International Journal of Learning, Teaching and Educational Research
Vol. 10, No. 4, pp. 51-70, March 2015

Does Inquiry-Learning Support Long-Term Retention of Knowledge?

Sarah Schmid and Franz X. Bogner
Z-MNU (Centre of Math & Science Education),
Institute of Biology Didactics, University of Bayreuth,
University Campus, NWI, D-95447 Bayreuth, Germany

Abstract. Structured Inquiry-Based Science Education (IBSE) provides the theoretical base for our 9th grader lesson, labeled "The Hearing of Sound". Participation in the 3 consecutive lessons enables participants to explore the phenomenon of hearing. Participants complete matching hands-on experiments, learn about the theoretical background of these experiments, describe observations and formulate explanations. The study followed a quasi-experimental design with 138 students. The participants' content knowledge on the subject was monitored 4 times: 2 weeks prior to the lesson (T0), on the day of the lesson (T1), 6 weeks (T2) and 12 weeks after it (T3). Students gained a significant short-term and a long-term increase in knowledge scores after 6 weeks. Furthermore, students showed a constant level of content knowledge when tested after 12 weeks, indicating that students did not forget information within the last six weeks. Furthermore, our inquiry lesson was suitable for both genders, as well as students with both high and low pre-knowledge. In their pre-knowledge boys outperformed girls. However, there was no influence of gender on the knowledge score after the lesson (T1, T2, T3). Conclusions for everyday teaching in school, by using inquiry teaching more frequently are discussed.

Keywords: Inquiry-based science education; secondary school; gender; long-term retention

Introduction

Inquiry approach

According to constructivism, knowledge cannot directly be transmitted from one person to another. Instead knowledge construction occurs through active thinking of the learner (Cakir, 2008). Inquiry based learning is a constructivist approach of learning. It enables the learner to construct concepts from experience and from verbal interaction. Through inquiry learning, students have

the opportunity of getting first-hand experience in doing science and, to develop inquiry skills (Tamir, 1985).
Inquiry can be defined as "the intentional process of diagnosing problems, critiquing experiments, and distinguishing alternatives, planning investigations, researching for information, constructing models, debating with peers and forming arguments", (Linn, Davis, Bell, 2004, p.XVi). Therefore, inquiry-based teaching is increasingly favored as an improvement tool in science education (Osborne & Dillon, 2008; Rocard et al., 2007). For example, the German national science education standards propose four main competence domains: subject-specific content knowledge, communication, judgment, and methodological knowledge (KMK, 2005a, 2005b). Within "methodological knowledge", many inquiry-learning related activities are listed and their importance for teaching Biology and Physics is underlined. Although the benefits of using inquiry-based science education (IBSE) are controversial (see Furtak et al., 2009), the proposed benefits have often been confirmed (Wilson et al, 2010; Lynch et al., 2005; Minner and Levy, 2009). Concerning reduction of the increase of knowledge and the gender gap, Geier, Blumenfeld, Marx, Krajcik, Fishman, Soloway, and Chambers (2008) and Secker and Lissitz (1999) found results promoting inquiry based science. However, especially studies regarding long-term retention are rare, which is why the present study was undertaken.

Inquiry-based teaching can be further subdivided into levels of student's autonomy. Blanchard et al. (2010) provide a clear table of the inquiry levels, based on Abrams, Southerland and Evans (2007) and adapted from Schwab (1962) and Colburn (2000), which describes four levels of inquiry: Level-0/verification, level-1/structured, level-2/guided and level-3/open. The higher the level, the more student autonomy is linked to a block of investigation. For example, in level-0 the teacher decides the question, the method for data collection and also interprets the results. In level-1, interpretation is already up to the students. In level-2, interpretation as well as finding a method for investigation is up to the student. In level-3, the formulation of a question to investigate is also the students' responsibility. Abrams et al. (2007) stated that for level-3 students need prior experience with inquiry-learning to learn appropriately. In our quasi-experimental study, for comparison reasons, all students needed to work on the same content. Therefore, the inquiry lessons were constructed as "level-1" or "structured" inquiry, where the students focus on understanding the link between experiment and theory, and as well on the interpretation of the results they obtain (Staver and Bay, 1987). Similar to Quintana et al. (2004), we see scaffolding as a key element of cognitive achievement. The teacher coaches, structures tasks in a meaningful way and gives hints, without explicitly giving the solutions to the students. Furthermore, we agree with Mayer (2004), who stresses that a certain way of guidance is always useful in learning instructions. In his review of studies between 1950 and 1980, he found evidence in favor of guided approaches to learning, like guided discovery, whereas unguided, problem-based instructions such as pure discovery learning did not work well. Mayer (2004) defines guided discovery methods [like structured or guided inquiry] as receiving problems to solve where a teacher provides "hints, direction, coaching, feedback and/or modeling to keep the student on track". The main activity of a teacher's work takes place before the actual lesson. It is his/her responsibility to develop and provide high qualitative learning material. That is, a teacher's role is reduced to the promotion of inquiry-learning by asking open-ended questions like "Tell me what you think about this."; giving students time to think of an answer; responding to questions by repeating and paraphrasing what students said without praising or criticizing; avoiding telling students what to do; and maintaining a disciplined classroom (Colburn, 2000).

Inquiry-learning material

"Clearly, the contemporary view of how students learn implies content that is deeper than facts and information, a curriculum that is richer than reading, […], and teaching that is more than telling": In line with this view of Bybee (2002; p.29), inquiry material should contain the content knowledge that a teacher wants to impart, but leaves space to explore the content. This can be done with the help of questions raised during engagement with the material. A student needs to find a satisfactory answer to such questions, by using his or her motivation und curiosity. This is in line with the claim of Novak (1988) that "most students are not aware that learning is a responsibility they must accept. Teachers have a responsibility to select meaningful material and seek to share his meaning with students, but only the student can choose to learn. They can choose to learn by rote or to learn meaningfully". The questions raised in the material should have a connection with the everyday life of students and, therefore, connect school contents with relevant scientific problems that might occur throughout a student's life. Consequently Bransford, Brown and Cocking (2000; p. 139) had stated: "Ideas are best introduced when students see a need or a reason for their use - this helps them see relevant uses of knowledge to make sense of what they are learning".

In the strategy of inquiry-based science learning, students are the main actors during a lesson. Hands-on working is an important part of inquiry-learning that helps to facilitate the understanding of theoretical knowledge and ideas. However, we agree with Mayer (2004) who stated that "activity may help promote meaningful learning, but […] the kind of activity that really promotes meaningful learning is cognitive activity (e.g., selecting, organizing, and integrating knowledge). Instead of depending solely on learning by doing […], the most genuine approach to constructivist learning is learning by thinking". This is what is meant by the difference of a "cookbook"-level-0-instruction and higher level inquiry. Furtak et al, (2009) stated that the "procedural facet of inquiry-based science" does not involve students proceeding mindlessly through scripted laboratory procedures. Lord and Orkwiszewski (2006) agreed that "While active learning suggests students are physically participating in the lesson, inquiry-learning requires that they are also mentally participating in it".

The structured inquiry lesson we invented for this study was designed to turn the responsibility for learning to the students, by centrally involving them in the thinking process about what the given experiments where for and in linking their observations to the theory provided. The lesson topic was the human ear and acoustics. Students were expected to learn how sound is created, how it is transferred to our ears as well as about the ear's anatomy, including the organization and function of the ossicles and the organ of Corti. As they also learned about the characteristic of sound waves, they were expected to use this newly acquired knowledge when learning about the function of the organ of Corti and the basilar membrane, as well as incorporating this knowledge to derive causes why noise is dangerous for our ears or the natural limitation of human hearing to a certain range of frequencies. In summary, the main learning goal of our inquiry-based lesson was the learning of science – defined as "acquiring and developing conceptual and theoretical knowledge" (Hodson, 2014, p.2537).

Within the three hours, four-person-groups worked on their own, using a provided working booklet. From time to time the teacher was questioning a student group, without involving the attention of the other groups, why they decided to write down a certain answer and let them explain their opinion. It was not sufficient to know what the result of an experiment was, but the material yielded on students understanding why this was the correct answer. Therefore, the teacher only turned to another group, if the students could explain why their answer was correct from their viewpoint. In case a student of the group did not know why the answer the group had decided for, was correct, he let another group member explain it to him. In case the group as a hole was not yet sure of why they decided for their answer, the teacher encouraged them to reread the text and/or observe the experiment again and would come back to them in a couple of minutes, meanwhile questioning another group. Summed up, students needed to read information texts to their group mates fetch and put together experimental set-ups, conduct experiments and answer questions regarding their observations and interpretations of them in regard to the theory provided. It was the students' responsibility to read the texts, fetch the setup, built it up and conduct the experiment in order to be able to answer the provided questions. When students e.g. asked "What do I need to fetch for this experiment?", the teacher would advise them to read the work booklet again. If questions like "Is this the correct answer?" would occur, the teacher would let them explain why they think it may or may not, and therefore let them reconsider their own answer. The students had the opportunity to test the experiments as many times as they liked and also to adapt them to their needs, if they were curious about it expanding the given questions. The teacher only asked groups to stay focused, when their exploration would shift to mere play without longer concentrating on the questions to answer or when they were behind the other groups – although both rather rarely took place. Provided material consisted of simple devices like springs, cards, rulers or paperclips etc. but also musical instruments, laptops and freeware software to display graphs of sound waves produced by the students. The phenomena in the texts were carefully selected to connect to students' everyday experiences: like, when questioning how sound is created to think about a rattling window when a truck passes by to focus their attention to vibration. Or when introducing sound intensity and how to protect your inner ear by enlarging the distance to the sound source, with questioning what opportunities they had in a discotheque, when they stand near the speakers. Or when questioning, why the eardrum is larger than the oval window, leading to recognition of a force concentration, giving examples of e.g., why pins are designed the way they are and let them explain observations made with a matching experiment in regard to force concentration and pressure increase.

Our study objectives were three-fold: (i) Classes that learned lesson contents via our structured inquiry unit were analyzed for their capability to comprehend and remember the lesson content, i.e. their content knowledge was repeatedly monitored. Although inquiry is postulated to support many soft skills, in our study we focused merely on cognitive achievement (since it is the main component on which students are assessed for during their school life): If IBSE did not lead to a satisfying learning outcome concerning content knowledge, its soft skill component would be insufficient justification for its use as a daily teaching strategy. Therefore, the first hypothesis was that participating in the provided structured inquiry unit would lead to a significant knowledge increase. (ii) Second, we hypothesized that learning science contents through structured inquiry would lead to a deeper understanding of the content and hence to secure long-term retention. (iii) Third, we hypothesized that our structured inquiry lesson may lead to significant learning in both genders.

Methods

Our study followed a quasi-experimental design (Mertens, 2010) in which 138 9-graders (Gymnasium), from ten classes and 4 schools participated. The sample size contained 47.83% females. The mean age of the participants was 15.1 years (SD=+/-0.55). To all students the taught topic was new. Before grade 9, human senses (NT 5.2.2) only are taught in an overview-like manner with some introduction to air and sonic (NT 5.1.2) in grade 5. A control sample of 64 students from 3 classes was monitored, with 42.19% females and mean age 14.80 years (SD=+/-0.52). This control sample did not take part in the inquiry classes and received no instruction on the topic during data acquisition. A repeated measurement ANOVA was applied. The control group did not learn through repeated completion of the knowledge questionnaire: there was no significant impact on the knowledge score at the four measurement times, p=n.s., F(3, 192)=1.09, partial eta-squared=0.017.

The applied knowledge scale was embedded into a larger questionnaire which was completed within approximately 40 minutes. All questionnaires were completed within a 14 week schedule. Total data acquisition was completed within one year.

The intervention "The Hearing of Sound" consisted of three consecutive school lessons of 45 minutes each, forming an interdisciplinary inquiry unit about how humans hear and what sound is. The content knowledge covered, therefore, combined issues from Biology and Physics. A commented excerpt of the working booklet is attached in the appendix.

The four topics of the lessons were:

1. "What is sound?": a) Sound creation & Movement, b) Frequency & Amplitude
2. "How do we hear?": a) The outer ear, b) The middle ear
3. "How do we distinguish frequencies?": a) Resonance and Eigen Frequency, b) The inner ear
4. "The limits of hearing": a) Natural limits of hearing, b) The risks of noise

The topics within the "working booklet" were arranged roughly according to the following scheme:

1. Title focusing the students on the sub-topic
2. Introductory text about a phenomenon students may have encountered in their life
3. Raising the question about how that phenomenon could be explained
4. Giving an example that helps students to understand the phenomenon
5. Asking questions to be investigated with an experiment or model (setup provided)
6. Students investigate these questions with the experiment or model
7. Students are asked to combine their results with the knowledge gained through the experiment or model, link it to the science behind it, answer questions, as well as fill in clozes to verify their knowledge.

Each group member was given a special task that was rotated between the four group members to ensure every student was engaged and to prevent "the couch potato phenomenon", described by Lord (1999). These roles were: a) reading the text aloud, b) fetching the right experimental material from the box, c) conducting the experiment d) writing down the group's conclusions. The student groups worked without tight time constraints.

The teacher had the role of a guide. The teams worked on their own and were asked to contact the teacher just when something could not be solved within the group. The inquiry lesson was the only source of information on the topic. Students were not aware of the repeated testing cycle. Beside the participation in the three hour course, no additional learning phase took place. Therefore, the acquired knowledge is dedicated to the learning during participation in the structured inquiry course. All lessons were conducted by the first author to minimize teaching style dependent effects, and to verify that the teaching material was not changed due to personal preferences of teachers, and was therefore used in the same way in each class.

Questionnaires for evaluating students' content-knowledge were applied about two weeks prior to the school lesson (T0), directly after the school lesson (T1), as well as six (T2) and twelve weeks (T3) after it. We also opted for a further long-term measurement after a period of twelve weeks, since many studies report a decline of content-knowledge in their retention-test (e.g., after six weeks), but the development after this time gap is rarely researched (Bogner, 1998; Dean and Kuhn, 2006).

Figure 1: **Testing schedule. Back area= lesson. T0= two weeks prior; T1= directly after; T2= six weeks after; T3= twelve weeks after the lesson.**

The ad-hoc content knowledge questionnaire consisted of 20 multiple choice items with four possible answers each, only one of which was correct. The item-difficulty-index (Zöfel, 2002) was calculated for each item; subsequently 3 questions were excluded as they were correctly answered by over 80% of persons in the pre-test and thus were not meaningful for analyzing knowledge gain (Zöfel, 2002).

As repeated application of an identical test might influence achievement scores due to repetition (Keeves, 1998), students were never aware of any testing schedule or of any repeated testing situation. Additionally, the order of the questions and the position of the right answer were randomly distributed for each time point. All questionnaires were handed out as paper and pencil tests to be completed under controlled conditions. The maximum possible score of the knowledge test was 17. In classical test theory the mean-item-difficulty is the percentage of participants that answered the item correct. Over all testing time points it was 50.3% (T0=30.2%, T1=67.0%, T2=52.9% T3=51.1%). Classical test theory was applied for Cronbach's Alpha. The multiple choice test had a mean reliability index of .67 (Cronbach's Alpha; T0=.632, T1=.709, T2=.647, T3=.679). All items aim for measuring content knowledge and had a positive discrimination index.

What is the frequency of a particle that vibrates 1000 times in 2 seconds?

☐ 1000 hertz
☐ 2000 hertz
☐ 500 hertz
☐ 250 hertz

Which characteristic of a sound determines how much the basilar membrane will be moved?

☐ The frequency
☐ The rhythm
☐ The amplitude
☐ The speed

When a sound wave travels through the room ...

☐ ... it causes a wind to blow that carries the sound particles
☐ .... the wave carries "sound particles" from the sound source to the ear
☐ ... it causes a repetition of refraction and condensations in the air
☐ ... the wave changes the frequency

'Ultrasound' is a sound that is ________ than humans are able to hear.

☐ Louder
☐ Higher
☐ Faster
☐ Slower

Figure 2: **Example questions from the multiple choice knowledge questionnaire.**

For statistical analyses SPSS (version 20) was used. Missing data was excluded list-wise. Raw-data was transformed to gain homogeny of variance needed for parametric tests (knowledge scores: log(10)+1; knowledge gain: log(10)+13). All statistical tests concerning knowledge scores or gain take these transformed data as source. However, for a more meaningful interpretation graphs are based on the untransformed data. Data was analyzed by t-test, repeated measurement ANOVA (mixed design) and one-way independent ANOVA.

Results

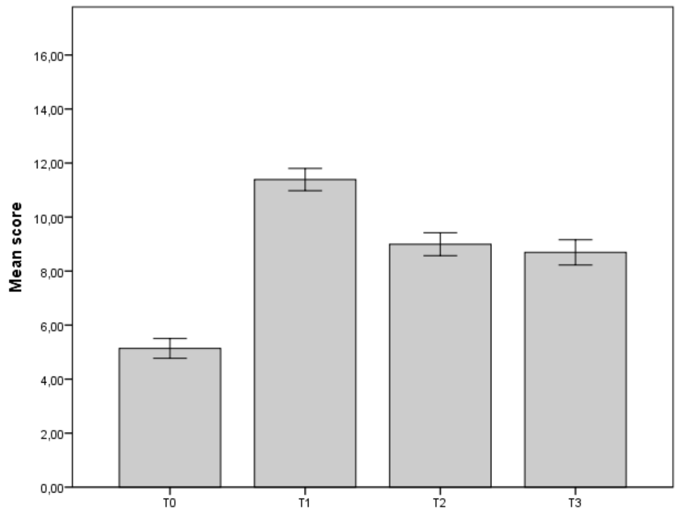


Figure 3: **Bar chart of the mean number of right answered knowledge questions grouped for each time point of measurement. T0 =2 weeks prior, T1=directly after, T2=6 weeks after, T3=12 weeks after the inquiry lesson. Max=17 right answered questions. Error bars = 95% CI.**

Repeated measurement ANOVA was applied to test if the consecutive time points of measurement differ significantly. Knowledge scores differed significantly over the time points: $p<0.0001$, $F(2.1, 286.22)= 173.52$, partial eta-squared =0.561.Students learned significantly through the inquiry lesson. The knowledge score of T0, two weeks before the lesson, was significantly lower than the knowledge score of T1, directly after the lesson, $p<0.0001$, $F(1,136)=351.24$, with a large effect size partial eta-squared=0.721. Afterwards, some content knowledge was lost again, comparing T1, directly after the lesson and T2, six weeks after the lesson, $p<0.0001$, $F(1,136)=105.36$, eta-squared=0.437. However, a further six weeks later (T2 -T3), the students lost no more knowledge, meaning that they are capable of recalling as much after twelve weeks as they could recall after six weeks, p=n.s., $F(1,136)=0.05$, partial eta-squared=0.0001.

A second repeated measurement ANOVA was used to test, if the knowledge scores for short and long-term learning were significant. The inquiry intervention lead to a significant short term increase in knowledge scores directly after the lesson (T0-T1) $p<0.0001$, $F(1, 137)= 322.41$, with a large effect size of partial eta-squared=0.702,as well as to a long-term increases in knowledge scores after six weeks (T0-T2) $p<0.0001$, $F(1, 137)= 148.12$, with a large effect size of partial eta-squared=0.521, and after 12 weeks (T0-T3) $p<0.0001$, $F(1, 137)= 135.50$, with a large effect size of partial eta-squared=0.497.

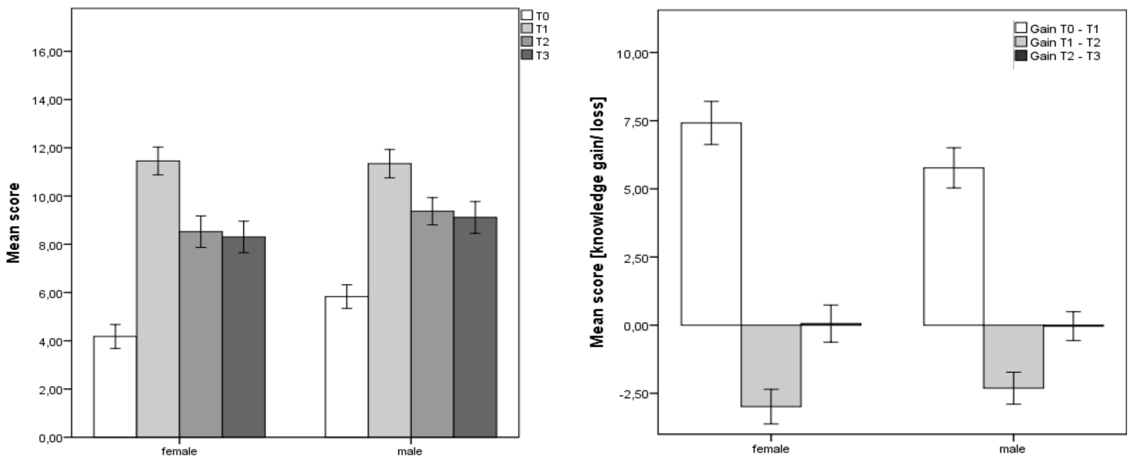


Figure 4: ***A/*****Left: Bar charts of the knowledge score at successive time points of measurement, separated for male and female. X-axe: Gender, Y-axe: mean number of questions answered right; max=17. B/Right: Bar charts of the median gain (or loss) of knowledge, grouped for the time span between consecutive time points of measurement. F.l.t.r.: T0-T1, T1-T2, T2-T3.T0=2weeks before, T1=directly after, T2=6 weeks after, T3=12 weeks after the inquiry class. Error bars= 95% CI.**

Figure 4a shows the knowledge level throughout the different time points of data acquisition for boys and girls. To test a potential gender discrepancy regarding pre-knowledge score (T0), a t-test was applied. As the genders differ in their pre-knowledge, with boys outcompeting girls (meanT0♂= 0.8115, SE=0.20; mean T0♀=0.65, SE=0.23, $p<0.0001$, $r=0.35$), a repeated measurement ANOVA with mixed design was applied for further analyses. The interaction of gender and the time-points of measurement of knowledge scores was significant, $p=0.002$, $F(2.11, 286.22)=6.39$, partial eta-squared=0.045. This means that the overall knowledge scores are different for boys and girls.

In detail, gender produced a highly significant effect when comparing two weeks before the lesson (T0), and directly after the lesson (T1), $p<0.0001$, $F(1,136)=11.22$, partial eta-squared= 0.076. This is due to boys starting with a higher pre-knowledge than females (see t-test above).However, gender did not influence further learning. There was no influence of gender on the knowledge score directly after the lesson (T1) to six weeks after the lesson (T2), *p*=n.s., $F(1, 136)=1.44$, partial eta-squared= 0.010. Nor was there a significant gender effect in knowledge level six weeks (T2) and twelve weeks (T3) after the lesson, *p*=n.s., $F(1,136)=0.21$, partial eta-squared= 0.002. This means that boys and girls reached the same knowledge level, despite different pre-knowledge, as well as that the levels of recall of content knowledge were the same for both genders.

Figure 4b shows the knowledge gain of boys and girls. For analysis a one-way independent ANOVA way applied. The knowledge gain from T0 to T1 was greater for girls than for boys. This difference in knowledge gain was significant $F(1,136)=4.0$, $p=0.048$, $r=0.17$. The higher direct benefit from the inquiry lesson for girls was due to the females starting with a significantly lower pre-knowledge in T0 (see above/ figure4a). The knowledge loss from T1 to T2, as well as from T2 to T3 were not significantly different for boys and girls, $F_{T1-T2}(1,136)=0.147$, p=n.s, $r=0.03$, $F_{T2-T3}(1,136)=0.003$. p=n.s., $r=0.001$.

Discussion

Inquiry lesson leads to substantial learning

First of all, our data (figure 3) follow typical learning pattern with a short-term peak and long-term decrease (e.g., Heyne and Bogner, 2012; Geier and Bogner, 2010): students learn and partially forget content-knowledge, but the level never drops below the before lesson level. So students learned substantially through a structured inquiry lesson. As no additional learning phase on the course content took place, the knowledge reflected in the test-scores must have been learned through the active participation in the course. This finding is in line with Von Secker and Lissitz (1999), who described teacher-centered instruction as negatively associated with general science achievement, while mean science achievement is expected to increase by about 0.4 SD for every SD increase in the amount of emphasis placed on laboratory inquiry. Marx et al. (2004) also showed that in a year-long study in an urban school area, where students were historically low achievers in science, students were able to significantly improve scores via inquiry as demonstrated by post-tests after the school year. Lederman, Antink and Bartos (2014) also described inquiry-based learning as very helpful for learning science contents as well as hands-on work as effectively supporting conceptual understanding. Bases on data of the PISA-study 2006, increases in science achievement are especially provoked by structured inquiry; i.e. the same level our inquiry-based unit was based on, where students conduct hands-on activities and draw conclusions upon them. Interestingly, higher, more open levels of inquiry resulted in much lower science achievement (Jiang and McComas, 2015).

Inquiry lesson leads to formation of long-term retention

Even after twelve weeks, students could recall as much as they did six weeks after a lesson, indicating a constant level for six weeks. This result strongly suggests that learning through inquiry-lessons substantially contributes to the formation of long-term retention.

There are very few studies about long-term effects of inquiry lessons. Anderson (1997) stated that neurocognitive theory supports learner-centered science instructions: "Active involvement by the learner maximizes activation of schemata, […] especially when the tasks are perceived as being facilitatory for problem-solving […]. Multimodal learning promotes more stable schema formation by simultaneously activating different cortical modules and enhances linkages among the [neural] networks. This stabilizes the information and makes it more accessible for reconstruction later […]". This would explain why students retained information from week 6 to week 12 after the inquiry intervention. The level of knowledge formed within the first six weeks after a lesson would remain long-term. Nuthall (1995) observed how students use their knowledge and experience to respond to achievement tests by using multiple-choice questionnaire before an intervention and directly after. Selected interviews two weeks after the intervention yielded reasons why specific answers had been given and where students obtained their information from (e.g., school or family and media). A further 12 months later, participants completed the questionnaire again, and thought "aloud" about how and where their information originated from. In total, recalling relevant item answers was closely associated with the recollection of the episodic and/or semantic content of the original learning experience. Since the original learning experience was so important for recall, this suggests why our structured inquiry lesson apparently helped students not to forget information between week 6 and 12. As they took responsibility for working through the learning material and conducting experiments with their group, students were actively involved at every stage of knowledge building about a subtopic of the lesson. For instance, they read about the phenomenon, tested models, were asked how the model relates to the real phenomenon, were asked to decide if the fillings of a cloze were right or wrong, were urged to discuss problems and results with their peers and write their answers and conclusions down. This probably provided a broad base from which to retrieve a correct answer to the content-knowledge test. As Nuthall (1995) concluded, "students who have learned more in classrooms may perform better on achievement tests, not because they have learned more answers, but because they have more alternative parallel ways of solving the problems posed by the test items".

Matching results also come from Blank (2000), who described students as recalling significantly more content-knowledge, two and six months after the end of an inquiry intervention, when they had been asked to make explicit their prior knowledge and to discuss the status of their conceptions and ideas throughout the course (meta-cognitive group), compared to students who completed the same inquiry lessons, but were not asked to record their thoughts. The actual learning gains of both groups were similar, directly before and after the treatments. However, long-term retention two and six months later differed: Students of the metacognitive learning cycle could recall significantly more. The students in our inquiry lesson were required to think actively to make sense of the written information, the experiments and their observations, while discussing the content with their peers. We conclude that active participation by hand and brain made the experiences during the lesson very "vivid" in the memory of the students, so that the learned content could be retrieved relatively easily, even 12 weeks after the intervention.

Although our inquiry lesson lasted just for half a day, significant learning outcomes and long-term effects on content learning were achieved. Together with the study of Geier et al. (2008), that indicates that a higher frequency of inquiry units results in higher achievement, this should be a signal towards the more frequent use of well-organized inquiry lessons. The meta-analysis of Furtak et al. (2009), also found indications that an inquiry treatment of between five and seven weeks had a higher impact on students learning than those lasting up to one week, supporting the higher frequency.

Inquiry can close the gender gap and is suitable for strong and weak students
Although males scored significantly higher on pre-knowledge before the lesson, our inquiry lesson made this gender difference in content-knowledge disappear. Immediately after the lesson, the female's mean score no longer differed from the male's one, although both genders had significantly increased knowledge through the lesson. Even with lower pre-knowledge females were able to fully catch up.

Supporting results come from Geier et al. (2008), who found that participation in at least one inquiry course resulted in a reduction of the gender gap. African-American boys, who are reported as often being outperformed by girls in US urban schools, closed the gender gap by gaining relatively more from an inquiry unit. This is in line with our results. Therefore, our study adds to the body of research that shows that inquiry-learning can help to reduce the gender gap, especially by being relatively more helpful for the disadvantaged group of learners without ignoring the needs of the rest of the class.

Achievement gaps in science classes can probably explain a high percentage of the skewed decision between the genders for later career opportunities in science. We would therefore welcome more inquiry-based science lessons throughout a student's school life, to increase decisions for a career in science. Especially if a more equal distribution in science careers between genders is desired by policy makers and politics, structured inquiry-teaching could be the solution. Inquiry does not especially suitable for a certain gender, rather inquiry is suitable for all students, but it reduces the barriers for learning and retention. The student centered way of teaching inquiry, where students can work and learn in small groups, in their own pace, and with less competition gives room to students to not to be afraid of asking questions or discussing with peers by tasks that are engaging, foster curiosity and motivation. These are circumstances especially beneficial for students that otherwise may not even want try to investigate the learning content, because they feel not capable or are afraid of social pressure. These may be the reasons why the learning environment of inquiry based teaching seems to be especially fruitful for thitherto disadvantaged students.

4. *Pre-knowledge is not crucial for achievement*

Females not only caught up directly after the intervention, but their later scores did also not differ from the boys' in week 6 and 12. Thus, although girls needed to learn more new facts, they were able to remember these new facts for at least 12 weeks. As they did not forget more than the boys, this indicates that they were capable of remembering more facts in total (T0-T3. As a conclusion, girls were able to gain more from our inquiry lesson which brought both gender to the very same level of content-knowledge without disadvantaging either. This result precludes pre-knowledge as a limiting factor for learning by inquiry and demonstrates that this teaching method can be an especially positive way of learning when the class consists of strong or weak students. Therefore, we can confirm Johnson and Lawson (1998), who found that prior knowledge did not account for a significant amount of variance in final science exams of college students, regardless whether taught with [open] inquiry or expository methods.

Conclusion

Our study showed the hypothesized positive effects of inquiry-based science teaching, leading to a better learning outcome and showing its suitability for both genders, but especially supporting female learning needs (Rocard et al 2007). Our inquiry-based lesson was shown suitable for students with both high and low pre-knowledge. Additionally, it also seems equally suitable for both genders, as both ended up with a similar learning outcome despite significant differences in pre-knowledge. In summary, this inquiry lesson led to a well-informed class, despite individual differences between participants. It also led to a class that evidently has successfully organized long-term retention, as they are capable of recalling information and have understood principles needed to answer content knowledge questions. With regard to what students go to school for, and why teachers are actually teaching, an at least 3-month recall ability seems to be a very strong argument in favor of structured inquiry.

This long-term retention was formed using a structured inquiry approach, and we agree with Mayer (2004), who states, "Activity may help promote meaningful learning, but instead of behavioral activity per se (e.g., hands-on activity, discussion, and free exploration), the kind of activity that really promotes meaningful learning is cognitive activity (e.g., selecting, organizing, and integrating knowledge). […] the most genuine approach to constructivist learning is learning by thinking. Methods that rely on doing or discussing should be judged not on how much doing or discussing is involved but rather on the degree to which they promote appropriate cognitive processing. Guidance, structure, and focused goals should not be ignored".

To achieve long-lasting knowledge students need to be given demanding learning goals. It is quite an effort to create learning environments that are more than a lecture talk and more than discovery learning. Students need directions and the possibility to think for themselves. This needs time, but it seems that the actual value of doing and understanding something oneself is that these facts can be remembered better (Nuthall, 1995; Blank, 2000). A very important point when students are engaged in active understanding is that a teacher has confidence in his students. If a student asks how a certain word is spelled correctly, the teacher should not divulge the answer, but teach the student to look it up in the dictionary. The student will not only learn and remember how the word is written, but he will learn additionally how to solve similar questions. Essentially this is what structured or guided inquiry is about. It's about teaching content through the understanding of mechanisms that can be applied in general, and it teaches students to take responsibility for their learning. Staying with the example of the dictionary, in a modern classroom, walking to the shelf to look up a word must not be classified as disturbing the lesson, but as taking advantage of the learned mechanism.

With the background of this and other studies that show that learning with structured or guided inquiry is beneficial, there is a demand for change in classrooms and schools. To be able to teach by inquiry-learning, teachers need to be given professional development on the difference between what is "too much" and "too little" guidance for their students. This is not only necessary for class management, but in the creation of learning tasks or worksheets with which the students can learn long-term. Divulging the answer too quickly and using "cookbook" labs will probably lead to the "couch potato"-effect (Lord, 1999) and students who learn by rote learning rather than by understanding. When lab time is to be applied of which students should do more than do the given steps of an experiments and copy the meaning of the results in their documents, time is needed. Therefore there is a need to extend the time in which students are confronted with science subjects in school. Science should be taught in e.g., two consecutive hours instead of two single hours, and there should be time for "science days" or "research weeks".

It is probably necessary to reduce the gender gap long before university entry. Inquiry-learning seems to be a good tool to narrow or close this gender gap e.g., in middle school. Politics now need to emphasize the goals they proposed in the teaching standards in 2005 (KMK, 2005a, 2005b), and provide teachers and schools with matching support to increase their expertise e.g., in teaching activities similar to inquiry-learning (see "methodological knowledge"-goals).

Inquiry is not the only teaching method that can be used. It is clear that it cannot be used to the full, e.g., because of time constrictions. But teachers can use inquiry by focusing on certain aspects whenever they see the chance to take advantage of it. For example, the focus can be set on formulating hypothesis on the base of background information, planning experiments for a desired outcome, predicting outcomes of experiments, collecting data, structuring data, evaluating data, formulating explanations and results, comparing predictions to results, communicating results, or discussing results and interpretations. Teachers should also demand that their students be more pro-active when the circumstances allow for it – for example, when inquiry-based lessons are undertaken. Already a single event of a three hour inquiry lesson had a positive effect on the long-term retention of content knowledge in the study presented here. As other studies, like the one of Geier et al. (2008) and Stohr-Hunt (1996), the present study suggests a cumulative effect of the frequent use of inquiry during a student's school life. More frequent use of structured or guided inquiry in school will probably increase the benefits for students in science classes. It needs to be clear that students need not remember each single activity or experiment years after they have left school, but they need to remember how they can approach a certain type of problem, or how they can assess scientific debates in the media relevant to their lives.

Acknowledgments

Our study was funded by the European Commission (PATHWAY, FP7-Science-in-Society-2010-1, THEME SiS-2010-2-2.1-1; 266624). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the European Commission.

References

Abrams, E., Southerland, S. A., & Evans, C. A. (2007). Inquiry in the classroom: Necessary components of a useful definition. In E. Abrams, S. A. Southerland, & P. Silva (Eds.), Inquiry in the science classroom: Realities and opportunities. Greenwich, CT: Information Age Publishing. Anderson, O. R. (1997). A neurocognitive perspective on current learning theory and science instructional strategies. *Science Education, 81*(1), 67-89.

Blanchard, M. R., Southerland, S. A., Osborne, J. W., Sampson, V. D., Annetta, L. A., & Granger, E. M. (2010). Is inquiry possible in light of accountability?: A quantitative comparison of the relative effectiveness of guided inquiry and verification laboratory instruction. *Science Education, 94*(4), 577-616

Blank, L. M. (2000). A metacognitive learning cycle: A better warranty for student understanding? *Science Education, 84*(4), 486-506.

Bogner, F. X. (1998). The influence of short-term outdoor ecology education on long-term variables of environmental perspective. *The Journal of Environmental Education, 29*(4), 17-29.

Bransford, J. , Brown, A., and Cocking, R., eds. 2000. *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

Bybee, R. (2002) Scientific inquiry, student learning and the science curriculum. In R. Bybee (Ed), *Learning science and the science of learning*.(pp. 25-35) NSTA Press, Arlington, VA.

Cakir, M. (2008). Constructivist Approaches to Learning in Science and Their Implications for Science Pedagogy: A Literature Review. *International journal of environmental and science education, 3*(4), 193-206.

Colburn, A. (2000). An inquiry primer. *Science Scope, 23*(6), 42-44.

Dean Jr, D., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education, 91*(3), 384-397.

Furtak, E., Seidel, T., Iverson, H., & Briggs, D. (2009*). Recent experimental studies of inquiry-based teaching: A meta-analysis and review.* European Association for Research on Learning and Instruction. Amsterdam, Netherlands.

Geier, CS., Bogner, FX. (2010). Student-centred anti-smoking education: Comparing a classroom-based versus an out-of-school setting., *Learning Environments Research*, 13(2), 147-157.

Geier, R., Blumenfeld, P. C., Marx, R. W., Krajcik, J. S., Fishman, B., Soloway, E., & Clay-Chambers, J. (2008). Standardized test outcomes for students engaged in inquiry-based science curricula in the context of urban reform. *Journal of Research in Science Teaching*, 45(8), 922-939.

Heyne, T., & Bogner, FX. (2013). Drug Prevention by Increasing Self-Esteem: Influence of Teaching Approaches and Gender on Different Consumption Groups. *World Journal of Education, 3*(1), 1-12.

Hodson, D. (2014). Learning Science, Learning about Science, Doing Science: Different goals demand different learning methods. *International Journal of Science Education, 36(*15), 2534-2553.

Jiang, F. & McComas W.F. (2015). The Effects of Inquiry Teaching on Student Science Achievement and Attitudes: Evidence from Propensity Score Analysis of PISA Data. *International Journal of Science Education, 37*(3), 554-576.

Johnson, M. A., & Lawson, A. E. (1998). What are the relative effects of reasoning ability and prior knowledge on biology achievement in expository and inquiry classes? *Journal of research in science teaching, 35*(1), 89-103.

Keeves, J. P. (1998). Methods and processes in research in science. In: B. J. Fraser & K. G. Tobin (Eds.), *International Handbook of Science Education, Part Two* (pp. 1127-1153). Dordrecht, Netherlands: Kluwer Academish Publishers.

Lederman, N. G., Antink, A., & Bartos, S. (2014). Nature of science, scientific inquiry, and socio-scientific issues arising from genetics: A pathway to developing a scientifically literate citizenry. *Science & Education, 23*(2), 285-302.

Linn, M. C, Davis, E. A. and Bell, P. (Eds), (2004). *Internet environments for science education* (p.XVi). Routledge, London: Lawrence Erlbaum Associates.

Lord, T., &Orkwiszewski, T. (2006). Moving from didactic to inquiry-based instruction in a science laboratory. *The American Biology Teacher, 68*(6), 342-345.

Lord, T.R. (1999). Are we Cultivating 'Couch Potapoes' in Our College Science Lectures? *Journal of college science teaching, 29* (1); 59-62.

Lynch, S., Kuipers, J., Pyke, C., &Szesze, M. (2005). Examining the effects of a highly rated science curriculum unit on diverse students: Results from a planning grant. *Journal of Research in Science Teaching, 42*(8), 912-946.

Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., Fishman, B., Soloway, E., Geier, R., & Tal, R. T. (2004). Inquiry-based science in the middle grades: Assessment of learning in urban systemic reform. *Journal of Research in Science Teaching, 41*(10), 1063-1080.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59*(1), 14.

Mertens, D. M. (2010). *Research methods in education and psychology. Integrating diversity with quantitative and qualitative approaches*. Thousand Oaks, Calif: Sage.

Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction—what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching, 47*(4), 474-496.

Novak, J. D. (1988). Learning science and the science of learning. *Studies in Science Education, 15*(1), 77 – 101.

Nuthall, G., &Alton-Lee, A. (1995). Assessing classroom learning: How students use their knowledge and experience to answer classroom achievement test questions in science and social studies. *American Educational Research Journal, 32*(1), 185-223.

Osborne, J., & Dillon, J. (2008). *Science education in Europe: Critical reflections*. London: The Nuffield Foundation.

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., et al. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences, 13*, 337–386.

Rocard M., Csermely P., Jorde D., Lenzen D., Walberg-Henriksson H. and Hemmo V., (2007) *Science Education Now: a renewed pedagogy for the future of Europe*. European Commission, ISBN - 978-92-79-05659-8

KMK (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland) (Ed.). (2005A). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss* [Biologie education standards for middle school graduation]. Neuwied: Luchterhand.

KMK (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland) (Ed.). (2005B). *Bildungsstandards im Fach Physik für den Mittleren Schulabschluss* [Physics education standards for middle school graduation]. Neuwied: Luchterhand.

Schwab, J. J. (1962). The teaching of science as enquiry. In J. J. Schwab & P. F. Brandwein, *The teaching of science*. Cambridge, MA: Harvard University Press.

Staver, J. R., & Bay, M. (1987). Analysis of the project synthesis goal cluster orientation and inquiry emphasis of elementary science textbooks. *Journal of Research in Science Teaching, 24*, 629-643.

Stohr-Hunt, P. M. (1996). An analysis of frequency of hands-on experience and science achievement.Journal of research in Science Teaching, 33(1), 101-109.

Tamir, P. (1985). Content analysis focusing on inquiry. *Journal of Curriculum Studies*, 17(1), 87-94.

Von Secker, C. E., &Lissitz, R. W. (1999).Estimating the impact of instructional practices on student achievement in science. *Journal of research in science teaching*, *36*(10), 1110-1126.

Wilson, C. D., Taylor, J. A., Kowalski, S. M., & Carlson, J. (2010).The relative effects and equity of inquiry-based and commonplace science teaching on students' knowledge, reasoning, and argumentation. *Journal of Research in Science Teaching, 47*(3), 276-301.

Zöfel, P. (2002). *Statistik verstehen: ein Begleitbuch zur computergestützten Anwendung.* Pearson Deutschland GmbH.

Appendix

Exemplary workbook material. Topic 1 "What is sound?", part a) "Sound creation & Movement". Extracted is the part on the movement of sound. The comments added to this excerpt intend to let the reader better understand how students where actively engaged in the learning activity. Remember that all student groups worked on their own. They were not told by the teacher what to next within the three hours and could not ask for a simple yes/no-reply if their conclusion was right or wrong as the teacher would ask them back why they would think so either way and let them explain to him, instead of providing information to them. Therefore, students needed to read their booklet with attention, as skipping paragraphs and guessing answers would leave them without feedback if they guessed right or wrong. If the teacher would from time to time ask a team about why they wrote down a certain answer and they could not explain why they decided so, he would recommend them to read the text again. Students darting to ask "is this the right answer?" to get through the material faster soon understood that if they would not know how to explain their answer they would not get a satisfying feedback but were asked to invest effort into the activity. Students mostly changed from asking for a right/wrong-feedback to asking if their way of understanding was correct or stopped asking the teacher for help but instead discussed within their team.

…

3. **Do air molecules and sound travel the same distance?**

After the air particles at the sound source are set in vibration by the vibration of the sound source, the air particles next to them vibrate as well and these bring their neighbors to vibrate as well, and so forth. The sound is moving in a wave pattern through the air. Scientists call it sound wave.

Comment: Before students made assumptions on how the sound might be moving from the sound source by plugging a rubber band which was stretched over a small box. The short paragraph above summarizes the broad theory for sound movement through a sound wave.

Question:

But how does the sound cross the broad distance from the sound source to your ear? We will compare the sound with something we already know to understand that.

Comment: Here students are introduced to a new and more explicit question about sound movement. Before introducing another experiment, students read about a phenomenon with similar function. The Mexican wave serves as a functional model of how a wave is moving, although the "particles" the wave consists of, do only oscillate around their starting position. Students sometimes have the misunderstanding that air particles travel all the way from the sound source to the recipient. Therefore students' focus is set to analyze what happens to a single air particle when a sound wave is created in the following experiment afterwards.

3.1 Comparison: The Mexican wave

A sound wave behaves like a Mexican Wave: the wave moves forward and travels big distances while the air particles that move when the wave passes only move around their original position. People only move their arms to create the Mexican Wave, they do not go anywhere themselves.

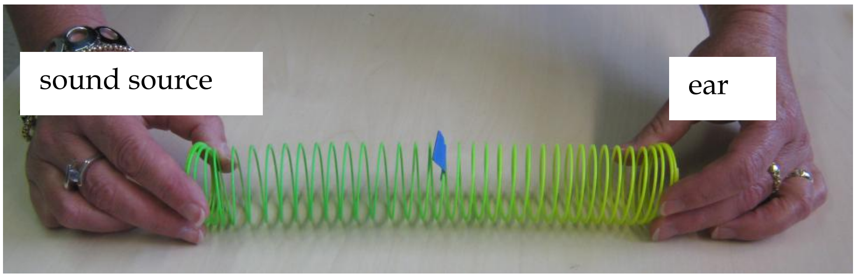
Figure 1: Fans do a Mexican wave at a sports game.

Photo not part of display

3.2 Experiment 2: The spring

Imagine the spring as the air between the sound source and your ear. Each ring of the spring is an air particle. One air particle was marked by us. The sound wave moves rythmically when there is a sound, and it stops when there is silence.

Question: You still want to find out how the sound does travel from the sound source to your ear.



Excecution:
- Place the slinky with the marked air particle on the table.
- Stretch the slinky. Hold both ends of the slinky (sound source and sound receiver)
- Strongly push one end of the slinky rhythmically in the direction of the other hand, and pull it back to its starting point. Repeat the movement.

Observations:

1. Does the marked part of the spring travel all the way from the sound source to the sound receiver side?
   ____
   ____

2. When you stop the movement, where is the marking positioned in the spring?
   ____
   ____

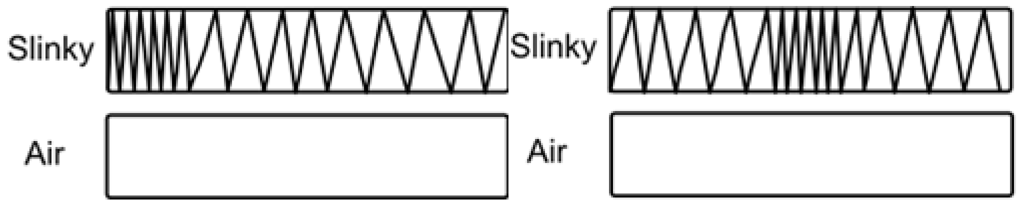
3. Each air particles in a sound wave behaves like the marked point in the slinky when a sound wave is moving across a room.
   Is the following statement right or wrong? Give reasons!
   "Each air particle of a sound wave moves through the whole room, from the sound source to your ear."
   ____
   ____
   ____

Comment: In the next passage students need to transfer their understanding of the traveling sound wave, which they gained through the experiment to the questions below. They thereby learn about the density of the air particles at certain time-points in the sound wave.

3.3 Sound waves

Do you see the similarity between the wave in the spring, the movement of the Mexican wave and the movement of a sound wave through the air? Let us exploit these similarities!

1. Draw the denseness of air particles (use little dots) in the tube under the corresponding pattern of the moving slinky. Both, the compression of the spring and of the air tell you how the sound wave moves.



2. Try to make sense of the following text by filling in the missing words, and choosing one of words in brackets by crossing out the other.
   "A sound wave is actually a wave of ____________ and ____________ of air particles in a certain part of the air. Scientists describe it as well as pressure wave. In the zones where the air is compressed the air pressure is therefore (higher/lower) as the normal background air pressure. In the zones where the air is rarefied the air pressure is (higher/lower) as the normal background air pressure."

…