Assessing the Development of Expertise in an Historical-Based Science: The Case of Integrative Archeology

Inbal Flash-Gvili, Jeff Dodick, Science Teaching Center, The Hebrew University of Jerusalem, Israel
finbal@cc.huji.ac.il, jdodick@vms.huji.ac.il

Developing inquiry-based skills in science students is a key focus of science education. This study examines the process whereby such skills are acquired by exploring the case of graduate training in the discipline of integrative archeology. To do this, we focus on the questions that the students posed during their field research. Our results show that some students attempt to bypass an initial phase of their research by trying to solve the final goal of reconstructing human behaviors, rather than focusing on material remnants that make such reconstruction possible. In the field, this was observed when the students attempted to reconstruct events via their own logic, rather than using the technical tools at their disposal to analyze the site's (chemical) properties. However, as their expertise grows, the students learn about the importance of foundational questions; this is reflected in a change in the frequencies of their questions, which in turn reflects a change in their research strategies.

Introduction

Scientific inquiry can be defined as "the diverse ways in which scientists study the natural world and propose explanations based on the evidence derived from their work" (NRC, 1996 p. 23, 2000 p.1); thus, the NRC defines inquiry-based learning as a process where "students actively develop their understanding of science by combining scientific knowledge with reasoning and thinking skills" (NRC, 1996 p. 2). The same source also claims that: "Inquiry into authentic questions generated from student experiences is the central strategy for teaching science" (NRC, 1996 p. 31). These definitions create a picture of a type of learning strategy, which should reflect the nature of science via scientific inquiry. Yet, the question remains as to what exactly are 'the correct' characteristics of scientific inquiry in practice?

A simple model of scientific inquiry is based on *the scientific method*. Broadly speaking, it is described as a process in which the following events occur in sequential order: (1) observing, (2) developing a question, (3) developing a hypothesis, (4) conducting an experiment, (5) analyzing data, and (6) stating conclusions. This process continues as new questions are generated based on previous findings. This model, with different variations is repeated in many of the US standards documents (NRC, 1996, 2000).

However, since its development, this model, with its rigid series of stages, has been criticized by philosophers and science educators, who have produced alternative lists of process skills that were considered more flexible. An example of such an alternative list is presented in the *Science: A Process Approach* program (AAAS, 1967). One example of an alternative process skill is controlling variables, a strategy which has received enormous attention in research on the development of scientific reasoning (Kuhn and Dean, 2005). Still, these lists are problematic as they define *the scientific method* as a coherent entity with regards to different scientific disciplines (such as geology or physics). This problem is reflected in the way scientific inquiry is defined by the NRC (1996; 2000). From one side, it talks about multiple methods; yet, concurrently it adheres to a single scientific method. Thus, if one of our science education goals is to create future experts by enhancing their ability to acquire scientific inquiry skills, we need to teach in a much more authentic way that reflects the methodological diversity of the sciences. As Sternberg (2003) suggests:

> If we wish to teach and identify expert students, we need to identify expertise in a way that is closely aligned with the way experts are identified in the disciplines students study. For starters, this means having students do tasks, or at least meaningful simulations, that experts do in the various disciplines. Second, it means teaching them to think in ways experts do when they perform these tasks.

Following Sternberg's (2003) recommendations, if we want to understand the development of expertise, we must first characterize the tasks that experts do in different disciplines, as well as the inquiry skills that they apply while engaged in those tasks. One research method that has been used to understand expert inquiry skills in the sciences is by studying the practices of scientists while they pursue their research agenda in real time (the *in vivo* method). Thus, in this study, we ask the following question: how do inquiry skills develop within a team of graduate student researchers in the *historical-based* science of integrative archeology? To answer this question we followed these students as they conducted their research.

There are many instruments for assessing the development of scientific inquiry skills, and in this study we used (among other sources of data) the scientific questions our research subjects generated about their field-based work. Students' ability to ask high-level questions about scientific phenomena has been shown to be a good indicator of science learning in students from the high school to university level (Brill & Yarden, 2003; Chin, 2004; Hofstein et al, 2005; Marbach-Ad & Sokolov, 2000). However, such question-asking must be put in context; to do this, we will also describe the research process of the team as a whole, as well as add important observations from the field and our interviews with the team members which further validate this analysis.

Understanding the Development of Scientific Expertise

Our understanding of what it means to be an expert in a domain has grown ever since DeGroot's (1946) classical studies of chess players. Thus, Chi and Glaser (1988) described experts as excellent performers who have superior short and long-term memory and represent problems in deeper, more principled way than novices who tend to build superficial representations. Concordantly, it was found that experts spend more time on constructing a good problem representation, while novices applied a suboptimal trial and error strategy (Van Gog et al., 2005).

Many theories were developed to explain these differences in performance (e.g.: Ericsson et al, 1993; Gobet and Simon, 1996). One such theory, the Model of Domain Learning (MDL) suggested by Alexander (1997), is important to our work because it portrays the nature of developing expertise in authentic academic domains rather than extracting it from carefully chosen problems in diverse nonacademic domains. In contrast to the traditional expert-novice theories, this model assumes that there are no sharp contrasts between experts and novices; rather, it is a graded transition toward expertise. It also considers both cognitive and motivational aspects as important factors in the expertise learning process, in contrast to traditional models which are "coldly cognitive" (Pintrich et al., 1993). Thus, it focuses on three components (i.e., knowledge, strategic processing, and interest) which are interrelated and play a role in this transition. In this paper, we focus on the first two components.

According to this model acquiring expertise has three stages (i.e., acclimation, competence, and proficiency/expertise). The most critical differences between experts and novices is the development of a broad and deep knowledge base, a shift in the kind of strategies used, from surface-level to deep-processing, and the increase in individual interest which permits experts to maintain a high level of engagement over time. The concurrent development of these components allows experts to be actively engaged in problem finding, posing questions, and instituting investigations that (sometimes) push their domain boundaries.

In terms of expertise studies examining scientific thinking in particular, the research have gone through considerable changes since its earliest period where the focus was on testing circumscribed aspects of the scientific discovery process (e.g. Wason, 1968). Within the last two decades, researchers have used simulated discovery tasks in complex domains in order to track participants as they explore, test hypotheses via experimentation, and acquire new knowledge in the form of revised hypotheses (Schauble, 1996).

However, such empirical research has a number of problems when used to generalize about scientific reasoning, such as the fact that scientific research is a collaborative enterprise taking months or even years to complete (Dunbar, 1995). Thus, researchers in the field of cognitive psychology (Dunbar 1995; Nersessian et al., 2003) and education (Bond-Robinson and Stucky, 2005; Feldman, 2008; LaPidus, 1997) have entered the laboratory to study scientists as they pursue research in real time.

Unfortunately, almost all of this research is weighted towards laboratory-based, experimental sciences; in fact, the only studies centered on field-based sciences are by Bowen and Roth (2007) in their examination of field ecologists. Thus, in this study we decided to focus on a field-based science with a strong *historical* component in order to better understand how expertise is learned in such fields.

As opposed to experimental-based sciences, such as chemistry which pursue experiments on natural phenomena under controlled laboratory conditions, the goal of *historical sciences* such as geology is to reconstruct past phenomena based on (mostly) un-manipulated evidence gathered as traces from the field. In fact, historical-based scientific methodologies were specifically developed to cope with problems that could not be solved experimentally. This has required the development of a whole set of methodological tools (both theoretical and practical) that help historical-based scientists overcome the constraints of their field-based evidence (Dodick and Argamon, 2009).. It has also affects the way graduate students are trained in the expert methodologies that historical scientists use to solve research problems, as we will show here.

Methods

The Research Sample

The research field of the team we investigated is *Integrative archeology*; a relatively new scientific domain, its goal is to use the (chemical and physical) properties of the materials accumulating at an archaeological site to better understand the humans who inhabited that site. Such materials include minerals, and organically produced materials such as bones, teeth, shells and plant remains.

The team's advisor (senior Professor) is trained in geology, biology and chemistry and has published extensively in the scientific literature. His team, at the time of this study consisted of two interacting groups:

expert scientists (2 junior colleagues), and novice scientists (1 M.Sc. and 3 Ph.D. students). Additionally, technicians, post-docs and summer students joined the team during summer fieldwork sessions.

As he advocates an interdisciplinary approach, the advisor accepts graduate students from most scientific fields; this approach is reflected in the team's background. "E" and "R", the two junior colleagues are trained in physics and archeozoology, respectively. Amongst the Ph.D students, "L" completed an M.Sc in Molecular biology; "D" has an M.Sc in Archeology and "A" in Materials Science. In this study we focus our analysis on the advisor and his 3 Ph.D. students.

Data Collection and Analysis

The team pursues fieldwork at least once a year in order to collect samples for their analyses. To collect our data, we joined the groups' excavations (which typically lasted three-weeks during the summer) during a 4.5-year research period (2004-2009). We videotaped nearly 50 hours of interactions among research team members as they gathered and analyzed data in the field. Extensive field notes complement the video data. We also recorded the groups' seminars where members presented their works.

Semi-structured interviews were held with the team members after each field season. The interviews informed us of the researchers' individual feelings and understandings about various aspects of their research. It was also used to refine and validate our understanding of the field observations. All of the interviews were recorded and transcribed.

The interviews provided us with rich amounts of data touching upon a considerable number of themes which were analyzed using Shkedi's (2004) constructivist (ethnographic) method of qualitative research. In the first stage of this analysis, primary categories were developed following a first reading of the interviews allowing an initial coding. Careful attention to these categories produced a focused emergent framework, and the data was then recoded according to that framework. Our analysis concentrated on the major research challenges that the group members faced, as well as the strategies they employed to cope with these challenges.

A second analysis was done on the interviews and field observations using the questions that the advisor and students generated spontaneously about their research. Questions are identified as "an interrogative sentence or a declarative sentence with an embedded interrogative" (J. Dillon pers. comm.).

Two coding schemes were used in this analysis: (1) An emergent scheme (Table 1) which reflects both the questions connected to the research goals of the team, as well as to the strategies the advisor mentioned as promoting research. (2) A deductive scheme (Table 2) based on Dillon's (1984) classification of (scientific) research questions. Inter-rater reliability of these two schemes averaged 90%.

Table 1: Emergent categories of the questions asked by the team members.

Category and Definition	Example
Validity questions are concerned with the extent to which the results of a research study can be generalized to situations beyond those involved in the study (*External validity*) **or** about the extent to which extraneous variables have been controlled by the researcher, so that any observed effects can be attributed solely to the "treatment" variable (*Internal Validity*).	What happened to the organic material since its initial deposition? (This question concerns whether the material changes were caused by chemical degradation.)
Methodological questions are concerned with procedures needed to solve a specific scientific problem.	How do we date plaster?
Human Behavior questions are concerned with attributes connected to human artifacts (e.g. Ceramic tools) or activities (e.g. Ways of cooking).	What were these ceramic vessels used for?
Materials questions are concerned with attributes of materials (e.g. color, IR spectra measurements).	What will be the ν-*Ratio* of the samples?
Space and Time questions are concerned with the spatial and temporal relationship between items such as Loci (areas), Layers, and Samples.	Which Archeological layer is above which?

Table 2: Deductive scheme based on Dillon's (1984) categories for classifying scientific research questions

Category and Definition	Example
Property questions are concerned with the properties of a specific "thing" (such as its shape, color, location).	What are the organic materials that were in this ceramic vessel?
Comparison questions deal with the comparative attributes between two "things".	What is the difference between the Calcite which was heated and the one that wasn't?
Contingency questions deal with the relationship (relational, correlative, or causal) between two "things".	Are the Copper rich sediments in situ or are they a product of post-depositional process of melting in water and going through the sediments?

Results and Discussion

The Team's Research Process

Even though fieldwork happens only for a short period each year, it is critical to the team's research process (Figure 1). In the first research stage, physical and chemical traces of human activity are collected, analyzed in the (portable) lab and than reapplied to the field in order to locate a relevant research problem. In the second research stage, specific traces are collected in selected locations, followed (sometimes) by simulations in the lab and reapplication of the results to the field in order to solve the research problem. Both stages require at least two iterations, because lab analyses often require the team returning to the field and re-sampling a specific area in order to clarify and validate results emerging from the initial analysis.

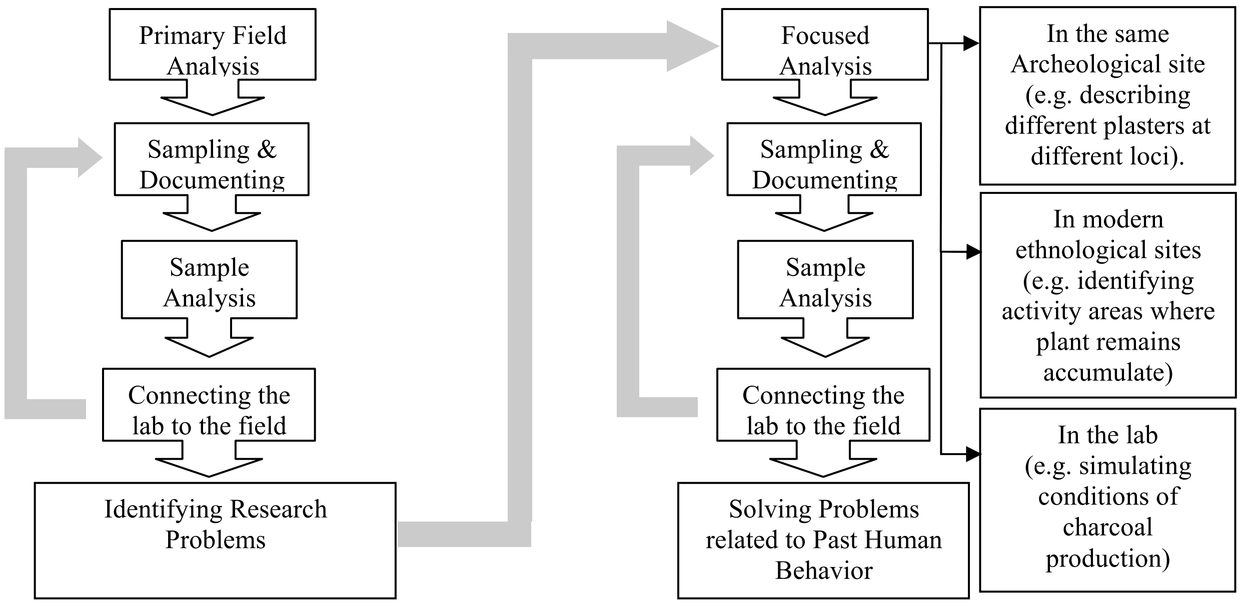

Figure 1: A schematic model of the team's research process

An example of a typical research problem was based on the observation that at certain site locations many of the sediments analyzed by IR-spectroscopy seemed to have been chemically altered by exposure to high temperatures. Solving this issue involved fully characterizing the phenomenon at different geographical and stratigraphical locations of the site, as well as modeling the effect of heat on sediments under both lab (using ovens) and field conditions (using natural fire); such simulations are intrinsic to the methodology of historical sciences. The aim of this problem-solving process was to create a tool enabling the researchers to reconstruct human behaviors from field materials, both at this site, and possibly at other sites as well.

Assessing the Development of Expertise amongst the Students

Our data indicates that the advisor strongly stresses (24 times in a total of five interviews) the importance of problem identification as possibly the most critical element of field research and the most difficult one to teach (Dodick and Flash-Givili, 2008). Typical is his comment after the 2004 field season.

> This is "the state of the art" or "the name of the game" in science! It's not to collect observations, to summarize the details and to publish it; however, in these complex situations, without a direction, to identify the “thread” that will lead to something significant. If this would be something that I could teach in a lecture or a class...it would have been useful, than I think I don't need to teach further; the rest is technique."

We frequently noticed the advisor teaching this principle to his students. A typical example is his feedback (from field season 2006) to two of his students who immediately tried to solve the questions connected to a specific section even before mapping it and identifying the questions it elicited. "This is the wrong picture! In other words, even if we where there [at the problem solving stage] it's not for... I mean, if it's a serious problem we will do it, we will solve it, we will ask these questions, but maybe it's not even stratified..". Later, when asked to reflect on this event, the advisor said: "The big message is that you have to use the tools in order to do it in iterations, every time having a smarter question. Not in an automatic way." Thus, 'listening' to the questions presented by the field is a leading research strategy for him.

These observations led us to hypothesis that if we analyzed the questions posed by the students we would be able to better understand the progress they made in their research. Hence, we characterized the frequencies of questions asked by the three Ph.D. students’ in the interviews (Table 3). Note that in 2006, the researchers surveyed a second site, in addition to the main focus site with which they began their research. It is important to take into consideration though, that this quantitative account is just a raw indicator of trends. This is not meant to be a fully quantitative study because of the small group size.

<u>Table 3: Changes in the types of questions posed by the 3 Ph.D. students after each field season based on our emergent categories (presented as percentages of the total)</u>

		Interview 1	**Interview 2**	**Interview 3**	**Interview 4**
		2004-2nd year*	2005-3rd year	2006-3rd year	2007-4th year
		N=13	N=24	N=17	N=15
	Materials	38%	17%	40%	40%
Student A	Hum. Behavior	62%	29%	12%	27%
(PhD)	Space & Time	0%	37%	24%	20%
	Validity	0%	0%	18%	0%
	Methodology	0%	17%	6%	13%
		2006 - 1st year*	2007-2nd year*	2008-3rd year	2009-4th year
		N=15	N=20	N=19	N=12
	Materials	20%	35%	52%	50%
Student L	Hum. Behavior	13%	5%	5%	0%
(PhD)	Space & Time	53%	20%	0%	8%
	Validity	7%	10%	16%	0%
	Methodology	7%	30%	27%	26%
		2005-4th year	2006-4th year		
		N=30	N=23		
	Materials	13%	17%		
Student D	Hum. Behavior	30%	17%		
(PhD)	Space & Time	33%	43%	NA (Ph.D.	NA (Ph.D.
	Validity	17%	17%	completed)	completed)
	Methodology	7%	6%		

Looking at the data, a number of important trends are seen. In the case of “A” there was a strong tendency at the beginning of her work to focus on Human (cultural) behavior. This is well reflected in the following quote from an interview after field season 2004 (the second year of her Ph.D.):

> I want to work on things: ceramics, tools, kilns, installations, technology, to understand technology, to see what happens to it, which materials did they use? Why did they use them? Where did they take it from? Those kinds of things…

These types of questions, as she noted, are not encouraged by the advisor as he prefers asking questions which will enable his team to create a universal tool, or research strategy which can be applied to many sites. However, even though she recognized his intentions, she still insisted on focusing on human behavior questions. This may have been due to her misconception that her immediate task, as an archeological problem-solver was to reconstruct human behaviors of the past. This also correlates with our field observations which showed that at the beginning of her Ph.D. this student worked very closely with the traditional archeologists, and this too may have swayed her thinking towards human behavioral problems. In fact, her advisor mentioned that at the beginning of her work, he needed to "separate" "A" from the archeologists".

However, the percentage of this question type tended to tail off through the third year of her Ph.D., as she began to understand the importance of analyzing the properties of field materials both in terms of content and their spatial-temporal relationships. This is clearly seen in the following quote from field season 2007, which was the fourth and final year of her Ph.D. research.

> Today I don't need someone to explain to me to know that this [layer] cuts this and this cuts the other, that this is early and this is late. This might sound trivial, but it wasn't at all like that for me at the beginning. I also have a little bit more understanding of the processes, of how they occur. Which kind of by-product you expect from a pile of dung which had disintegrated, or from metal that stayed in the ground. What to expect if I have a collection of metals? How the does the sediment around it look? All these little contexts are much clearer to me today.

In contrast to "A", "L" asked a smaller frequency of human behavior questions even at the beginning of his Ph.D. and the frequency of such questions rapidly declined through his research. Concomitantly, we see an increase in the percentages of materials, validity and methodological questions; this increase connects well to his interest in methodological problems, in which the ultimate goal is to create a tool that might be used to better reconstruct the temporal background of a site. As he noted in the first year of his Ph.D. studies: "If we could invent a scientific physical method that will really prove that one ceramic artifact is present before the other [...] it's a very nice idea to do". Indeed, "L's" research focused on validating a new tool for dating plaster compounds. Interestingly, at the end of 2008, his research focus shifted toward characterizing the chemical properties of plaster with no connection to developing a new tool. This change was forced upon him by circumstances in the field, as he encountered some major difficulties in the dating project, while at the same time he started identifying extraordinary examples of plaster in a specific site.

In essence, the shift in "A" and "L's" questions represent the influence of the advisor's research philosophy which emphasizes the importance of starting a project with a field-based material analysis before jumping to questions of human behavior or trying to invent a new tool. However, at least in the beginning of their studies, his students had a tendency to pose broad questions (i.e. what did people do here?) and immediately tried to solve them. Moreover, they were inclined towards using "actualistic" logic to answer their questions even before collecting data in the field. (Actualistic logic is used to reconstruct the past based on knowledge of present conditions. It is a widely used in all historical sciences). Indeed, the advisor often had to warn them "not to use their logic to come to a conclusion". Instead, he emphasized the importance of "using logic to ask a question" and "to try and get the data to support it".

Finally, "D" like her colleagues shows a decrease in human behavior questions, but unlike "A" and "L" she shows almost no change in materials-based questions. This may be due to her archeological background, which made it hard for her to look at the field from a materials perspective. Indeed, she admitted that although she had learned much in her Ph.D. she still needed to fill in considerable gaps in her understanding of chemistry. This connects well to the way her advisor characterized her progress towards her Ph.D., which was in the direction of becoming a "lab specialist"; this, contrasts with his research strategy which integrates both field and lab findings.

Another research element emphasized by the advisor (2006 interview) is connected to understanding the spatial and temporal context of the field, as he noted: "Excavation for everybody is the bread and butter of Archeology. I excavate, I can decipher the stratigraphy and can complete the picture. If you are doing mistakes in deciphering the stratigraphy than it's a big mess, that nobody can work out". In Table 3 we see that "A" asked no spatial-temporal questions at the beginnings of her Ph.D. because as she admitted she "didn't even understand where to start observing". This is supported by the lack of any stratigraphic diagrams in her field notebook at this time. It was only in the third year of studies that she began to ask relevant stratigraphic questions from her colleagues, and even began to draw field-based diagrams. In contrast with "A", "L", was highly focused on developing a good understanding of the spatial-temporal relationships of the field, and this is seen by the large frequency of questions he asked from the beginning of his studies. In fact, his advisor (interview, 2006) noted that he was “ready made for this type of work”. Finally, Student “D” was very focused on spatial-temporal relationships. This correlates well with her previous research experience in archeology.

We also used Dillon's (1984) scheme for classifying the questions; here, questions are classified to three levels: Properties (I), Comparisons (II), or Contingencies (III). The purposes of this teams’ research are identifying archeological materials and reconstructing human behavior; in contrast, the other types of questions (Validity, Methodology, and Space and Time) serve these purposes. For this reason, we decided to use Dillon’s (1984) scheme to focus in on materials and human behavior questions.

Our results show a very strong trend in (almost) every year, and every student towards level I Properties questions, both in the case of materials and human behavior questions. "A"'s Properties questions averaged 64% of her total materials-based questions across all the years we observed her, while her Properties questions devoted to human behavior averaged 71%. This situation repeats itself with "L" (77% and 100% respectively) and "D" (86% and 70% respectively). This supports the fact that the advisor is very careful in establishing the baseline properties of archeological materials before jumping to causal relationships (Level III questions). For example, one of his big research projects has been examining the material contents of an ancient (garbage) pit. However, much of the discussion connected to this work has been on defining the nature of “pits” in general. A similar issue arose around the definition of “metallurgic kilns” which was part of “A’s” research.

These results show that a good scientific research program does not necessarily have to concentrate on causal questions; establishing a phenomenon's properties is critical in of itself. This is especially important in field-based, historical sciences, where one doesn't control the evidence offered by the field; this forces the scientists to collect every important trace, while validating its properties prior to answering any causal question. In agreement with this observation, student "L" described the most important quality of this group's method as "Differentiating between [the phase of] documenting the data from its interpretation; first explain what was found, bring pictures, document everything, do everything properly and only after that stage, write in the conclusion what you think it is". This emphasis is also supported by the team’s publications, most of which have a strong descriptive character. The advisor hopes that his method will be applied to other archeological problems, providing his work with greater validity.

Finally, thematic analysis of the advisor’s interviews shows an abundance of strategies that he uses to identify research problems demonstrating that he is highly engaged in problem identification. Such strategies are either derived from the *disciplinary culture* (i.e. Identifying gaps in experts' knowledge) or the *data* itself (i.e. Following unexpected findings; Focusing on prominent data). However, at least at the beginning of their studies, most of the students’ strategies were not derived from *data* but rather from the *disciplinary culture* wherein they focused on problems previously identified by others; in other words they often expected their problem to be well defined by others, before collecting data. This is not surprising in that this is the first time that these students received an opportunity to fully develop a research program. Still, as they progress through their studies, the students adopt a more experiential based approach (i.e. focusing on data they know how to analyze, based upon 'what the field offers'). This shift demonstrates their transition towards greater expertise.

Our results correlate well with Feldman (2009) where he classified the highest level of graduate science research as creating "Knowledge Producers" who can, among other things, independently formulate research questions. Similarly, Alexander (1997) also defined the highest level of academic research (in general) as proficient experts who are actively engaged in problem finding. It also connects well with Van Gog et. al.’s (2005) research which showed that in contrast to novices, experts invest far more time representing a problem before solving it.

Nevertheless, our case study shows that such problem identification can be circumvented by the students’ misunderstanding of where their research truly begins, as they believed that their primary goal was solving human behavioral problems, when in fact they (first) needed to do a complete material analysis. Simply put, they attempted to by-pass a significant research stage. By misinterpreting their research goals they also sidetrack the process of formulating the questions that would ultimately guide their studies by immediately jumping into a problem solving mode. This is reflected by how their questions change over time.

These results emphasize that the first goal of inquiry learning, i.e., "identifying questions that can be answered through scientific investigation" (NRC, 1996) is a complex skill that novices need to develop before they become scientific experts. Thus, even before dealing with the second goal of inquiry learning, "design and conduct a scientific investigation", amongst which controlling variables is listed as a major skill, we need to invest more time in supporting students with experiences that will develop their ability to formulate effective research questions (Kuhn and Dean, 2005); we certainly cannot expect them to develop this skill by themselves.

Obviously this is a small, focused case study so more research needs to be done to provide increased validity to our results. Thus, we have started a comparative study analyzing another field based science – ecology. This will permit a fuller description of some of the in-depth problems that students experience while learning to become expert field scientists.

References

Alexander, P.A., (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher*, 32(8), 10–14

American Association for the advancement of Science (1967). *Science – A process approach*. New York: Xerox Division, Ginn and Company.

Bond-Robinson, J., Stucky, A.P. (2005). Grounding scientific inquiry and knowledge in situated cognition. In: *Proceedings of the 27th Annual cognitive science society* (pp. 21–24), Stresa, Italy.

Bowen, G.M., Roth, M.W. (2007). The practice of field ecology: insights for science education. *Research in Science Education,* 37, 171–187.

Brill, G., & Yarden, A. (2003). Learning biology through research papers: A stimulus for question-asking by high-school students. *Cell Biology Education*, 2, 266– 274.

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

Chin, C. (2004). Students' questions: Fostering a culture of inquisitiveness in science classrooms. *School Science Review*, 86(314), 107– 112.

DeGroot, A. D. (1946). *Het Denken van den Schaker. (Thinking Processes in Chess Players.)* Noord Holland, Den Haag, The Netherlands.

Dodick, J., & Argamon, S. (2009). Understanding scientific methodology in the Historical and Experimental sciences via language analysis. *Science and Education*, 18 (8), 985-1004.

Dodick, J., and Flash-Gvili, I. (2008). Challenges to graduate student research in the historical based sciences. Proceedings of the National Association of Research in Science Teaching Meeting, Baltimore, MD.

Dunbar, K. (1995). How scientists really reason: scientific reasoning in real-world laboratories. In: Sternberg RJ, Davidson J (eds), *Mechanisms of insight* (pp. 365–395). Cambridge MA: MIT Press,.

Ericsson, K.A., Kramp, R.Th., & Tesch-Romer, C. (1993) The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363-406.

Feldman, A., Divoll, K., Rogan-Klyve, A. (2009). Research Education of New Scientists: Implications for Science Teacher Education. Journal of Research In Science Teaching, 46 (4), 442–459.

Gobet, F., & Simon, H.A. (1996) Templates in chess memory: A mechanism for recalling several boards, *Cognitive Psychology*, 31, 1–40.

Hofstein, A., Navon, O., Kipnis, M., & Mamlok-Naaman, R. (2005). Developing students' ability to ask more and better questions resulting from inquiry-type chemistry laboratories. *Journal of Research in Science Teaching*, 42, 791-806.

Kuhn, D., & Dean, D., Jr. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science*, 16, 866–870.

LaPidus, J.B. (1997). *Doctoral education: Preparing for the future*. Washington, DC: Council of Graduate Schools.

Marbach-Ad, G., and Sokolove, P.G. (2000). Can undergraduate biology students learn to ask higher questions? *Journal of Research in Science Teaching,* 37, 854–870.

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

National Research Council (2000). *Inquiry and the national science education standards*. Washington, DC: National Academy Press.

Nersessian, N.J., Newstetter, W.C., Kurz-Milcke, E., Davies, J. (2003). A mixed-method approach to studying distributed cognition in evolving environments. In: *Proceedings of the international conference on learning sciences* (pp. 307–314).

Pintrich, P. R., Marx, R., & Boyle, R. (1993). Beyond "cold" conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research*, 63, 167–199.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. Developmental Psychology, 32, 102–119.

Shkedi, A. (2004). Words of Meaning: Qualitative Research-Theory and Practice. University of Tel Aviv Press: Tel Aviv.

Sternberg, R.J. (2003). What is an "expert student?" *Educational Researcher*, 32(8), 5-9

Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology*, 19, 205-221.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology,* 20, 273-281.

Acknowledgements: This research was supported by Israeli Science Foundation Grant 731/06, the Levi Eshkol Institute for Social Sciences Research Award of the Hebrew University of Jerusalem and the Municipality of Jerusalem Student Research Scholarship.