Elementary students' laboratory record keeping during scientific inquiry

Garcia-Mila, Merce; Andersen, Christopher; Rojo, Nubia E.

Postprint / Postprint
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
www.peerproject.eu

Empfohlene Zitierung / Suggested Citation:
Garcia-Mila, M., Andersen, C., & Rojo, N. E. (2010). Elementary students' laboratory record keeping during scientific inquiry. *International Journal of Science Education*, *33*(7), 915-942. https://doi.org/10.1080/09500693.2010.480986

Elementary students' laboratory record keeping during scientific inquiry

Journal:	*International Journal of Science Education*
Manuscript ID:	TSED-2009-0346.R3
Manuscript Type:	Research Paper
Keywords:	elementary school, reasoning, laboratory work
Keywords (user):	Inquiry, Record keeping, Metacognition

Elementary students' laboratory record keeping during scientific inquiry

Abstract

The present study examines the mutual interaction between students' writing and scientific reasoning among 6th grade students (age 11-12 years) engaged in scientific inquiry. The experimental task was designed to promote spontaneous record keeping compared to previous task designs by increasing the saliency of task requirements, with the design goal of making the relationship between record keeping and inquiry strategies more explicit and visible. Compared to previous studies, this new task design resulted both in a higher amount of record keeping overall and in a higher quality of information, which is interpreted to be a result of increased participants' metatask and metastrategic knowledge arising from greater engagement with the task. The study found a significant relationship between the quality of students' record keeping and the inquiry strategies that were investigated. However, this relationship varied depending on the type of inquiry strategy. Strategies that are employed during the design of experiments [i.e., factorial combination and control of variables (CVS)] were statistically related to the number of complete comments (plans and intents), but not with the total number of comments. In contrast, the study found that for strategies employed while evaluating evidence (i.e., drawing inferences), student production of quality records is a necessary but not sufficient condition for effective evidence evaluation; in addition to recording high-quality information, students must also review their records (both from design and evaluation phases).

Science and Writing

The claim that writing in the classroom can foster learning across the curriculum has been a long-standing topic of research in education and psychology. However, the previous research has not clearly defined the conditions in which writing promotes learning. In their meta-analysis of the literature, Bangert-Drowns, Hurley, and Wilkinson (2004) conclude that writing has a small positive impact on academic achievement, but the benefits of classroom interventions that incorporate 'writing to learn' are dependent on the context and strategies used. For example, interventions that include metacognitive strategies such as prompts to reflect on their ongoing learning on a content area and comprehension failures and successes are more likely to result in enhanced learning, whereas interventions that include longer writing assignments are less likely to be beneficial.

There is similar interest in the use of writing to support learning in science education, on the part of both researchers (Hand, Prain, Lawrence, & Yore, 1999; Keys, 1994, 1999a, 1999b; Keys, Hand, Prain, & Collins, 1999; Klein, 2000; 2004; 2006; Prain, 2006; Prain & Hand, 1996; Prain & Waldrip, 2010; Rivard & Straw, 2000) and practitioners (Klentschy, 2008; Tierney & Dorroh, 2004). The science education research literature is inconsistent concerning the benefits of writing for science learning, even more so than the research literature is with regard to the benefits of writing for learning across the curriculum. In his review of the use of writing in secondary school science, Prain (2006) points out that even though there is widespread agreement on the benefits of *talking* about science to foster learning, there is no consensus concerning how and why *writing* benefits science learning. In his own review, Klein (2000) highlights the need for a theory to explain the potential of writing for learning in science.

The studies mentioned above view writing in science classrooms as a means of enabling understanding of scientific concepts, inquiry processes, and practices among secondary and college students. However, there is not yet clear evidence as to whether this benefit applies equally to novice writers, such as elementary school students (Bereiter & Scardamalia, 1987; see Klein, 2000, for a review). Most of the previous research across all ages has focused on the writing of narrative prose. While the preparation of scientific reports is an important part of the research process, the present study examines a form of writing that occurs during the scientific investigation itself: the recording of laboratory notes.

The laboratory notebook is considered a thinking tool for the students where language, data, and experience operate jointly to form meaning for the student, thus where students can apply language arts not only to develop a deep understanding of science content but also to attain scientific literacy (Amaral, Garrison, & Klentschy, 2002; Klentschy, 2008; Klentschy & Molina-De La Torre, 2004; Rivard & Straw, 2000; Shepardson & Britsch, 2001; Saul, Reardon, Pearce, Dieckman, & Neutze, 2002). Under the assumption that laboratory notebooks become a thinking tool for the students, Amaral et al. (2002) go a little further and claim that students should be provided with the opportunity to write to themselves in their laboratory notebooks. This group of researchers and practitioners claim that the students' laboratory notebooks should be embedded into the science curriculum. They maintain that the student laboratory notebook is not a mere record of data that students collect, facts they learn, and procedures they conduct but a record of students' reflections, questions, predictions, claims linked to evidence, and conclusions, all structured (Klentschy, 2008). And although these embedded activities could start as early as kindergarten, students need time and practice using laboratory notebooks to attain expertise. In order to help students to learn how to write in their laboratory notebooks, embedded writing prompts such as questioning, predicting, clarifying to promote comprehension monitoring and summarizing become necessary.

According to this view of science learning integrated with note taking activities in laboratory notebooks, we think that there is a prior condition that should be satisfied before a

Deleted: science note
Deleted:
Deleted: science note
Deleted: ese
Deleted: science note
Deleted: state
Deleted: science note
Deleted:
Deleted: science note
Deleted: science note
Deleted: science note

student is asked to use a notebook during scientific inquiry. In order for students to benefit from these advantages and avoid writing in their notebooks to simply fulfil the teacher's demands, they need to be aware of the utility and the benefits of note taking while doing scientific inquiry. When elementary school students are asked to solve a scientific problem and are not specifically asked to take notes, do they do so? Are they aware of the benefits of note taking or of what notes to take? And also, in what ways is this note taking related to their inquiry strategies? These are the questions we address in the present paper. More concretely, the purpose of this study is to investigate the relationship between elementary students' inquiry strategies and their laboratory record-keeping practices, with a special focus on the children's awareness of the benefits of these practices.

Deleted: that
Deleted: they
Deleted: e
Deleted: l

The Use of Inscriptions in Science

As a theoretical starting point, we begin with Latour's claim (1990) about the use of 'inscriptions' in empirical research, as well as the related studies that use Latour's ideas to show how young students may benefit from inscriptional practices (e.g. Lehrer, Schauble, Carpenter, & Penner, 2000). In this line of research, 'inscription' can refer to geometrical representations, maps, diagrams, graphs, tables, texts, and chemical, algebraic, or numerical notations that are used to represent the world and freeze those aspects that are essential to build theories (Latour, 1990). According to Lehrer and Schauble (2006) and others, these external representations are not mere copies of what one sees, but rather are the products of adapting, selecting, magnifying, and fixing the conventions of representational systems to build arguments.

Deleted: several authors (e.g.
Deleted: &
Deleted: ,
Deleted: add that
Deleted: ,

Along with historical work examining scientists' laboratory notebooks and the advancement of science, the recent theoretical interest in the relationship between inscriptions and cognition (Olson, 1994; Wells, 1999) leads us to highlight three important functions beyond the communicative. The first is the *mnemonic* function, as established in Tweney's study on Faraday's rigorous laboratory record keeping 'to prevent any change on what is remembered' (Tweney, 1991, p. 305). The second is the *organizational* function, which enables the management, organization, and structuring of the information involved in the empirical research, with the goal of making the information objective and facilitating awareness of certain relations that would otherwise be invisible (Wells, 1999). Finally, the third is the *epistemic* function. Science advances by creating, manipulating, and transforming inscriptions as semiotic objects that create meaning (Lemke, 2002). The epistemic function is very well illustrated in Gruber's analysis (1974) of Darwin's notes and Holmes' report (1987) on Krebs' findings in biochemistry. These two analyses show the mutual adaptation that occurs between the internal and external representations through the revision of notes while the research work is in progress. More concretely, Darwin's successive draft diagrams illustrate very clearly the progress in the search for the missing link between primates and humans in his theory of evolution. Based on historical studies that demonstrate how inscriptions contributed to the advancement of science, we join Klein and Olson (2001) as they pose the question of whether inscriptions maintains the same effect, moment-by-moment, on the development of scientific thinking in elementary school students. If so, how?

Deleted: i

Recent years have seen a growing interest in the analysis of the relationship between inscriptions and students' conceptual development on the one hand (Keys, 1994, 1999a, 1999b; Klein, 2000, 2004; Lehrer, Schauble, Carpenter, & Penner, 2000) and between inscriptions and the development of scientific reasoning on the other (Eberbach & Crowley, 2009; Ford, 2005; XXX, 2007; Kanari & Millar, 2004; Klaczynski, 2000; Masnick & Klahr, 2003; Wu & Krajcik, 2006).

Inscriptional Practices and Concept Development

Research that relates conceptual development and inscribing shows contrasting results. A Lehrer et al. (2000) study of third graders used a task embedded in a year-long science and mathematics curriculum. Using Latour's (1990) expression 'cascade of inscriptions', the authors provide a qualitative analysis of the interaction and mutual progress of the children's concepts of plant growth and their representational practices. Keys (1994, 1999a, 1999b) also worked with students over several months in a naturalistic setting. She asked seventh through ninth grade students to make observations, gather and interpret data, and write a report. Using the Bereiter and Scardamalia (1987) model of writing and Halliday's (1978) linguistic framework, she analyzed the reports from both a content and a linguistic point of view, respectively. Her results (Keys, 1999a) showed that the students who integrated inferences and data into their reports were more the exception than the rule. Also, very few students were able to elaborate on their initial ideas by using language to generate new meaning from the investigation. Keys (1999b) maintains that these students approached the task of inquiry report writing by relating their investigative findings in a rote manner with little reflection on the meaning of the data. Similarly, but with much younger students, Ford (2005) showed how third graders recorded rigorous but irrelevant descriptions of minerals while making very few associations between their observations and the related concepts.

From a more quantitative perspective, Klein (2000) asked elementary school students (grades 4, 6, and 8) to conduct science experiments while taking journal-style notes. Klein focused on the effects of writing on concept learning. He found that the extent to which students reviewed what they had done and written, using their experiments and the text to develop knowledge about the tasks, seemed to be a crucial factor in explaining gains in knowledge. These contrasting results on the relationship between writing and concept development could be interpreted in terms of involvement and reflection on the task. In the study by Lehrer et al., students' involvement and reflection may have been fostered by the teachers' long-term and continuous scaffolding, which was absent in the other cases. Children and early adolescents may require prompts and scaffolds to remind them of the importance of record keeping for scientific discovery (Zimmerman, 2005).

Inscriptional Practices and Scientific Inquiry

The extended literature on the development of scientific inquiry strategies contrasts with the little attention that is devoted to children's awareness of the benefits of laboratory record keeping when investigating a scientific problem. To review the research that looks into the relationship between record keeping and the scientific inquiry strategies, we need to specify the strategies to which we refer. We view science knowledge acquisition as the ability to consciously articulate a theory, understand the type of evidence that supports or contradicts it, generate such evidence, and justify the confirmation or disconfirmation of such theory (Kuhn, 1989; Kuhn, Garcia-Mila, Zohar and Andersen, 1995). This approach is well illustrated in Duschl, Schweingruber, & Shouse's (2007) definition of scientific investigation, envisioned as something that involves numerous procedural and conceptual activities such as:

> Asking questions, hypothesizing, designing experiments, making predictions, using apparatus, observing, measuring, being concerned with accuracy, precision, and error, recording and interpreting data, consulting data records, evaluating evidence, verification, reacting to contradictions or anomalous data, presenting and assessing arguments, constructing explanations (to oneself and others), constructing various representations of the data (graphs, maps, three-dimensional models), coordinating theory and evidence, performing statistical calculations, making inferences, and formulating and revising theories or models (p.130).

Deleted: asking

According to the citation by Duschl et al. (2007), our claim is that it is important to examine the entire process of scientific investigation when studying the development of scientific inquiry strategies. This is due to the interrelationships between the parts of the investigative process. For example, even if the generation of data is done via an experimental design in which a variable is isolated, those data will not be effectively used unless inferences are drawn using a valid strategy that considers that a controlled comparison is being made. Similarly, general conclusions can only be made when all possible combination of variables are tested, that is, only when conclusions refer to inferences based on the complete problem space. Also when studying the relationship between the inquiry and inscriptional practices, these need to be examined across the entire process of scientific investigation. A chart that structures the factorial combinations of variables may be used as a tool for experimental design allowing the organization of the complete problem space[1] when constructing successive experiments that isolate and control variables as well as a tool for evidence evaluation, allowing for the controlled comparison of evidence across multiple experiments. More concretely, when elementary students engage in a self-directed inquiry task, what data do they generate? Do they use the factorial combination strategy? Do they cover the complete problem space· if given the chance? Do they design controlled experiments, that is, those based on the control-of-variables strategy? Or, even further, do they make inferences based on those controlled comparisons when they are asked to evaluate the evidence they have generated? Also, and most important to this study, do they record information and review it? A large number of studies have discussed the main biases that preadolescents, adolescents, and adults show when they are asked to solve an inquiry task (see Duschl et al., 2007; Schauble, 1990; Zimmerman, 2000, for reviews), but only a few have examined data recording during scientific inquiry in particular.

In their classic study, Siegler and Liebert (1975) investigated the effects of record keeping on the design of a factorial experiment. They used trained students from grades five through eight and asked them to draw tree diagrams. They then examined how these diagrams helped the students to investigate all possible combinations of variables. They found a significant correlation between record keeping and the number of combinations designed; those whose training was more focused on drawing tree diagrams were more likely than peers in other conditions to generate all combinations. Similarly, Toth (2000) analyzed note taking

[1] According to Klahr and Dunbar (1989), the problem space investigated is the total possible number of unique combinations of variables that would constitute the database from which inferences can be made.

in relation to the design of controlled comparisons. She found that preadolescents' strategies improved when they used predeveloped tables. However, when subjects developed their own inscriptions, the benefit disappeared. The author claims that the expressive use of inscriptions requires a minimal level of metatask knowledge (i.e., knowledge about the structure of the domain, the goal of the task, and the cognitive state of the interpreter). Again, it seems that with appropriate scaffolding, children benefit from inscriptions.

A different question is whether this benefit remains in more naturalistic tasks that do not include record keeping in their instructions (Tweney, 1991) or provide scaffolding for inscribing. For instance, Carey et al. (1989) showed that spontaneous record keeping was more the exception than the rule among seventh graders asked to determine which factor (yeast, flour, sugar, salt, or warm water) caused bubbling in a mixture. Similarly, Kanari and Millar (2004) had 10-, 12-, and 14-year-olds work on two causal reasoning tasks that involved the management of a covariation effect and a non-covariation effect, both believed to covariate. Although their study referred to general inquiry strategies, they looked into record keeping during inquiry and found that students rarely took notes for those results that confirmed their prior expectations, but repeated significantly more experiments and recorded more data points for the non-covariation variable, as if by repeating they would succeed in making the non-covariation data fit their prior theory.

In our own previous work (XXX, 2007), we presented four different inquiry tasks with the goal of determining the causal structure of the underlying multivariable system. This was a self-directed investigation in which the students were provided with a notebook to record anything they wanted. We aimed to find out whether the students would keep records, what records they would keep, and whether they would review those records given that design was microgenetic with the task lasting 10 weeks (split into two phases).

Deleted: ten

Children were asked to work on four different problems: the Boat task (a physical model) and the Cars task (a computerized simulation) in the physical domain, and the School task and the TV task (both paper-based) in the social domain. All were designed to be isomorphic in regard to the structure of the tasks' effects (causal, noncausal, and interactive). That is, all four tasks consisted of a causal system of five variables that presented the same arrangement of task effects.

Fifteen 10-year-old (fourth grade) students and 17 community college students worked individually in two 30-45 minute sessions per week, one in the physical and one in the social domain, for a total of 10 weeks. At the sixth of the 10 weeks, the alternate and counterbalanced tasks in each domain were replaced for the remainder of the study. All participants thus encountered all four tasks by the end of the study. In order to maintain a naturalistic setting, no specific instructions about note taking were provided. Rather, in the first session each participant was given a notebook with his/her name on it and told that it would be available in each session in case they needed it.

We found a lack of spontaneous record keeping among fourth graders when compared to adults. Only half of the children kept records compared to all but one of the adults. Also, on average, adults took three times as many records as the children, with the latter never taking a single complete note.[2] Most importantly, the children's recording decreased over time, dropping to about half when comparing the initial and final phases. The children were also observed to review their notes rarely.

The problem space of each task was considerably large (48 different combinations of variables) and the students moved from one session to another without making any connections or seeing any need to integrate results across sessions, even though there were five sessions for each task. This lack of continuity would explain why these children did not

[2] A complete note was defined as any note that referred to an experiment that contained all of the antecedents if it was an intent (before the experiment was done) and all antecedents and the outcome if it was an assertion (a record of an experiment already performed).

see any need for reflection or revision of the notes. Because data collection and interpretation were done within the same session, children might believe that the inquiry demands for each session occurred at the conclusion of the session.

A critical analysis of this study and the others previously mentioned in this section revealed that the tasks used did not seem to provide enough feedback regarding the need to take notes and what to take notes on. For these children, record keeping may not have seemed to be of any utility. This could be due to the fact that their knowledge about the task (metatask knowledge) was limited (Toth, 2000), especially in relation to their cognitive state (i.e., memory limitations). Children's metatask knowledge may be limited in the sense of lacking: (a) the need to address the complete problem space; and (b) the need to compare outcomes gathered not only within but also across several sessions. Both refer to elements of memory: The first is related to their working memory and the need to mentally organize all possible combinations of variables, while the second is related to their long-term memory and the need to remember all of the experiments and their outcomes. Toth (2000) concluded that children's note taking was not of sufficient quality. This is the same interpretation made in Eberbach and Crowley's (2009) review of observational skills, which indicated that this lack of spontaneous record keeping might be due to the fact that children's observational records typically include information that is incomplete.

The above results raise several questions. Would children take more notes if the task design induced them to, with the possibility that, in taking more notes, they would receive more feedback on the benefits of note taking? Would a design that implicitly induces participants to review their notes help to provide this feedback?

To address these questions, we modified two aspects of the previous study. First, the participants designed the experiments at a different session than the one in which they interpreted the data. The use of a biological system (plant growth) required children to wait until the following session to see if the planted seed had grown and how much it had grown. This aspect of the task served not only to encourage the writing of notes (for a clear future need) and the review of notes (to recall information from the past), but also to highlight the usefulness of looking across sessions at the continued growth of the plants rather than viewing each session as a self-terminating event. Second, the number of factors in the task was reduced because we did not want participants to get overwhelmed and distracted with a larger size of the problem space over the course of their investigation. In addition, the continued growth of the plants further encouraged participants to engage in analysis of the same factors over time rather than change their focus every session.

According to this critical analysis of our prior work, the goal of the present paper was to determine whether the quantity and quality of note taking would be affected by using a task that lasted several sessions instead of one that self-terminated in a single session. In addition, the variable to be observed was of a cumulative nature and depended on the results of prior sessions (with records taken repeatedly over the different sessions, emulating the work of scientists). This would increase awareness of the need for and utility of note taking and, as a result, the relationship between record keeping and scientific inquiry strategies would hopefully become visible and explicit. As previously mentioned, record keeping is an important process of scientific inquiry and one that is typically neglected in research. Access to data gathered in several sessions, the history of changes in the variable, and consultation of cumulative records become essential in scientific reasoning. We hypothesized that experimental design and/or inference-making strategies would be elicited from and enhanced by information recording. Notes may help to structure the factorial combinations of variables, thereby allowing for the organization of all possible combinations of variables when constructing successive experiments that isolate variables. In addition, notes may serve as tools for evidence evaluation by allowing for the controlled comparison of evidence across

Deleted: ,

multiple experiments. For these benefits to accrue, however, students must first realize the need to take notes. This leads to our main hypothesis: Changes in the task design will encourage students to increase their record keeping. Furthermore, the increase in record keeping will provide feedback for students to regulate the records' usefulness and, eventually, their quality.

Method

Participants

Participants belonged to an intact classroom of 34 sixth graders (17 girls and 17 boys) from a public charter school of a middle SES neighborhood in the city of Barcelona (Spain). Their mean age was 11.6 (range 11.0-13.0). All students participated individually in two 30-45 minute weekly sessions during a four-week period (for a total of seven sessions). This twice-weekly interview protocol was conducted as a within-subjects design (i.e., each participant serves as his/her own control as change in performance is analyzed). The inquiry task took place in a lab large enough to accommodate all of the participants' plants. Interviews were conducted by a native speaker in the students' first language (either Catalan or Spanish).

Deleted: u

Deleted: ;

Task

Formatted: Left

The task was presented as part of the science curriculum. Participants were told that they were going to participate in a four-week tutoring program to develop inquiry skills. They were asked to investigate which of three factors caused a given plant to grow faster (Wisconsin Fastplants, 1999).[3] The factors presented to the participants were type of light (artificial or natural), type of fertilizer (chemical or organic), and type of seed (Rosette or

[3] Fastplants are a species of a fast-developing cabbage (Brassica) that completes its life cycle in 14 days.

Brassica). The problem space that results from the combination of the variables and their respective effects are presented in Appendix A.

The scientific problem was introduced to the participants as follows:

> *The Canadian Government has discovered a seed called Brassica that is very effective to feed the cattle. Our local Government is very interested in testing that seed under different conditions. Also, the local Government has another seed (Rosette), very similar to Brassica, that might work as effectively as Brassica, but it is much cheaper. Your task during the following four weeks is to determine the best conditions for the plant to grow and also whether the Rosette could work as well as the Brassica[4]. There are three factors that we have been asked to study: the type of seed (Brassica or Rosette); the type of fertilizer (chemical or organic) and the type of light (natural or artificial).*

Formatted: Font: 10 pt

Formatted: Font: 10 pt

Design and Procedure

The task, as presented in the previous Section, lasted seven sessions organized in four weeks. The participants were told that in order to solve the scientific problem they would be allowed to design 10 experiments.[5] Since the plants take several days to begin showing the effects of the different factors in their growth, and also the fact that we aimed at capturing the effect of participants' own feedback in their inquiry process, we organized their work according to the following sequence: In session 1 the students only designed four experiments; in session 2 they observed these four experiments and designed four more experiments; in session 3 they observed the growth of all eight previously designed experiments. In session 4, participants were asked to observe the plants growth again and design the last two experiments. From session 5 onward, the children only observed, discarding four experiments in session 5 and four experiments in session 6. In session 7,

Deleted: plants

Deleted: ,

[4] These instructions are based on the engineering model of investigation (''produce the best outcome'') vs. the scientific model ("find out how the system works") (Schauble, Klopfer, & Raghavan, 1991).

[5] The term "plant" is used as equivalent to "experiment". Each plant is designed according to the three factors under inquiry, therefore it can be considered an experiment regardless of what other plant it is compared to.

participants observed the last two experiments, drew conclusions, and wrote a final report. Since the total number of possible combinations was eight (2x2x2: three variables of two levels each), participants could have completed the problem space by the end of session 2 (see Appendix A).

In the first and the last sessions, students' theories were assessed to test their content learning. After the initial theory assessment, students were invited to begin the investigation by choosing the levels of each of the three factors for the first plant. In each session, the interviewer asked the participants a range of questions ('What are you planning to find out with this experiment?' 'What do you think the outcome will be?' 'What have you found out?' 'How do you know that … is better than …?'). Similar to our prior study (XXX, 2007), no specific instructions about note taking were provided in order to maintain the naturalistic setting and to allow the analysis of students' spontaneous note taking, as well as their awareness of the utility of taking notes and what notes to take for scientific inquiry. Thus, in the first session, each participant was given a notebook with his/her name on it and told that it would be available each session in case they needed it. In addition, participants were provided with the necessary materials (pots, soil, fertilizers, seeds, and stickers on which to write their names and the date in order to identify their pots).

Deleted: i

Results

The results of the present paper focus on the analysis of the relationship between note taking and the strategies involved across the whole cycle of self-directed scientific inquiry. The strategies under investigation were structured into two groups: those involved in experimental design and those involved in evidence evaluation. For the former, we analyzed the factorial combination strategy and the strategy to design controlled comparisons to gather

Deleted: correlational

data. The second part of the results section will focus on the latter group–the strategy of making valid inferences based on evidence evaluation of the self-gathered data.

Record Keeping

Any inscription separated by blank spaces or lines was considered a record (XXX, 2007). Only records that self-referred to experimental activity were taken into account for this analysis. The occasional irrelevant comments (e.g., ‘my mom shows me how to take care of the plants and she will be happy when I tell her what to do’) were not included in the note taking analysis. These comments comprised fewer than 2% of the total entries. One child out of 34 took no notes at all. Most of the inscriptions made by the children[6] were text notes with different levels of structuring. Some were linear sentences with little or no structuring (see Figure 1), while others (7/34) progressed from linear text to text structured in charts and lists (single- or double-column lists, see Figures 2 and 3, or Figure 4).

Participants’ entries were coded as comments and assertions. Comments referred to intents and plans, which were considered complete if they included all the antecedents. Intents and plans were written between sessions 1 through 4, those sessions in which children were asked to design experiments. When the record explicitly referred to an observation of a specific experiment that could either include an inference or not, it was coded as an assertion. These records were coded as complete if they contained enough information to mentally replicate the experiment. That is, any record referring to an experiment that contained, at minimum, all of the antecedents, the outcome, and a reference to the time was coded as a complete assertion (see XXX, 2007). The total number of records was double-coded, and reliability was 89%. Disagreements in coding were resolved by discussion. Since participants gathered data in sessions 1 through 4, their comments and plans concentrated on these

[6] All names are pseudonyms.

sessions. They observed the data in sessions 3 through 7 and therefore mainly recorded their assertions in entries written at those times.

No record keeping training was provided. The goal of the study aimed at analyzing students' spontaneous note taking and with it, awareness of the utility of notes for scientific inquiry and of what notes are best to take. Thus, students' were expected to take complete notes that could be used in further encounters with the task. Figures 2 and 3 show an optimal recording in which plans and intents progress into assertions within the same entry. In the second session, David made a diagram to apply the factorial combination strategy and listed all the experiments generated by the diagram. He then structured his notebook into eight entries, one for each of the experiments he was planning to design. These entries were coded as plans. In the following sessions, he recorded the data (i.e., the plant growth) for each session in different colors, with each color corresponding to a different data set (see the color code in the upper right hand corner of the page, Figure 3 shows page 5 of his notebook). The recording strategy shows good awareness of the benefits of writing economical notes (Lee & Karmiloff-Smith, 1996) and the benefits of the history of the notes (Lehrer & Schauble, 2000). This recording strategy was seldom used by participants in the study (only 3 out of 34 used it) in spite of the fact that nonlinear note taking has proved to enhance learning (Makany, Kemp, and Dror (2009).

Deleted: and

Deleted: -

Record Keeping and Experimental Design Strategies

Factorial combination strategy. As noted previously, the problem space for the experimental task was eight. That is, children could design eight different factorial combinations by varying the two levels of seed factor (Brassica and Rosette), two levels of fertilizer factor (organic and chemical), and two levels of light factor (artificial and natural) (i.e., 2x2x2=8). Since children were told to design four experiments per session, they could design eight different combinations of factors and thus complete the problem space (PS) by session 2. They did not design any experiments in session 3, though they did design two more experiments in session 4. Fourteen children completed the problem space in session 4, while twenty children never completed it. The mean percentage of PS investigated by session 4 was 86 (SD=15.3).

Given the type of study design we constructed, we were unable to establish a cause-effect relationship between children's record keeping and their efficiency in investigating of the problem space. However, it is interesting to note that the children's recording between session 1 and session 4 was significantly related to their performance in applying the factorial combination strategy when designing experiments. We observed that the 14 (out of 34) students who designed the eight different combinations in the total 10 experiments (four in session 1, four in session 2, and two in session 4) had a significantly higher number of complete comments (intents and plans) (see Figure 2) when compared with those who did not design all eight combinations (20/34) (see Figure 1). The mean number of complete intents and plans for those who completed the problem space by session 4 was 11.3 (SD=8.2). The mean for those who did not complete it was 3.4 (SD=4.8) (see Table 1 for a summary of the results).The Mann-Whitney U non-parametric test for comparison of means yielded statistical significance (U=49.5, $p<.001$). It is worth mentioning that the same analysis performed on the total number of records (complete and incomplete) was not significant. Thus, it is the fact that the record is complete (rather than simply the fact that the note is taken) that is related to the factorial combination.

Insert Table 1 about here

Insert Figures 1, 2 and 3 about here

Was this efficiency in designing all possible combinations accompanied by an awareness of the need to design controlled comparisons to make valid inferences? Furthermore, was the strategy used to determine causality accompanied by inferences based on all of the controlled comparisons one could make from the complete database? How did participants perform as a whole in designing controlled comparisons?

Control-of-Variables strategy (CVS). Because the task setting of the study focused on the complete inquiry cycle, we can address not only the students' performance with regard to the factorial combination strategy, but also their awareness of designing controlled comparisons. We structured students' performance into three groups: those who did not show any signs of awareness of the need to design controlled comparisons, those who applied the strategy without any explicit awareness, and those who explicitly mentioned the need or the benefit of designing two controlled instances. We observed 21 students who did not show signs of CVS and 13 who did. Of these 13, six used strategies that were coded as implicit and seven used strategies coded as explicit. More concretely, a CVS was coded implicit if the participant designed a pair of experiments whose variables were controlled for except for the one whose effect was the intended to find out about, but when they were asked for the goal of that particular experiment, they did not mention relate the varying factor with the goal of the experiment. On the other hand, the strategy was coded as an explicit CVS when the participant mentioned that the factor under investigation was the only one that varied. The following examples transcribed from the students' verbal protocols (and translated from Spanish) show both implicit and explicit CVS application.

Deleted: the fact

Example of implicit CVS in which Peter designed two experiments, one with brassica, natural light, and organic fertilizer (BNO) vs. and the other with brassica, natural light, and chemical fertilizer (BNC):

-Exp. What do you plan to find out?

-Peter: *If it grows the same or worse. That is, if it does give any fruit. I don't know.*

To see the differences between putting chemical fertilizer or organic.

Example of explicit CVS in which Ruben (Figure 4) designed BNO vs. RNO:

-Exp. What do you plan to find out?

-Ruben: *This time I planted Rosette. This way I'll be able to test the difference between one seed and the other. I'll put them (the two seeds) in the same conditions and that way I'll be able to see if one with the same conditions grows more than the other.*

Like the analysis of note recording and the factorial combination strategy, we performed an analysis comparing the mean number of complete records (plans and intents) for those who demonstrated CVS in session 2 (pooling the implicit and explicit together) and those who did not. Since the number of complete records (session 1-2) was not normally distributed, we performed the Mann-Whitney U test once again and found that it yielded significant differences between means. Mean rank of complete records for those who did not use CVS was 14.02 (N=21; mean=2.02, SD=3.6). Mean rank for those who did use CVS was 23 (N=13; mean=6.38, SD=4.27) (U = 63.5, p= .005) (see Table 1). When the students took incomplete records about plans or intents, the number of records was not related to the quality of their experimental design strategies. Therefore, we again see that the strategies used to gather data (more concretely, the strategies used to design controlled comparisons) are related to record keeping strategy. This demonstrates that only the complete records were related to scientific inquiry.

As we have seen in both types of experimental design strategies (factorial combination and CVS), the statistical relationship between records and experimental design strategies was significant only when complete records (rather than all records) were considered in the analysis. The next section addresses the issue of whether this finding also applies to the relationship between record-keeping and inference-making strategies.

Deleted:

Record-Keeping and Inference-Making Strategies

Deleted:

The data analyzed in this section will concentrate on sessions 3 through 6. Session 2 was excluded because many participants were unable to observe plant growth. By session 3, however, all of the children had some kind of plant growth from which to make inferences. Session 7 was also excluded because the students' inferences were mixed with general conclusions elicited by the experimenter.

Children could make two types of inferences that were coded with two levels of strictness. According to the strictest criterion, the participant needed to draw an inference of inclusion if the outcome varied or an inference of exclusion if the outcome did not vary (based on two experiments whose factors were identical except for the one about which the inference is made). We call this type of inference a valid inference based on the control-of-variables strategy (Chen & Klahr, 1999; Kuhn et al., 1995; Tschirgi, 1980). For example, if we wanted to test whether the type of light was causal, we could design the following two experiments and observe the plants' growth rates:

Brassica, artificial light, & organic fertilizer=12 cm in 2 weeks

Brassica, natural light, & organic fertilizer= 5 cm in 2 weeks

A valid inference would then be that the type of light is causal, with artificial light being better than natural light.

The number of inferences based strictly on controlled comparisons was very low. Between sessions 3 and 6, two children made four inferences, one child made three inferences, three children made two inferences, seven children made one inference, and twenty-one children made no inferences. Because the biological domain gathers probabilistic data, the fact that the task belongs to the biological domain could be interpreted as playing a role against our students' willingness or need to make inferences based on CVS. As shown from the data gathered by the children, our experimental setting was highly susceptible to uncontrolled variables. It was clear that the children were aware of this susceptibility. This could be the reason for the children's low use of the control-of-variables strategy (used by less than 10%). In contrast, we observed a high number of inferences of the generalized type. This type of inference is not based on comparison of any specific instances, but instead generally refers to an entire database of (uncontrolled) instances. Children may focus their attention on one variable, and one of its levels may be perceived as being associated with a different outcome or range of outcomes than the other level (Kuhn et al., 1995). For example, after observing the database formed by all (uncontrolled) instances involving Brassica and Rosette, the mean height of Brassica was clearly higher than the mean height of Rosette. Participants could have then concluded that Brassica is better than Rosette, although the Rosette with artificial light may be higher than Brassica with natural light. Although we are aware that these inferences are not valid according to the deterministic sciences, we applied this less restrictive criterion to code valid inferences, and we considered these generalized inferences as valid. We considered them superior to the inferences coded as clearly invalid, such as those based on theory, those that were non-justified, or those that were invalidly justified (e.g., inferences based on a single instance or on several instances not involving any comparison). Again, all verbal protocols were double-coded and reliability reached 85%, with disagreements resolved by discussion.

Since our goal was to study progress in making valid inferences and how this progress could be related to writing, we compared the proportion of valid inferences (CVS and generalized) from sessions 3 and 4 to those from sessions 5 and 6. In sessions 3 and 4, participants observed clear growth from a total of 16 experiments (8 in each session). In addition, they observed 10 experiments in session 5 and 6 experiments in session 6 for a total of 16. The mean proportion of valid inferences in sessions 3 and 4 was compared to the mean proportion of valid inferences in sessions 5 and 6. The means were 0.38 (SD=.40) and 0.54 (SD=.37), respectively. The statistical comparison between means yielded significant differences [Wilcoxon signed ranks test (N=34) =-1.8, p (one-tailed)=.035, effect size, *d*=.40].

Therefore, we observed slight progress with practice in strategies of making valid inferences. This preadolescents' level of performance is similar to that reported in the developmental literature on scientific reasoning skills (see Zimmerman, 2000 for a review). Likewise, this performance can be related to the students' theory change from the initial session to the final one. Table 2 shows a summary comparing the students' initial and final theories about the plants causal system. We find that in general students reached a good level of correctness of their theories, mostly because they confirmed their prior theories. However, it is also interesting to note the high proportion of students (26/33) that disconfirmed their prior theory about natural light better than artificial light. The statistical comparison between the correctness of final theories and initial ones was statistically significantly (Wilcoxon Signed Rank, $Z = -4.5$, $p < .001$). This shows a relevant content learning outcome that must be interpreted along with the progress in the inference-making strategies (see Table 2). The next issue is the check how much this improvement can be related to note recording.

Deleted: of it due to the fact that

Insert Table 2 about here

In terms of record keeping, the analysis focused on assertion notes that involved inference making. The mean number of participants' total notes-assertions was 13.7 (range 2 to 38), and the mean percentage of complete notes-assertions was 55% (see Figures 3 and 4 for example). Neither the mean number of notes nor the mean number of complete notes was significantly correlated with the mean proportion of valid inferences.

Deleted: -

Thus, we proceeded to test whether the students' use of their notes rather than note taking per se was related to scientific reasoning. Some students would take notes and never show any explicit sign of reviewing them by mentioning it in the individual interview, while others would verbally express the need to go back and check their previous notes. With this goal in mind, a new variable was defined to measure whether the students reviewed their notes. Participants were coded as note-reviewers if they explicitly mentioned the need for and action of reviewing their notes to design a new experiment or made claims in the oral interview when asked for their findings (e.g., 'let me check my notebook because I do not remember' or 'I need to see what happened in my previous experiments to see if artificial light is better…, let me see what I wrote'). Participants were also coded as note-reviewers when their notes showed some kind of data organization and structuring (e.g., when they added observations of different sessions under the same experiment heading, as in Figure 4), assuming that adding the day 3 observation under the day 4 observation with the new date implied minimal revision and comparison of prior results. A comparison of the means of the total number of valid inferences (from sessions 3 through 6) yielded significant differences between the reviewers (N=11) and the non-reviewers (N=22)[7]. The mean number of valid inferences was 4.6 (SD=2.9) for the former and 2.7 (SD=2.3) for the latter. A one-way ANOVA used to compare the means of the two groups yielded statistical significance, $F(1,32)=3.99$, $p= .05$.

Deleted: ¶

Insert Table 3 about here

Deleted: Because the results obtained here are also correlational

Formatted: Indent: First line: 28.35 pt

Due to the fact that the design was not experimental, we cannot establish a cause-and-effect relationship between reviewing one's own notes and making valid inferences. However, we have obtained a complete picture of the relationship between note recording during scientific inquiry and the strategies involved in this inquiry process (experimental design and inference making). As for the former, we observed that notes had to be complete in order for them to be related to the two core strategies of experimental design (factorial combination and control-of-variables strategy). On the other hand, to find a relationship between note recording and inference making, participants had to take notes as well as review them. The critical factor was not the number of notes or their completeness, but rather the fact

Deleted: -

Deleted: -

[7]One child did not make any entries.

that they were being reviewed. These results confirm Klein's (2000) finding about the importance of note revision in science learning.

Discussion

We begin the discussion section by addressing Klein and Olson's (2001) question of whether inscriptions have the same moment-by-moment effect on elementary school students' development of scientific thinking as it has historically had among scientists. More specifically, is there any relationship between elementary students' record keeping and scientific inquiry? As noted in the review of the literature, some studies have found a relationship between writing and scientific learning when specific instructions and appropriate scaffolding are provided (Lehrer & Schauble, 2000, 2006; Wu & Krajcick, 2006). However, this relationship has been difficult to demonstrate in studies with task instructions that do not prescribe writing. Elementary students rarely take notes, and when they do, the notes are incomplete and inaccurate. Consequently, the notes cannot adequately fulfill the notational functions mentioned above: mnemonic, organizational, and epistemological (XXX, 2007).

Deleted: ;

To demonstrate the relationship between elementary students' record keeping and scientific inquiry strategies, this paper used a task that satisfied several criteria hypothesized to foster spontaneous note taking and thus make the mentioned relationship between record keeping and inquiry strategies explicit and visible. The criteria were as follows: (1) the task had to be a self-directed inquiry that involves the entire cycle of investigation (hypothesizing, data gathering, data assessment, inference making, and drawing conclusions); (2) the instructions did not make record keeping mandatory; (3) the design was microgenetic and lasted multiple sessions, thereby providing practice and engagement with the task in order to increase metacognition; and (4) the topic of the task (in this case, plant growth) was chosen so that the effects of a given variable could not be observed on the same day that the experiments

Deleted: -

were designed. The plant required a few days before any growth could be observed, which we expected to allow participants to connect observations and emphasize the history of cumulative change in the variable being observed (Lehrer & Schauble, 2000); in doing so, we intended to make the need for and benefit of taking notes more explicit.

Our results can be summarized by three general findings. First, the way that the task was set up succeeded in eliciting spontaneous record keeping among the elementary school students. These results are in contrast with other findings in the field (Carey et al., 1989; Everback & Crowley, 2009; Duschl et al., 2007; XXX, 2007; Kanari & Millar, 2004) that show that without appropriate scaffolding, young students take spontaneous notes only occasionally while conducting scientific investigations. We observed an increase in overall record keeping compared to other studies, as is evident from the following three trends: (1) the fact that only one student did not take any notes at all; (2) the mean number of total notes was much higher than in other studies; and (3) there was an increase in the number of notes taken by the students in the final sessions as compared to the first one. The latter finding is not consistent with our own previous study (XXX, 2007) in which half of the children did not take any notes at all and those that did took far fewer (considering that the children worked over 20 sessions compared to 7 in the present study). Most importantly, the students in our 2007 study reduced their note taking by half during the 20-session inquiry process, while the present study showed an increase in note taking. Our main claim in interpreting these results is that changes in the task succeeded in eliciting the students' awareness of the utility of notes. This pattern of increase in the number of notes is arguably related to the fact that the notes in this study were of higher quality (i.e. more complete) and thus provided more empirical satisfaction when looking for the necessary information needed in the notebook. This experience served to foster better recording, which is related to the second finding we wish to highlight. Children in our 2007 study did not record any complete notes. In the present study,

Deleted: Garcia-Mila & Andersen

Deleted: Kanary

however, more than half of the students' notes were complete (on average). This can be interpreted as an indication of higher engagement with the task, increasing the metatask and metastrategic knowledge. Regular feedback was provided through practice to distinguish between what was useful and what was not. Therefore, our results thus far show notes of improved quantity and quality. In prior studies, elementary and junior high school students (Carey et al., 1989; Everback & Crowley, 2009; Duschl et al., 2007; XXX, 2007; Kanari & Millar, 2004) do not regularly and/or spontaneously take notes when they are presented a scientific problem. It is as if they perceive neither the need nor the benefits of note taking in their problem-solving process. Along with Toth (2000), our claim is that the use of inscriptions requires metatask knowledge. That is, a deep comprehension of the task demands such as knowledge about the structure of the domain, the goal of the task, and the knowledge about what one will need to know in repeated encounters with the task. The studies mentioned above are either presented as a scientific problem self-terminated in a single session, or when they are not, they may be wrongly understood by the students as such. It is in this sense that microgenetic designs, by providing repeated encounters with the task in a short period of time, increase participants' self-regulation facilitated by the feedback generated by the task itself (Kuhn, 2002). That is, when a student like David, in session 2 is asked what he wants to investigate, he realizes that in order to design all possible experiments generated by the factorial combination, he needs to rely not only on written records, but also on a diagram that solves the combination of variables. He shows a good awareness of the cognitive demands of the task and of the appropriate tools to solve it. This awareness comes from the dissatisfaction generated in his second encounter with the task where he realised that his Session 1 notes were incomplete to fulfil the goal of the task (his notes simply included four numbers under the heading of the date and the term 'seed' near each number, and a general prediction for the first one: *It will have grown*).

Deleted: -

Deleted: -

Deleted: do

Deleted: neither

Deleted:

Deleted: -

Deleted: about

Deleted: "

Deleted: "

The third and primary question we sought to address was whether these notes are related to the students' emerging inquiry strategies. Our results showed a statistical relationship between the various inquiry strategies investigated and the students' record keeping. However, this relationship varied depending on the type of strategy analyzed. On the one hand, the experimental design strategies (factorial combination and controlled comparisons) were statistically related to the number of complete comments (plans and intents), but not to the total number of comments. More concretely, students who completed the problem space (by designing all eight different combinations of variables) had a significantly higher number of complete comments. The same analysis was performed on the total number of notes, yielding a non-significant result. As for the relationship between record keeping and the use of the control-of-variables strategy, those who used CVS had a significantly higher number of complete notes than those who did not. In addition, when the number of total notes was pooled in the analysis, the difference remained non-significant. The two main strategies of experimental design (factorial combination of variables and control-of-variables strategy) were related to good record keeping. Our claim is that the previously reported increase in record keeping made the relationship between inscriptions and scientific inquiry strategies visible, as it would have been impossible to see otherwise. As we mentioned above, the fact that the design primed the naturalistic approach made it difficult to establish cause-and-effect relationships.

Deleted: our analysis is correlational. Consequently, we cannot establish a cause-and-effect relationship.

Deleted: T

Finally, the last analysis aimed to check whether there was a relationship between record keeping and evidence evaluation. We related the proportion of valid inferences to the number of complete note-assertions as well as the total number of note-assertions. Unexpectedly, neither pairing yielded a significant correlation. Moreover, when we split the sample of participants into the note reviewers and the non-reviewers, we found that those who reviewed their notes had a significantly higher number of valid inferences. This confirms

Klein's finding that one of the factors that contributes significantly to science learning is note revision. The results of the present study show that notes must be complete in order for them to be related to experimental design strategies. For notes to be related to inference making, however, being complete is not enough; the notes must also be reviewed.

Our main claim is that the task presented two characteristics that arguably increased students' awareness of the necessity and benefits of inscriptional practices during scientific inquiry. The first characteristic refers to the fact that observation of the effects of the task factors on the plant growth was delayed with respect to the design (i.e., the effect could not be observed on the same day that the experiment was designed). This delay emphasized the need to take notes on the studies that already been designed, the results that were obtained, and the need to review the notes for future sessions. Also, the fact that the growing cycle lasted two weeks forced the students to gather data in an iterative manner rather than in a single session. Likewise, their conclusions had to be based on cumulative data. Hopefully, these data were correctly recorded over the different sessions. Iteration in task sequencing was proposed by Wu and Krajcik (2006) in their study with tables and graphs as mediating factors in children's investigations.

On the other hand, the idea of data gathering and deferring observation over several days is an issue pointed out by Lehrer et al. (2000) under the concept of *history*. These authors underline the importance of the history of inscriptions for research processes and learning in science classrooms in a double sense. First, they refer to the *history* of the inscription itself. The fact that the inscriptions kept evolving and adapting to the task and were reviewed, edited, restructured, and redimensionalized made them candidates in the children's inscriptional repertoire. The second sense refers to *history* as something that is preserved. It is not only useful to recover what has been recorded when needed, as Faraday describes (Tweney, 1991); it is also useful to trace all changes in the inscriptional process. The present task was designed to fulfill both senses of the concept of history.

To summarize, the task included both of the above characteristics (iteration and history) to promote record keeping during inquiry. Thus, we can say that it succeeded in eliciting students' awareness of the necessity and benefits of note taking. First, our participants may have become aware of the need to consult the data from different sessions to make inferences. If they did not take notes, they had to tax both their working memory (in an attempt to coordinate data during controlled comparisons) and long-term memory (in trying to recall the results of past sessions) to know what they had done and what results they had obtained. A good understanding of the demands of the task (metatask knowledge) and the reasoning involved (metastrategic knowledge) were needed to avoid the lack of record keeping. These two metacomponents provide an awareness of the utility of producing external representations to serve as a tool to bridge the gap between the students' mental limitations and the task demands. In the present study, the note-reviewing process provided positive feedback on how students' inquiries could benefit from their notes. This produced an effect that fostered the quantity and quality of note taking. According to Lehrer and Schauble (2000), this finding highlights the importance of recognizing and comprehending the function of the inscriptions, rather than having a great repertoire of graphical tools.

Deleted: -

Deleted: -

Deleted: -

Deleted:

The increased awareness of the necessity and benefits of note taking led to better use of representational practices and to better inquiry practices. In fact, our results indicate that children showed how experimental design and evidence evaluation strategies were related to the quality of their notes and to the fact that they reviewed those notes, thereby supporting results reported by other researchers (Klein, 2000; Siegler & Liebert, 1975; Toth, 2000). The relationship between making valid inferences and note reviewing was interpreted by the fact that the importance of evidence was highlighted in the task structure. By having a task outcome that had to be observed repeatedly over different sessions over time, a need to record results was created. Also, the delay between the antecedent and the outcome was hypothesized to increase both the child's expectations and his/her focus on the antecedent in relation to the expected outcome. The students' representational practices played a role in their improved inquiry strategies by making evidence more explicit and making the coordination of theory and evidence more feasible. By recording observations, the evidence is explicited and more easily becomes an object of cognition that can be compared across records (Olson, 1994; Wells, 1999). This comparison may also generate the need to organize and structure the data recording in diagrams or charts that facilitate comparison (Lemke, 2002). All of these activities were embedded in a design for which metacognition was argued to be the key of the interrelated development.

To design classroom activities for scientific practice and science learning, these must be embedded in regular classroom learning activities. Along with theoretical concept learning, experimental design, observations, and inference making should be included in tasks that are done regularly. These activities should last several sessions instead of self-terminating in a single session and should include demands that combine all phases of the inquiry cycle. The consequent revision of data and notes would enhance the need for writing, note taking, and/or diagram making in support of the inquiry. Metacognition is crucial in the knowledge acquisition process, and writing can be a tool used to foster it. Wu and Kracick's (2006, p. 90) note that 'engaging students in using inscriptions in an iterated matter seems to promote the enactment of inscriptional practices'. We would add a comment on iteration, not only for inscriptional practices, but combined with inquiry practices to develop their mutual interaction and promote scientific reasoning and learning through the development of metacognition. The importance of iteration is also pointed out by Newton (2000), who claims that in order for students to benefit from data logging, their attention must be shifted back and

Deleted: -

Deleted: -

forth toward interpretative work that encourages them to focus on data and data-logging. Such activities will give them responsibility for decision making and will make them aware of their roles in each task.

Deleted: -

References

Amaral, O., Garrison, L., and Klentschy, M. (2002). Helping English learners increase achievement through inquiry-based science instruction. *Bilingual Research Journal*, *26*(2), 213-239.

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writing-to-learn interventions on academic achievement: A meta-analysis. *Review of Educational Research, 74*, 29–58.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition.* Hillsdale, NJ: Lawrence Erlbaum.

Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). An experiment is when you try it and see if it works: A study of grade 7 students' understanding of the construction of scientific knowledge. *International Journal of Science Education, 11*, 514-529.

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variable strategy. *Child Development, 70,* 1098-1120.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (Eds.). (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: The National Academies Press.

Eberbach, C., & Crowley, K. (2009). From everyday to scientific observation: How children learn to observe in the biologist's world. *Review of Educational Research, 79,* 39-68.

Ford, D. (2005). The challenges of observing geologically: Third graders' descriptions of rock and mineral properties. *Science Education, 89,* 276-295.

XXX (2007).

Gruber, H. (1974). *Darwin on man: A psychological study of scientific creativity.* Chicago: University of Chicago Press.

Halliday, M. (1978). *Language as social semiotic: The social interpretation of language and meaning.* London: Edward Arnold.

Hand, B., Prain, V., Lawrence, C., & Yore, L. (1999). A writing in science framework designed to enhance science literacy. *International Journal of Science Education, 21,* 1021–1035.

Holmes, F. L. (1987). Scientific writing and scientific discovery. *Isis, 78,* 220-235.

Kanari, Z., & Millar, R. (2004). Reasoning from data: how students collect and interpret data in scientific investigations. *Journal of Research in Science Teaching, 41*, 748–769.

Keys, C. W. (1994). The development of scientific reasoning skills in conjunction with collaborative writing assignments: An interpretive study of six ninth-graders. *Journal of Research in Science Teaching, 31,* 1003-1022.

Keys, C. (1999a). Language as an indicator of meaning generation: An analysis of middle school students' written discourse about scientific investigations. *Journal of Research in Science Teaching, 36,* 1044-1061.

Keys, C. (1999b). Revitalizing instruction in scientific genres: Connecting knowledge production with writing to learn in science. *Science Education, 83,* 115-130.

Keys, C. W., Hand, B., Prain, V., & Collins, S. (1999). Using the science writing heuristic as a tool for learning from laboratory investigations in secondary school. *Journal of Research in Science Teaching, 36*, 1065-1084.

Klaczynski, P. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: A two process approach to adolescent cognition. *Child Development, 71*, 1347-1366.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12,* 1-48.

Klein, P. D. (2000). Elementary students' strategies for writing-to-learn in science. *Cognition and Instruction, 18*, 317-348.

Klein, P. D. (2004). Constructing scientific explanations through writing. *Instructional Science, 32*, 191-231.

Klein, P. D. (2006). The challenges of scientific literacy: From the viewpoint of second-generation cognitive science. *International Journal of Science Education, 28,*143-178.

Klein, P. D., & Olson, D. R. (2001). Texts, technology and thinking: Lessons from the great divide. *Language Arts, 78,* 227-237.

Klentschy, M. P. (2008). *Using science notebooks in elementary classrooms.* Arlington, VA: NSTA Press.

Klentschy, M., & Molina-De La Torre, E. (2004). Students' science notebooks and the inquiry process. In W. Saul (Ed.), *Crossing boarders in literacy and science instruction: Perspectives on theory and practice* (pp. 340–354). Newark, DE: International Reading Association.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96,* 674-689.

Kuhn, D. (2002). A multi-component system that constructs knowledge: Insights from microgenetic study. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 109-131). Cambridge, UK: Cambridge University Press.

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Monograph of the Society for Research in Child Development, 60*(4), Serial No. 245.

Deleted: E.

Latour, B. (1990). Drawing things together. In M. Lynch & S. Woolgar (Eds.), *Representation in scientific practice* (pp. 19-68). Cambridge, MA: MIT Press.

Lee, K., & Karmiloff-Smith, A. (1996). The development of cognitive constraints on notations. *Archives de Psychologie, 64,* 3-26.

Lehrer, R., & Schauble, L. (2000). Modeling in mathematics and science. In R. Glaser (Ed.), *Advances in instructional psychology. Educational design and cognitive science.* Mahwah, NJ: Lawrence Erlbaum.

Lehrer, R., & Schauble, L. (2006). Scientific thinking and science literacy. In W. Damon, R. Lerner, K. A. Renninger, & I. E. Sigel (Eds.), *Handbook of child psychology: Vol. 4. Child psychology in practice*. (6th ed., pp. 153-196). Hoboken, NJ: Wiley.

Lehrer, R., Schauble, L., Carpenter, S., & Penner, D. (2000). The interrelated development of inscriptions and conceptual understanding. In P. Cobb, E. Yackel, & K. McClain (Eds.), *Symbolizing and communicating in mathematics classrooms. Perspectives on discourse, tools and instructional design* (pp. 325-360). Mahwah, NJ: Lawrence Erlbaum.

Lemke, J. (2002). Enseñar todos los lenguajes de la ciencia: Palabras, símbolos, imágenes y acciones. In M. Benlloch (Ed.), *La educación en ciencias. Ideas para mejorar su práctica* (pp. 159-186). Barcelona: Paidós.

Makany, T., Kemp, J., Dror, I. E. (2009). Optimising the use of note-taking as an external cognitive aid for increasing learning. *British Journal of Educational Technology, 40*(4), 619-635.

Masnick, A. M., & Klahr, D. (2003). Error matters: An initial exploration of elementary school children's understanding of experimental error. *Journal of Cognition and Development, 4*, 67–98.

Newton, L. R. (2000). Data-logging in practical science: Research and reality. *International Journal of Science Education, 22,* 1247-1259.

Deleted:

Formatted: Font: Not Italic

Olson, R. D. (1994). *The world on paper*. Cambridge, UK: Cambridge University Press.

Prain, V. (2006). Learning from writing in secondary science: Some theoretical and practical implications. *International Journal of Science Education, 28*, 179-201.

Prain, V., & Hand, B. (1996). Writing and learning in secondary science: Rethinking practices. *Teaching and Teacher Education, 12,* 609–626.

Prain, V., & Waldrip, B. (Eds.). (2010). Representing science literacies [Special issue]. *Research in Science Education, 40*(1).

Rivard, L., & Straw, S. (2000). The effect of talk and writing on learning science. *Science Education, 84,* 566–593.

Saul, W., Reardon, J., Pearce, C., Dieckman, D., & Neutze, D. 2002. *Science workshop: Reading, writing, and thinking like a scientist* (2nd ed.). Portsmouth, NH: Heinemann.

Schauble, L. (1990). Belief revision in children: the role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology, 49*, 31-57.

Schauble, L., Klopfer, L. E., & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching, 28*, 859-882.

Shepardson, D., & S. Britsch. (2001). The role of children's journals in elementary school science activities. *Journal of Research in Science Teaching* 38 (1): 43–69.

Siegler, R. S., & Liebert, D. E. (1975). Acquisition of formal scientific reasoning by 10- and 13-year-olds: Designing a factorial experiment. *Developmental Psychology, 11*, 401-402.

Tierney, B., & Dorroh, J. (2004). *How to write to learn in science* (2nd ed.). Arlington, VA: NSTA Press.

Toth, E. E. (2000). Representational scaffolding during scientific inquiry: Interpretive and expressive use of inscriptions in classroom learning. In L. R. Gleitman & A. K. Joshi

Formatted: Indent: Before: 0 pt, Hanging: 27 pt

Deleted: .

(Eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 953–958). Mahwah, NJ: Lawrence Erlbaum.

Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development, 51*, 1-10.

Tweney, R. D. (1991). Faraday's notebooks: The active organization of creative science. *Physics Education, 26*, 301-306.

Wells, G. (1999). *Dialogic inquiry*. Cambridge, UK: The Press Syndicate of the University of Cambridge.

Wu, H.-K., & Krajcik, J. S. (2006). Inscriptional practices in two inquiry-based classrooms: A case study of seventh graders' use of data tables and graphs. *Journal of Research in Science Teaching, 43*, 63-95.

Wisconsin Fastplants: Growing instructions. (1999). Burlington, NC: Carolina Biological Supply Company.

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

Zimmerman, C. (2005). *The development of scientific reasoning skills: What psychologists contribute to an understanding of elementary science learning*. Washington, DC: National Research Council.

Table 1.

Distribution of Mean Number of Intents and Plans (and SD) according to Participants' Experimental Design Strategies

	Complete Problem Space in Session 1 through 4		Control-of-Variables Strategy in Session 1 through 2	
	No	Yes	No	Yes
N	20	14	21	13
Mean (and SD)	3.4 (4.8)	11.3 (8.2)	2.02 (3.6)	6.38 (4.27)

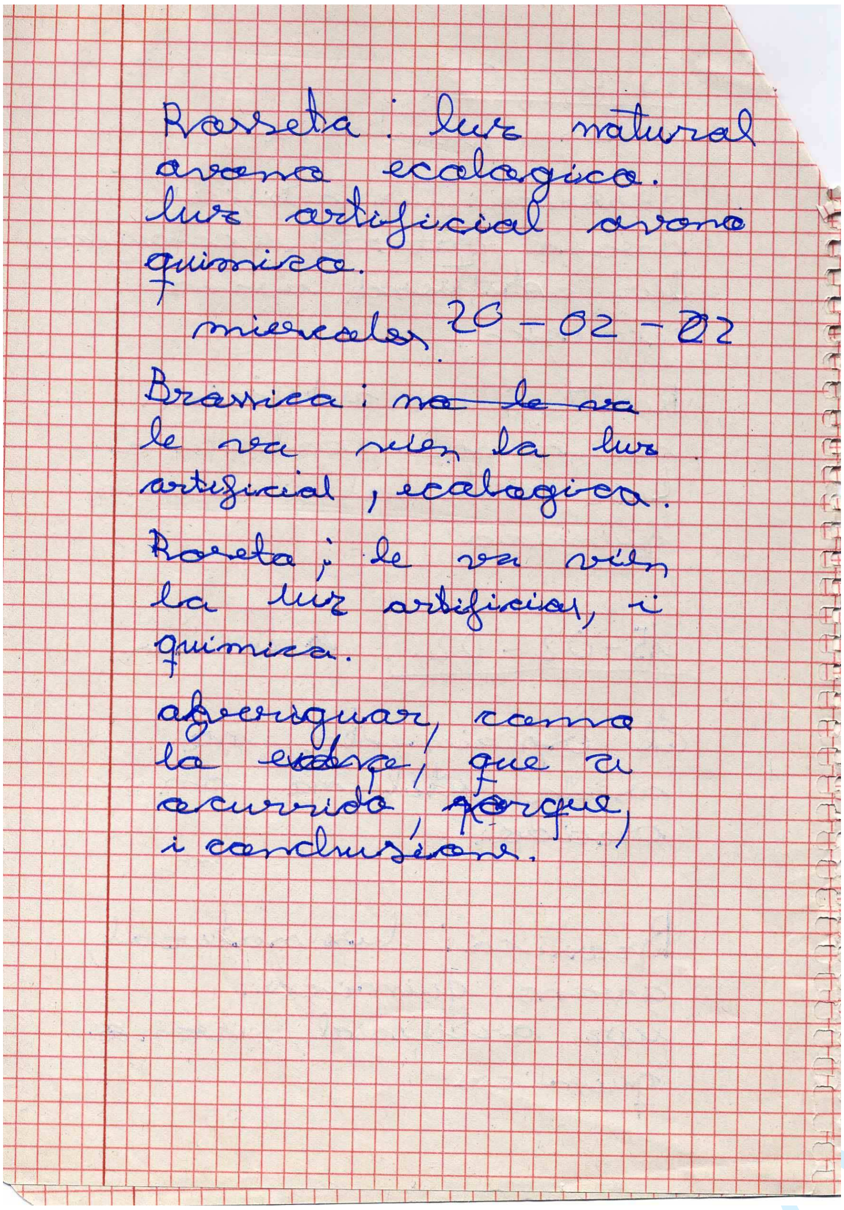
Table 2. Comparison of Participants' Initial and Final Theories

Variable	Initial	Final	Correctness
Type of seed			
Brassica>Rosette	22	32	correct
Rosette>Brassica	12	2	
Brassica=Rosette	0	0	
Type of light			
Natural>Artificial	33	6	
Artificial>Natural	1	26	correct
Natural=Artificial	0	1	
Indeterminacy		1	
Type of Fertilizer			
Chemical>Organic	9	4	
Organic>Chemical	25	29	correct
Chemical=Organic	0	1	

Table 3.

Distribution of Mean Number of Valid Inferences in Session 3 through 6 (and SD) according to Participants' Notes Revision

	Notes Reviewers	
	No	Yes
N	22	11
Mean (and SD)	2.7 (2.3)	4.6 (2.9)



Rosseta: luz natural
avona ecologica.
luz artificial avono
quimica.

miercoles, 20 - 02 - 02

Brassica: ~~no le va~~
le va vien la luz
artificial, ecologica.

Roseta; le va vien
la luz artificial, i
quimica.

averiguar, como
la estaba, que a
ocurrido, porque,
i conclusiones.

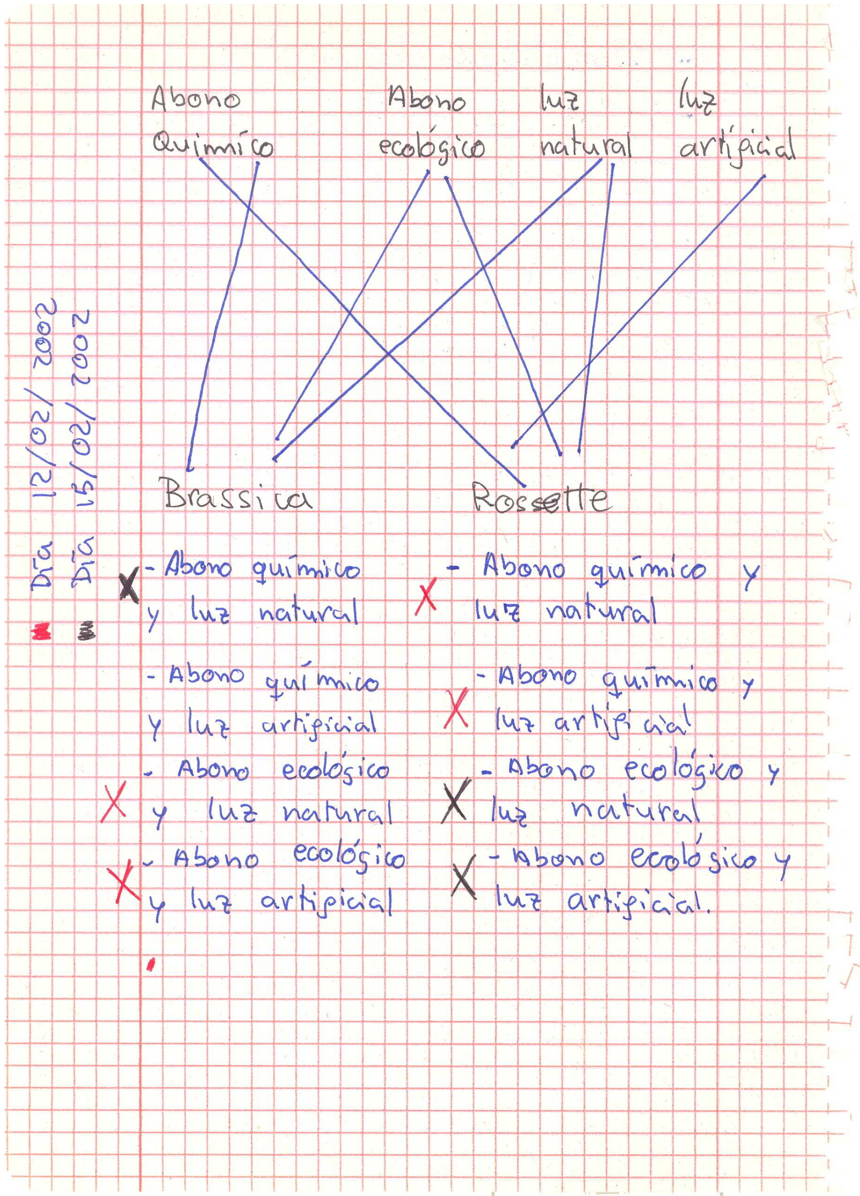
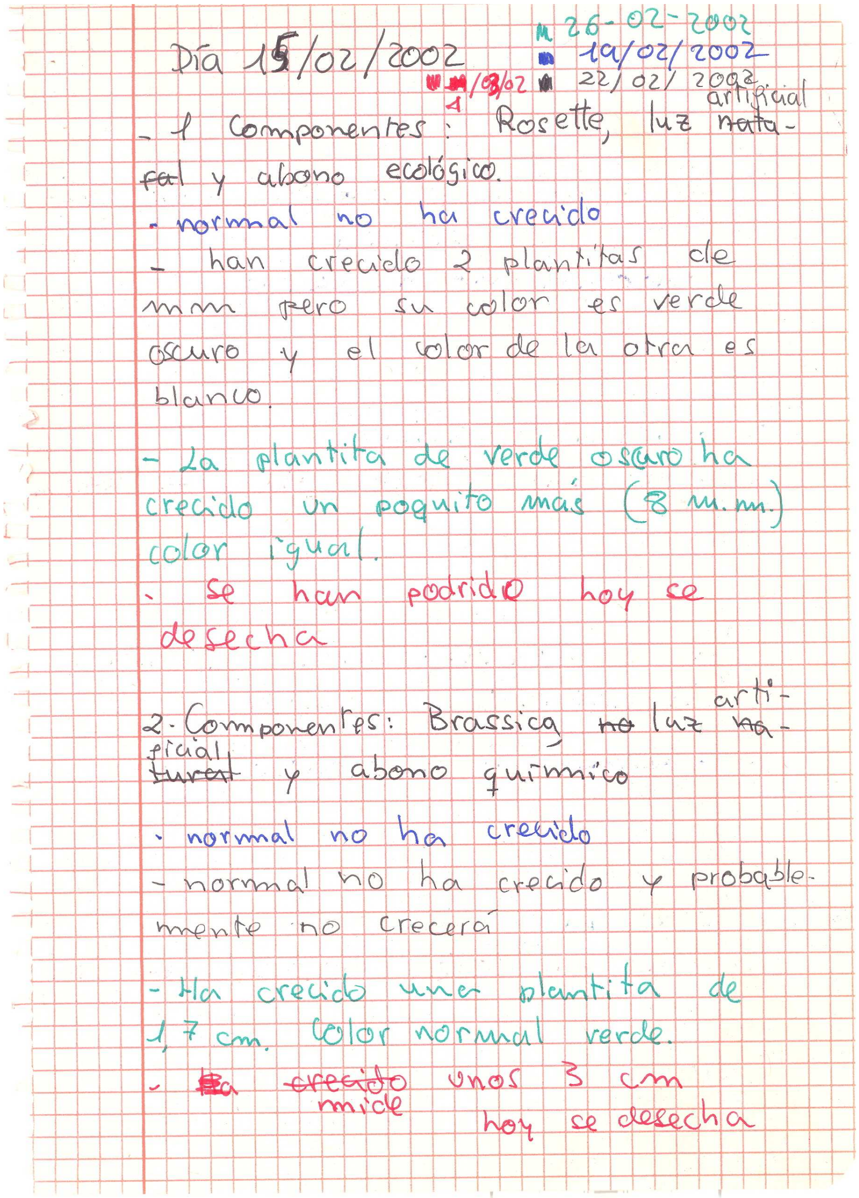


Figure 2. David's Diagram in Session 2 to Apply the Factorial Combination Strategy
157x217mm (300 x 300 DPI)

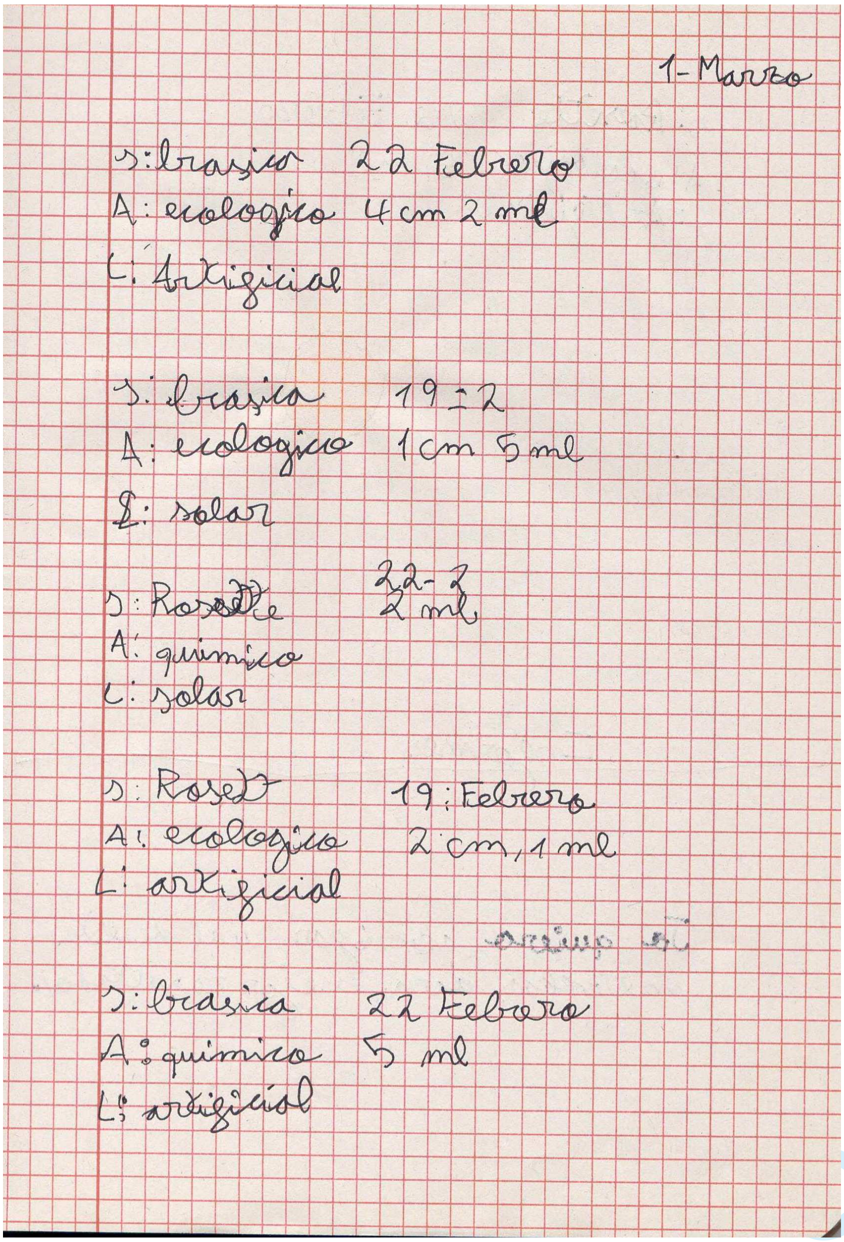

26-02-2002
19/02/2002
22/02/2002
Día 15/02/2002
/3/02
- 1 Componentes: Rosette, luz ~~natural~~ artificial y abono ecológico.
- normal no ha crecido
- han crecido 2 plantitas de mm pero su color es verde oscuro y el color de la otra es blanco.
- La plantita de verde oscuro ha crecido un poquito más (8 m.m.) color igual.
- se han podrido hoy se desecha

2. Componentes: Brassica ~~no~~ luz ~~natural~~ artificial y abono químico
- normal no ha crecido
- normal no ha crecido y probablemente no crecerá
- Ha crecido una plantita de 1,7 cm. Color normal verde.
- ~~Ha crecido~~ mide unos 3 cm hoy se desecha

Figure 3. David's Notes.
Text Translation from Spanish:
In the upper right corner David writes each data with a different colour. The page shows two entries, one for each experiment.
1. Components:
-normal. It has not grown (in blue)
-they have grown 2 little plants, mm, but their colour is dark blue while the colour of the others is light (in black)
-the little plants have grown a little more (8mm). Colour the same (in green)
-they have died. Today we discard (in red)
2. Components: Brassica, artificial light and chemical fertilizer
-normal. It has not grown (in blue)

-normal. It has not grown, and it will not probably grow (in black)
-little plants have grown, 1.7 cm, colour: normal green (in green)
-it measures 3cm. Today we discard (in red).

156x217mm (300 x 300 DPI)



1- Marzo

s: brasica 22 Febrero
A: ecologico 4 cm 2 ml
L: Artificial

S: brasica 19-2
A: ecologico 1 cm 5 ml
L: solar

S: Rosette 22-2
2 ml
A: quimico
L: solar

s: Rosett 19: Febrero
A: ecologico 2 cm, 1 ml
L: artificial

S: brasica 22 Febrero
A: quimico 5 ml
L: artificial

Appendix A.

Problem Space and Causal Structure of the Fastplant Growing System

Variable Effects	
Type of seed (Brassica-B or Rosette-R)	Brassica > Rosette
Type of fertilizer (Chemical-C or Organic-O)	Organic > Chemical
Type of light (Natural-N or Artificial-A)	Artificial > Natural

Outcomes for Each Plant of the Problem Space	
Do not germinate	Rosette-Natural light-Chemical fertilizer
Germinate/Height approx. 1cm	Rosette-Natural light-Organic fertilizer
	Rosette-Artificial light-Chemical fertilizer
	Brassica-Natural light-Chemical fertilizer
Germinate/Height approx. 5cm	Rosette-Artificial light-Organic fertilizer
	Brassica- Artificial light-Chemical fertilizer
	Brassica-Natural light-Organic fertilizer
Germinate/Height approx. 12cm	Brassica- Artificial light-Organic fertilizer