions, we analyzed performance relative to the Control condition. The gain scores for the 7th grade Control group (*M* = .09, *SD* = .22) were marginally higher than those of the 4th grade Control group (*M* = -.02, *SD* = .18), *t*(39) = 1.78, *p* = .08. Gain scores for the 4th grade experimental conditions were significantly higher than those for the 4th and 7th grade Control conditions, *t*s > 2.12, *p*s < .05, with the exception of the 4th grade participants in the Self-Read/Static Graphics condition, whose gain scores (*M* = .18, *SD* = .29) were not significantly greater than the 7th grade Control participants', *t*(39) = 1.18, *p* = .12. Gain scores were generally lower in the 7th grade experimental conditions, as revealed by the ANCOVA. Only the Self-Read/Static Graphics condition (*M* = .19, *SD* = .26) had higher gains than the 7th grade Control condition, *t*(53) = 1.72, *p* < .05. The other 7th grade experimental conditions had higher gains than the 4th grade Control participants, *t*s > 2.70, *p*s < .05, but not the 7th grade Control participants, *t*s < 1.35, *p*s > .09.

Molecular properties of the three states of water

The results were analyzed with a 2 (Grade: 4th vs. 7th) x 2 (Reading Condition: Tutor-Read vs. Self-Read) x 2 (Graphics Condition: Static vs. Dynamic) between-groups analysis of covariance (ANCOVA). The dependent variable was a pre-post gain score in the proportion of ideal correct responses to questions about the molecular properties of the three states of water. Standardized scores on the PPVT were used as a covariate in the analysis. Gain scores were arcsine transformed for the analysis to adjust for the unequal variances between the conditions. The descriptive statistics reported below, however, represent the original scale of measurement.

The results are shown in Figure 1. The ANCOVA revealed a main effect of Grade, *F*(1, 279) = 6.35, *MSE* = .31, *p* < .05, η_p^2 = .02, with 7th grade participants showing greater gain scores (*M* = .70, *SD* = .36) than 4th grade participants (*M* = .64, *SD* = .39). There was also a main effect of Reading Condition *F*(1, 279) = 46.72, *MSE* = .31, *p* < .05, η_p^2 = .14, such that participants in the Tutor-Read condition had greater gain scores (*M* = .77, *SD* = .31) than participants in the Self-Read condition (*M* = .46, *SD* = .41). There was, however, no effect of Graphics Condition, *F*(1, 279) = 0.54, *MSE* = .31, *p* = .46, η_p^2 < .01. The analysis also revealed a marginally significant interaction between Grade and Reading Condition, *F*(1, 279) = 3.15, *MSE* = .31, *p* = .08, η_p^2 = .01. This trend is due to the fact that 4th grade participants showed a larger difference in gain scores between the Tutor-Read condition (*M* = .77, *SD* = .30) and Self-Read condition (*M* = .36, *SD* = .41) than the 7th graders (*M* = .70, *SD* = .36 for Tutor-Read; *M* = .55, *SD* = .40 for Self-Read). No other interactions approached significance, *F*s < 1, *p*s > .30.

Performance of the experimental conditions was also compared to the control conditions. The gain scores for the 4th grade Control group (*M* = .15, *SD* = .24) were equal to those of the 7th grade Control group (*M* = .16, *SD* = .38), *t*(39) = 0.10, *p* = .92. Gain scores for the experimental conditions were significantly higher than those for each of the Control conditions, *t*s > 2.20, *p*s < .05, with the exception of the 4th grade participants in the Self-Read/Static Image condition, whose gain scores (M = .16, SD = .38) were only marginally greater than the 7th grade Control participants', $t(42) = 1.28$, $p = .10$.

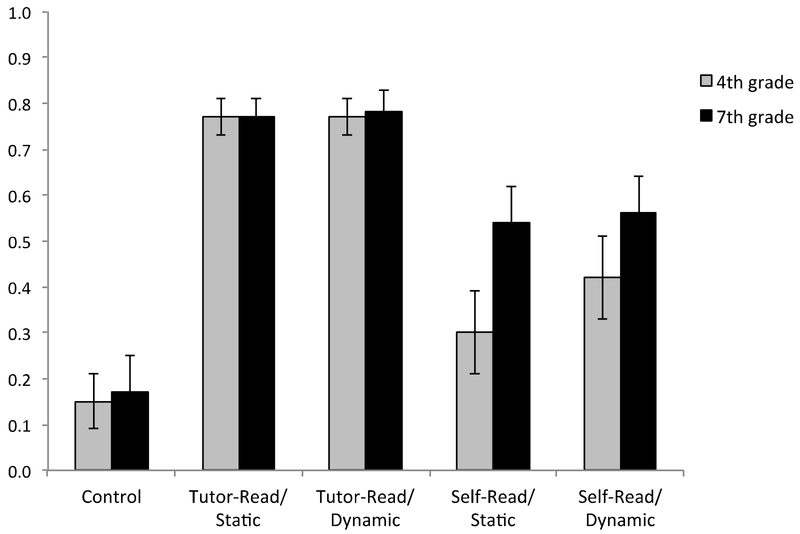


Figure 1. Mean gain scores in the proportion of ideal correct responses to questions about the molecular properties of the three states of water

Discussion

The present study examined the cognitive factors that influence children's learning about the observable and molecular properties of the three states of water by manipulating the delivery of a causally-coherent lesson (Tutor-Read vs. Self-Read) and the nature of the images that the children received (Static vs. Dynamic). The results revealed several important findings. First, children at both grade levels had some prior knowledge about the observable properties of solid and liquid water, and 4th graders showed the greatest improvement at posttest. Second, even though children at both grade levels began with little to no knowledge of molecular properties of the three states water, they were able to learn about these properties through the causally-coherent lesson. Third, even when vocabulary scores were statistically controlled, children in the Tutor-Read condition learned more than those in the Self-Read condition, and this difference was especially pronounced for 4th graders. Finally, at both grade levels children learned equally well regardless of the type of graphics (static vs. dynamic) they received.

In the Introduction, we characterized the two reading conditions in terms of their costs and benefits to different components of multimedia learning. Self-reading provides the benefit of self-pacing the lesson, but does not support integrating of the verbal and visual information. Tutor-reading forfeits control of the pace of the lesson, but the temporal contiguity of verbal and visual information processing supports integration of the two modalities. Our finding of overall greater learning gains in the Tutor-Read condition suggests that integrating the verbal and visual information was the greatest challenge to the children. When a tutor reads to the learner, integration, and thus learning, was enhanced.

It is interesting that the Tutor- vs. Self-Read effect was especially pronounced in the younger age group (though, as noted, this interaction was marginally significant). Younger children have poorer reading skill and metacognitive awareness than older children, and these variables could have contributed to the 4^{th} graders' exacerbated difficulties in the Self-Read conditions. Although we did not collect data on children's reading level and metacognitive ability, the Tutor- vs. Self-Read effect was found when children's vocabulary—a strong predictor of reading ability and general cognitive development—was statistically controlled.

In interpreting these results it is important to take into account potential limitations of the present study. The text that we used was unique because it was designed to be causally coherent. It is possible that children would have benefitted more from self-reading (and therefore self-pacing) if the text lacked this coherence (as is the case in most textbooks), because understanding the content would be more challenging. If the text lacked coherence, the learners would have been required to fill in gaps using their prior knowledge (McNamara et al., 1996), and the more controlled pace of self-reading could have facilitated this process. Another potential concern is that children in the Tutor-Read conditions may have been more engaged in the lesson than those who were self-reading. That is, it is possible that they paid closer attention to both the verbal and visual information and put more effort into integrating the two. However, a tutor was also present to oversee the children who were self-reading. It is equally plausible that children who were self-reading were more engaged because they had to read the text themselves, making the lesson more interactive for them. Finally, our analyses used the children's grade level/age as a proxy for cognitive control and capacity. Ideally, we would have a measure for each of these variables to test their contributions to the learning outcomes and to rule out other age- and context-related differences, such as everyday experience with water, parental/caregiver expertise in science, and the level of scientific discourse in the children's broader communities. These are important considerations for future research.

• • •

Acknowledgements

This work was supported in part by a grant from NSF 0529648. We are indebted to Marc Hernandez for his contributions. We would also like to thank the principals, teachers, and students involved in this study.

Florencia K. ANGGORO received her Ph.D. in Cognitive Psychology from Northwestern University. She is currently an Assistant Professor of Psychology at the College of the Holy Cross. She studies conceptual development, particularly the role of language, culture, and formal learning experiences in shaping children's concepts. Her particular interest is in concept acquisition in STEM domains.

Nancy L. STEIN received her Ph.D. in Psychology and Education from Stanford University. For over 35 years she has studied reading, models of comprehension, narrative and argument understanding, emotion, cognitive development, instruction, and learning science. Her current research in early science learning combines her expertise in education, instruction, and learning science with her background in mathematics and chemistry.

Benjamin D. JEE received his Ph.D. in Cognitive Psychology from the University of Illinois in Chicago. He is currently a Visiting Assistant Professor of Education at the College of the Holy Cross. He studies knowledge acquisition and expertise, with particular interest in the visual learning mechanisms that promote STEM learning.

References

Bar, V., & Galili, I. (1994). Stages of children's views about evaporation. *International Journal of Science Education, 16*(2), 157-174.

Duschl, R. A., Schweingruber, H. A., Shouse, A. W., National Research Council (U.S.). Committee on Science Learning Kindergarten Through Eighth Grade. National Research Council (U.S.). Board on Science Education & National Research Council (U.S.) (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, D.C.: National Academies Press.

Flavell, J. H. (2000). Development of children's knowledge about the mental world. *International Journal of Behavioral Development, 24*, 15–23.

Gobert, J., & Buckley, B. (2000). Special issue editorial: Introduction to model-based teaching and learning. *International Journal of Science Education, 22(9),* 891-894.

Goldberg, F., & Bendall, S. (1995). Making the invisible visible: A teaching/learning environment that builds on a new view of the physics learner. *American Journal of Physics 63*(11), 978-991.

Jose, T. J. & Williamson, V. M. (2005). Molecular visualization in science education: An evaluation of the NSF funded workshop. *Journal of Chemical Education, 82*(6), 937-943.

Klausmeier, H. J. (1992). Concept learning and concept teaching. *Educational Psychologist, 27*(3), 267. Larkin, J., & Simon, H. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65-99.

Mandler, J. M. (2008). On the birth and growth of concepts. *Philosophical Psychology, 21*(2), 207-230.

Mayer, R.E., & Anderson, B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology, 3*, 484-490.

Mayer, R.E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93*, 390-397.

Mayer, R. E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction. *Journal of Experimental Psychology: Applied, 11*(4), 256-265.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*, 43-52.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better?Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*(1), 1-43.

Metcalfe, J., & Shimamura, A. P. (1994). *Metacognition: knowing about knowing.* Cambridge, MA: MIT Press.

National Research Council (U.S.) (1996). *National science education standards: Observe, interact, change, learn*. Washington, DC: National Academy Press.

Project 2061 (American Association for the Advancement of Science) (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

Romance, N. R., & Vitale, M. R. (2001). Implementing an in-depth expanded science model in elementary schools: Multi-year findings, research issues, and policy implications. *International Journal of Science Education, 23*, 373-404.

Romance, N. R., & Vitale, M. R. (2010). Effects of an integrated instructional model for accelerating student achievement in science and reading comprehension in grades 1-2. Presented at the Annual Meeting of the American Educational Research Association, Denver, CO.

Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain, and Education, 2*, 114–121.

Shwartz, Y., Weizman, A., Fortus, D., Krajcik, J., & Reiser, B. (2008). Middle school science curriculum: Coherence as a design principle. Paper presented at the National Association of Research in Science Teaching.

Slotta, J. D., & Chi, M. T. H. (2006). The impact of ontology training on conceptual change: Helping students understand the challenging topics in science. *Cognition and Instruction, 24,* 261-289.

Stein, N. L., Hernandez, M. W., & Anggoro, F. K. (2010). A theory of coherence and complex learning in the physical sciences: What works (and what doesn't). In N. L. Stein & S. W. Raudenbush (Eds.), *Developmental science goes to school*. New York: Taylor and Francis, Inc.

Stein, N. L., Hernandez, M. W., Anggoro, F. K., & Hedberg, E. (under review). A developmental study of physical science learning: Modeling invisible processes and starting early.

Stein, N. L., & Levine, L. J. (1989). The causal organization of emotional knowledge: A developmental study. *Cognition & Emotion, 3*(4), 343-378.

Stein, N. L., & Miller, C. A. (1993). A theory of argumentative understanding: Relationships among position preference, judgments of goodness, memory, and reasoning. *Argumentation, 7*, 183-204.

Stein, N. L., & Trabasso, T. (1982). What's in a story: An approach to comprehension and instruction. In R. Glaser (Ed.), *Advances in Instructional Psychology, 2*, 212-267. Hillsdale, NJ: Lawrence Erlbaum Associates.

Thagard, P. (2000). *Coherence in thought and action*. Cambridge: The MIT Press.

Trabasso, T., & Bouchard, E. (2002). Teaching readers how to comprehend text strategically. In Block, C.C. & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 176-200). New York: The Guilford Press.

Trabasso, T., Secco, T., & van den Broek, P. W. (1984). Causal cohesion and story coherence. In H. Mandl, N. L. Stein & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 83-111). Hillsdale, NJ: Lawrence Erlbaum Associates.

Trabasso, T., & Stein, N. L. (1997). Narrating, representing, and remembering event sequences. In P. W. van den Broek, P. J. Bauer & T. Bourg (Eds.), *Developmental spans in event comprehension and representation: Bridging fictional and actual events* (pp. 237-270). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tversky, B., Heiser, J., Lozano, S., MacKenzie, R., & Morrison, J. (2008). Enriching animations. In R. Lowe & W. Schnotz (Eds.), *Learning with animation*. Cambridge: Cambridge University Press.

Vosniadou, S. & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology, 24*, 535-585.

Winston, P. (1986). Learning by augmenting rules and accumulating censors. In R. S. Michalski,

Carbonell, J. G., & Mitchell, T. M. (Ed.), *Machine learning: An artificial intelligence approach*. New York: Morgan Kaufmann.