Fostering Scientific Reasoning: A Meta-analysis on Intervention Studies

Katharina Engelmann, Frank Fischer, Ludwig-Maximilians-Universität München, Leopoldstr. 13, 80802 München, Germany
k.f.engelmann@psy.lmu.de, frank.fischer@psy.lmu.de

Abstract: Pedagogical intervention for scientific reasoning is a highly relevant topic in science education and outside the classroom. A systematic analysis of the success of interventions on scientific reasoning is still missing, leaving unanswered questions regarding the magnitude of the effect of interventions for scientific reasoning; and which factors in the intervention and the assessment explain differences between studies. Effect sizes taken from 15 studies were included in a meta-analysis. The results revealed a large effect of interventions on scientific reasoning (g = 0.80). Moderator analyses included learning activities and, surprisingly, showed that constructive activities yielded larger effects than interactive ones. The meta-analysis is limited by the number of studies included. Nevertheless, the results show that scientific reasoning can be fostered, though the success of the intervention depends on variables in its content, pedagogy, and assessment.

Objectives and Purpose

Scientific reasoning has become a prioritized topic in science education. Science in school aims to teach more that scientific knowledge; it aims to develop a scientific way of reasoning (Zimmerman, 2000). However, scientific reasoning does not only play a role in science education. A lot of information outside the classroom has been generated in scientific research. It is possible to process information derived from science without being able to reason scientifically. Nevertheless, an understanding of the processes and concepts of science; the skill to apply this knowledge; and the ability to reason scientifically enables people to understand scientific information within its context, including the assumptions and limitations that derive from its origin (Giere, 1979). Fostering scientific reasoning is therefore not only relevant for science education; it is also relevant for enabling people to participate in a world in which everyone is surrounded by science-based information.

Zimmerman (2007, p. 215) states that the "…issue of the best way to assess the effectiveness of instructional interventions [for scientific reasoning] will be the next issue in need of resolution". This meta-analysis presents a first approach to an exploration of the effects of interventions on scientific reasoning. Furthermore, it analyzes to what extent factors in the intervention and in the measurement of scientific reasoning can be identified to moderate the effect of the interventions.

Theoretical Framework

Conceptualizations of Scientific Reasoning

There are several different conceptions of scientific reasoning; however, as diSessa noted, any effort toward a closed set definition of scientific reasoning is an "…elusive and likely chimerical goal" (diSessa, 2008, p. 560). Heuristically, conceptualizations of scientific reasoning can be differentiated into approaches that focus on scientific reasoning as a process in scientific inquiry (e.g. Klahr and Dunbar, 1988; Lawson, 1995); approaches that focus on scientific argumentation (e.g. Kuhn, 1993, 2010); and approaches that focus on an understanding of the principles of science (Giere, 1979).

One of the most frequently cited approaches to scientific reasoning (according to *Google scholar*) is the conceptualization by Klahr and Dunbar (1988) who understand scientific reasoning to be a scientific discovery that is conducted through a dual search process in an hypothesis space and an experiment space. The model includes three main components: the *search in the hypothesis space*, during which an hypothesis is evoked by prior knowledge or induced by observations from experiments; *test hypothesis*, during which experimentation is used to evaluate a specific hypothesis; and *evaluate evidence*, during which it is analyzed whether all the results that are produced regarding an hypothesis allow its rejection or acceptance of the hypothesis. In addition, further models of scientific reasoning were developed that focus on procedural aspects e.g. in science education (Lawson, 1995).

Argumentation is considered to be "…common in science" (Osborne, 2010, p. 463) and also serves as a pedagogical method to facilitate scientific reasoning (Osborne, 2010). Kuhn (1993) connected scientific reasoning to reasoning as an argument (1991). She understands argumentation to be either rhetorical or dialogic. A dialogic argumentation contains a minimum of two people with different views, who join a dialog in which everyone offers justifications for their own view and counterarguments opposing the other view. A rhetorical argument contains a juxtaposition of two opposite claims and a reasoning process in which the truth or falsity of the claims is considered. Both forms of argument share a similar form of argumentative reasoning: at least one person notices the opposition between different claims and evidence is used in the following dialog or reasoning process to support or challenge claims. Each of the individuals involved in the dialog or reasoning process stays open with regard to the possibility that the original claim might be wrong, while allowing new evidence to have an impact on the evaluation of the assertion without dominating the reasoning process. In most cases, it appears that claims are not completely correct. In an ideal argument the evidence would be therefore "…weighted in an integrative evaluation" (Kuhn, 1991, p. 12) towards an integrative resolution. Therefore, scientific argumentation could be conceptualized as a rhetorical or dialogic argument about science, scientific constructs, and/or within a scientific context.

The nature of science contains a number of principles that are characteristic of the process of science. The American Association for the Advancement of Science (1989) understands the nature of science to be a particular way of observing, experimenting, validating, and thinking. Knowledge about the nature of science contains an understanding of scientific methods, the nature of scientific reasoning, and also a set of beliefs and attitudes about the world that serves as a foundation of science. Scientific reasoning, as an understanding and application of the nature of science, focuses on knowledge about scientific statements and their justification and arguments; the role of theories; statistical methods; the difference between causes and correlations; and values and decisions in science. While this description consists of mainly conceptual knowledge, understanding of the nature of science also includes the components of applying this knowledge within the scientific context; moreover, it also includes the application outside the scientific context in everyday situations (Giere, 1979). Here, the link between scientific literacy and scientific reasoning becomes apparent. Scientific literacy is considered to be an "…understanding of science concepts and processes with the assumption that such understanding would lead to an informed citizenry able to enact their knowledge in personal and societal issues" (Cavagnetto, 2010, p. 336).

The description of the differentiation between processes of scientific reasoning, scientific argumentation, and understanding of the nature of science includes a wide range of knowledge and skills. Therefore, it seems reasonable to adapt this differentiation to an analysis of differences between the contents of scientific reasoning in interventions that aim to foster scientific reasoning and test that measures the success of those interventions.

Fostering Scientific Reasoning

Empirical studies have begun to investigate the effect of interventions on scientific reasoning (e.g. Duncan & Arthurs, 2012); scientific inquiry (e.g. Gutwill & Allen, 2012); and scientific argumentation (e.g. Stark, Puhl, & Krause, 2009). Successful interventions were reported concerning children within a school context (e.g. Kuhn & Dean Jr., 2005), children outside a school context, e.g. in a museum (e.g. Gutwill & Allen, 2012), and also concerning adults in higher education (e.g. Duncan & Arthurs, 2012).

An analysis of pedagogical approaches to foster scientific reasoning can be conducted from the perspective of the intervention in which scientific reasoning is fostered and from the perspective of the assessment.

From the perspective of interventions, it could be hypothesized that factors that influence learning in general also influence the facilitation of scientific reasoning. The first factor is the differentiation in content of scientific reasoning as described in the conceptualization of scientific reasoning. Additionally, different types of knowledge could be targeted within each content of scientific reasoning. Moreover, Mayer (2012) suggests analyzing learning in terms of the knowledge type that is targeted in an intervention. He differentiates between *facts*, *concepts*, *processes*, *strategies*, and *beliefs*.

Apart from the content of the intervention, it can be assumed that pedagogical methods influence the success of an intervention. Chi (2009) suggests differentiating between the activities that learners undertake. She describes a framework (ICAP) distinguishing *active*, *constructive*, and *interactive activities*. *Active* activities include all activities during which the learner does something physically. *Constructive* activities require learners to produce something that goes beyond the information in the learning environment. *Interactive* activities can be differentiated into *instructional dialogs*, in which learner interacts with an expert or teacher, and peer dialogs, in which learners refer to each other and build on each other's contributions. In addition, it can be hypothesized that the technological support and the way in which the technology is supporting learners during the intervention influences the success of the intervention. Although the technological support became an important variable in the design of learning environments, a main conclusion from research on technology-enhanced learning is that the effects of technology are rarely main effects but rather interaction effects of technology with the pedagogical approach in which it is used. In recent years, many technology-supported interventions have been developed using a constructivist perspective by engaging learners in authentic activities, providing scaffolds for self-regulation and meta-cognition and encouraging collaboration (Rosen & Salomon, 2007).

From the perspective of assessment, the transfer distance that is demanded in the post-test could be understood to be represented by the degree to which the knowledge type included in the post-test was already addressed in the intervention (e.g. Barnett and Ceci, 2002). It can be hypothesized that the transfer distance is negatively related to the size of the effect or, in other words: the effect of an intervention is larger if what is measured in the post-test is more similar to what has been facilitated during the intervention.

Research Questions

The scientific community – especially in developmental psychology and science education - has been interested in scientific reasoning and related constructs (Zimmerman, 2000). Moreover, there seems to be an interest in fostering scientific reasoning skills across different content areas. Even though several single studies yielded positive effects of interventions on scientific reasoning, a systematic analysis of the success of interventions for scientific reasoning and possible moderating factors is missing. Interventions use different pedagogical approaches and aim to facilitate different aspects of scientific reasoning. Interventions and post-tests differ across studies in the content of scientific reasoning, the knowledge type, the technological support, learning activities, and the degree to which the knowledge type included in the post-test was already addressed in the intervention; consequently, it could be hypothesized that these differences explain parts of the variability between the effects of interventions on scientific reasoning.

Research question 1: What is the magnitude of the effect of interventions on scientific reasoning? What is the variability of the effects across intervention studies?

Research question 2: If there is variability of effect sizes across studies, to what extent do the content of scientific reasoning, the knowledge type, the technological support, and the type of learning activities included *in the intervention* explain the variability of the effects between studies?

Research question 3: If there is variability across studies, to what extent do the content of scientific reasoning, the knowledge type, and the transfer with respect to different knowledge types *in the post-test* explain the variability between studies?

Method

Literature search and selection

A systematic literature search was conducted to identify and retrieve publications concerning the facilitation of scientific reasoning. The literature search was conducted in the databases PsyINFO and ERIC using the search terms *scientific reasoning*, *scientific thinking*, *scientific discovery*, *scientific inquiry*, and *scientific argumentation*, which resulted in 2722 papers. The search was conducted in March 2013. The results were restricted to literature that contained at least one of the five search terms in the title in order to reduce the number of findings and eliminate mostly irrelevant literature, which resulted in 664 studies.

The studies included in the meta-analysis were selected using the following criteria: (a) empirical publication in a scientific journal or book, (b) in English or German, (c) published within the period from 01.01.1988 through 31.12.2012, and inclusion of a report of (d) an intervention and (e) at least one between-group comparison in a post-test separate from the intervention. Some studies did not report all the necessary information needed to conduct a meta-analysis. In these cases, the authors were contacted via email. In case they did not respond, whenever possible the existing data were used if sufficient to enable inclusion of the publication in the meta-analysis, otherwise the publication was eliminated. The selection of literature resulted in 15 studies to be included in the meta-analysis.

Literature coding

Coding schemes were developed in order to analyze the content of scientific reasoning in interventions and post-tests; the knowledge type in interventions and post-tests; the technological support in interventions; and the learning activities in interventions.

An intervention was operationalized as a difference in treatment between an experimental and a control group. Three studies included more than one intervention. Abdullah and Shariff (2008) included two experimental groups, of which only one (the HACL condition) was included in the meta-analysis because the description of the study suggested that this condition was the author's target condition with respect to the intervention. Gutwill and Allen (2012) included two experimental groups, of which the Juicy Question group was chosen for the same reason. Zion, Michalsky, and Mevarech (2005) had three experimental groups, of which the most inclusive intervention condition was chosen (the condition that included the attributes of the two other experimental conditions). Furthermore, two studies had more than one control group. In both cases, the name of the control group led our decision on inclusion in the meta-analysis (i.e. "pure control" in Gutwill & Allen, 2012 and "main control" in Kuhn & Dean Jr., 2005).

Content of scientific reasoning: The coding scheme for the content of scientific reasoning was derived from the conceptualizations of scientific reasoning and differentiated between the *processes of scientific reasoning*, *scientific argumentation*, and *understanding science*. Intervention that fostered *processes of scientific reasoning* included scientific processes such as the deduction or generation of hypothesis, or the generation or evaluation of evidence. Interventions that fostered *scientific argumentation* included an argumentation about science or scientific constructs. Interventions that fostered *understanding science* included knowledge about the principles, concepts, assumptions, and limitations of science and its application. One code was given to each intervention.

Knowledge type: The coding scheme for the knowledge type differentiated between *facts*, *concepts*, *processes*, *strategies*, and *beliefs* (Mayer, 2012) with respect to scientific reasoning. One or more codes were given to each intervention representing the knowledge types that were included in the intervention. In order to adequately analyze the data, the interventions were classified into (a) one group of interventions that included facts, concepts, or both (b) one group that included processes, strategies, or both and (c) a third group that included a combination of at least one item from each of the two prior groups.

Technological support: The coding scheme for the technological support differentiated between interventions that included technological support from a constructivist perspective, operationalized as creating an authentic learning environment, supporting cooperation, or supporting self-regulation; technological support that did not support any of these three aspects of a constructivist approach; and interventions that did not include any technological support. One code was given to each intervention.

Learning activities: The coding scheme for the learning activities (Chi, 2009) differentiated between *active*, *constructive*, interactive in an *instructional dialog*, and interactive in a *peer dialog*. One code was given to each intervention representing the most dominant activity.

All coding schema also included one residual category in case the description did not give enough information to determine which code was true. These data were treated as missing data in the moderator analysis.

The measurement of scientific reasoning was operationalized as the post-test(s) conducted after the pedagogical intervention. The coding scheme for *content of scientific reasoning* and *knowledge type* in the post-test was similar to the coding scheme for the intervention.

Transfer with respect to different knowledge types: The transfer was calculated by categorizing each study in respect to the distance between what was measured in the post-test to what was facilitated during the intervention. We differentiated between post-tests that included *no transfer* and posttest that included *transfer*. No transfer was assumed in cases where the post-test included only knowledge types that were already part of the intervention. Transfer was assumed in cases were the post-test also (or only) included knowledge types that were not part of the intervention.

Statistical analysis

The meta-analysis was conducted following the procedure suggested by Lipsey and Wilson (2001) for the fixed model. Wherever possible, the descriptive data were used to calculate the effect size (Hedges' g) by using the arithmetic means of experimental (X_G1) and control (X_G2) groups; the pooled standard deviation (σ_pooled); and the samples size for each group (n1 and n2): g=(XG1−XG2)/(σ_pooled), (σ_pooled)=√(σ12⍰(n1−1)+σ22 ⍰(n2−1))/(n1+n2−2).

Alternatively, The results of the inferential statistics (*t* or *F* values) were used to calculate the effect size using these formulas: g=t*√((n1+n2)/n1⍰n2) or g=√(F*(n1 + n2)/n1⍰n2).

Most studies reported one effect in respect to scientific reasoning and this effect was used in the meta-analysis. The arithmetic mean of the effect sizes was calculated for those studies that reported more than one outcome, except for one study (Duncan & Arthurs, 2012) who reported two outcome measures based on different but possibly not independent samples. In this case, the more reasonable effect size was chosen.

The specific parameters of the meta-analysis and the moderator analysis were calculated using the meta-analysis macros for SPSS from Wilson (2005). The effects were corrected for small sample bias and weighted using the inverse variance weight. The results were computed by calculating the mean of the effects and the confidence interval (CI). Furthermore, the homogeneity (Q) of the effects was tested.

Results

The first research question was: What is the magnitude of the effect of interventions on scientific reasoning? What is the variability of the effects across intervention studies?

The meta-analysis revealed a highly significant effect, suggesting a large mean effect size ($g = 0.80$, $CI_{95\%}$ [0.70, 0.90], $p < 0.01$). Figure 1 gives an overview of the effects included in the meta-analysis. The analysis of homogeneity was highly significant (Q (14) = 96.34, $p < 0.01$), showing that the sample is heterogeneous. The magnitude of the effects of interventions for scientific reasoning is high; however, the results show a high variability of the effects across intervention studies.

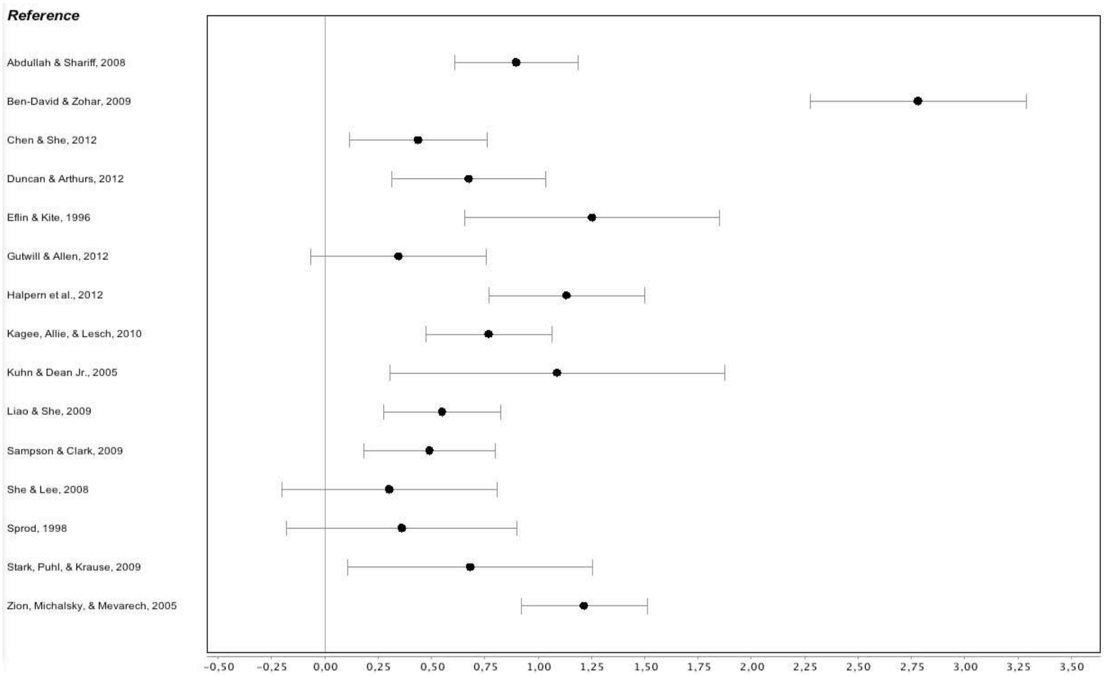

Figure 1. Forest plot of the effects included in the meta-analysis, indicating mean effect size and CI.

The second research question was: If there is variability of effect sizes across studies, to what extent do the content of scientific reasoning, the knowledge type, the technological support, and the type of learning activities included *in the intervention* explain the variability of the effects between studies?

The descriptive results of the subgroup comparison for the moderator variables in the intervention are shown in Table 1. The result of the subgroup comparison for the content of scientific reasoning revealed a highly significant difference between the groups ($Q(2) = 14.02$, $p < 0.01$). Interventions targeting the processes of scientific reasoning yielded larger effects than interventions targeting understanding science. Interventions on scientific argumentation yielded the lowest effects. The result of the subgroup comparison for the knowledge type also revealed a highly significant difference between the groups ($Q(2) = 32.01$, $p < 0.01$). The largest effects were found in interventions that included the combination of knowledge types; the lowest effect in interventions that included processes and strategies. The result of the subgroup comparison for the technological support revealed no significant difference between presence or absence of technological support ($Q(1) = 0.21$, $p > 0.05$). Furthermore, no significant difference was found between the groups after adding the differentiation between technology in the context of constructivist pedagogy and technology in the context of other pedagogies ($Q(2) = .37$, $p > 0.05$). The results of the subgroup comparison for the learning activities revealed a highly significant difference between constructive and interactive activities ($Q(1) = 17.68$, $p < 0.01$) and also highly significant differences after including the differentiation between instructional and peer dialog into the moderator analysis ($Q(2) = 18.77$, $p < 0.01$). interventions utilizing constructive activities yielded larger effects than interventions utilizing interactive activities. Moreover, the addition of the differentiation between peer dialog and instructional dialog increased the amount of explained variability between the studies.

Table 1: Descriptive results of the moderator analysis of the intervention

	Number of studies	g	Lower bound of the CI	Upper bound of the CI
Content of scientific reasoning				
Processes of scientific reasoning	3	1.28	0.99	1.58
Scientific argumentation	6	0.65	0.50	0.81
Understanding Science	3	0.84	0.65	1.04
Knowledge type				
Facts and concepts	5	0.73	0.56	0.90
Processes and strategies	5	0.46	0.29	0.64
Combination	4	1.20	1.01	1.38
No technological support	*6*	*0.83*	*0.67*	*0.99*
Technological support	*9*	*0.78*	*0.66*	*0.90*
Constructivist pedagogy	6	0.80	0.66	0.93
Other pedagogies	3	0.73	0.45	1.02

	Number of studies	g	Lower bound of the CI	Upper bound of the CI
Learning activity				
Active	0			
Constructive	5	1.09	0.93	1.25
Interactive	9	0.66	0.53	0.78
Interactive: instructional dialog	4	0.74	0.54	0.94
Interactive: peer dialog	5	0.60	0.44	0.76

The third research question was*:* If there is variability across studies, to what extent do the content of scientific reasoning, the knowledge type, and the transfer with respect to different knowledge types *in the post-test* explain the variability between studies?

The descriptive results of the subgroup comparison for the moderator variables in the post-test are shown in Table 2. The result of the subgroup comparison for the content of scientific reasoning revealed a highly significant difference between the groups ($Q(2) = 23.28$, $p < 0.01$). The inclusion of processes of scientific reasoning in post-tests yielded higher effects than understanding science. Post-test measuring scientific argumentation yielded the lowest effects. The result of the subgroup comparison for the knowledge type also revealed a highly significant difference between the groups ($Q(2) = 10.53$, $p < 0.01$). The largest effects were found in post-tests that measured facts and concepts, the lowest effect was found in post-tests that measured a combination of knowledge types. The result of the subgroup comparison for the transfer with respect to different knowledge types also revealed a highly significant difference between the groups ($Q(1) = 15.75$, $p < 0.01$). Post-tests that included no transfer yielded the largest effects than post-tests that included transfer.

Table 2: Descriptive results of the moderator analysis of the post-test

	Number of studies	g	Lower bound of the CI	Upper bound of the CI
Content of scientific reasoning				
Processes of scientific reasoning	4	1.25	1.04	1.46
Scientific argumentation	3	0.56	0.35	0.77
Understanding Science	6	0.75	0.61	0.88
Knowledge type				
Facts and concepts	5	1.03	0.85	1.20
Processes and strategies	4	0.78	0.59	0.96
Combination	4	0.61	0.43	0.79
Transfer with respect to different knowledge types				
No transfer	10	0.94	0.82	1.06
Transfer	3	0.47	0.28	0.67

Discussion

This meta-analysis carried out a systematic examination of interventions that aimed to foster scientific reasoning and found a large mean effect, suggesting that scientific reasoning can be successfully fostered by interventions. Furthermore, some variability between the studies may be explained by variables in the intervention and variables in the post-test of scientific reasoning.

Studies in which the interventions focused on processes of scientific reasoning such as hypothesis generating, experimenting, and evidence evaluation showed a larger mean effect than studies that focused on an understanding of science; studies aimed at facilitating scientific argumentation had the smallest effect size. The same pattern was found in the post-test of scientific reasoning. Measuring processes of scientific reasoning reached a larger mean effect than than measuring an understanding of science, while measuring scientific argumentation yielded the smallest effect size. The knowledge type also played an important role in intervention and post-test. The largest mean effect was achieved in studies that included a combination of knowledge types in the intervention. Regarding the post-test of scientific reasoning, the largest effects were achieved when only facts and concepts were measured.

The effects of the moderator analyses have to be cautiously interpreted, however, because the tests of significance refer to subgroup comparisons and do not provide a test of significance for comparisons between single variables. The moderator analysis shows, for instance, a significant difference between contents of scientific reasoning in interventions. However, it does not provide a test of significance between *processes of scientific reasoning* and *understanding science*.

The results of the moderator analysis with respect to technological support in the learning environment do not support the assumption that technology can increase the effect size, even when combined with a constructivist pedagogy (Rosen & Salomon, 2007).

The differentiation in learning activities between interventions revealed a significant difference, validating to an extent the distinction put forward in the ICAP hypothesis by Chi (2009). However, the assumption that *interactive* activities are more effective than *constructive* activities is not supported in the context of intervention studies on scientific reasoning. The larger effect of constructive, in comparison with interactive, activities might be explained by the topic of the intervention. It would be reasonable to assume that interactive activities, especially with peers, provide both aspects that support the learning process and aspects that are detrimental. The evidence provided by Chi (2009), supporting the hypothesis that interactive are more beneficial than constructive activities, compares situations only in which the interactive activity is similar to the constructive one, except for the presence of another individual. In this comparison, further aspects are neglected that occur in interactive activities at times; add to the *cost* side; and might negatively affect the learning process; such as time that is spent on coordination and eristic arguments. Scientific reasoning is a complex target for intervention; thus, the *cost* side of interactive activities might have become more influential. If this explanation is valid, ICAP needs to be differentiated with respect to potential collaboration costs that might be higher in more complex reasoning tasks without additional guidance. Future analyses could more closely examine the interaction process itself to test this modification of ICAP. In addition, future analyses should try to include the *active* category of activities in learning environments for scientific reasoning and argumentation, to comprehensively test the validity of ICAP in this context.

The results of the moderator analysis concerning the transfer suggest that the range of transfer between intervention and post-test explains some variability between the studies. The absence of transfer yielded larger effects than the presence transfer. This result is coherent with conceptualizations of transfer (e.g. Barnett and Ceci, 2002) and could additionally be interpreted as a validation of the sample of studies that was included in the meta-analysis.

Even though most results of the meta-analysis are highly significant, interpretations of the results have to be made cautiously. The main limitations of the meta-analysis directly result from the selection of studies. As Eisend (2004) points out, the research and publication process favors studies that report significant results, leading to a bias in the published studies. Furthermore, the sample included in the meta-analysis was limited to studies that included scientific reasoning and related search terms in the title, which might have enhanced a bias in favour of studies that included an successful intervention for scientific reasoning.

This meta-analysis provides a first overview of intervention studies concerning scientific reasoning and shows a large effect of intervention on scientific reasoning. Furthermore, we were able to identify moderator variables in the content, pedagogy, and assessment. Interventions for scientific reasoning which engage learners in constructive activities are more successful than interventions which engage learners in interactive activities; this result validates the distinction made by Chi (2009) but, at the same time, disconfirms the order of the activities. Furthermore, the type of content and knowledge fostered in the intervention, and measured in the post-test, influences the success of the intervention. Further research is needed to test directly the effects found in this meta-analysis.

References

References marked with an asterisk (*) indicate studies included in the meta-analysis.

*Abdullah, S., & Shariff, A. (2008). The effects of inquiry-based computer simulation with cooperative learning on scientific thinking and conceptual understanding of gas laws. *EURASIA Journal of Mathematics, Science & Technology Education, 4*(4), 387-398.

American Association for the Advancement of Science. (1989). *Project 2061- Science for all Americans*. Washington- AAAS.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*(4), 612-637.

*Ben-David, A., & Zohar, A. (2009). Contribution of meta-strategic knowledge to scientific inquiry learning. *International Journal of Science Education, 31*(12), 1657-1682. .

Cavagnetto, A. R. (2010). Argument to foster scientific literacy: A review of argument interventions in K-12 science contexts. *Review of Educational Research*, *80*(3), 336-371.

*Chen, C.-H., & She, H.-C. (2012). The impact of recurrent on-line synchronous scientific argumentation on students' argumentation and conceptual change. *Educational Technology & Society, 15*(1), 197-210.

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*(1), 73-105.

diSessa, A. A. (2008). A "theory bite" on the meaning of scientific inquiry: A companion to Kuhn and Pease. *Cognition and Instruction*, *26*(4), 560-566.

*Duncan, D. K., & Arthurs, L. (2012). Improving student attitudes about learning science and student scientific reasoning skills. *Astronomy Education Review,* 11(1), 010102–1–010102–11.
*Eflin, J. T., & Kite, M. E. (1996). Teaching scientific reasoning through attribution theory. *Teaching of Psychology, 23*(2), 87.
Eisend, M. (2004). *Meta-analyse – Einführung und kritische diskussion*, Discussion Paper of the Economics Department of the Free University Berlin, No. 2004/8.
Giere, R. N. (1979). *Understanding Scientific Reasoning*. New York,NY: Holt, Rinehart & Winston.
*Gutwill, J. P., & Allen, S. (2012). Deepening students' scientific inquiry skills during a science museum field trip. *Journal of the Learning Sciences*, *21*(1), 130-181.
*Halpern, D. F., Millis, K., Graesser, A. C., Butler, H., Forsyth, C., & Cai, Z. (2012). Operation ARA: A computerized learning game that teaches critical thinking and scientific reasoning. *Thinking Skills and Creativity, 7*(2), 93-100.
*Kagee, A., Allie, S., & Lesch, A. (2010). Effect of a course in research methods on scientific thinking among psychology students. *South African Journal of Psychology, 40*(3), 272-281.
Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive science*, 48, 1-48.
Kuhn, D. (1991). *The skill of argument.* New York,NY: Cambridge University Press.
Kuhn, D. (1993). Science as argument: implications for teaching and learning scientific thinking. *Science Education, 77*(3), 310-337.
Kuhn, D. (2010). Teaching and learning science as argument. *Science Education*, *94*(5), 810-824..
*Kuhn, D., & Dean Jr., D. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science (Wiley-Blackwell), 16*(11), 866-870.
Lawson, A. E. (1995). *Science teaching and the development of thinking.* Belmont, CA: Wadsworth.
*Liao, Y.-W., & She, H.-C. (2009). Enhancing eight grade students' scientific conceptual change and scientific reasoning through a web-based learning program. *Educational Technology & Society, 12*(4), 228-240.
Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks,CA: Sage.
Mayer, R. E. (2012). Information processing. In K. R. Harris, S. Graham, T. Urdan, C. B. McCormick, G. M. Sinatra, J. Sweller (Eds.), *APA educational psychology handbook, Vol 1: Theories, constructs, and critical issues* (pp. 85-99). Washington, DC: American Psychological Association.
Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, *328*(April), 463-467.
Rosen, Y., & Salomon, G. (2007). The differential learning achievements of constructivist technology -intensive learning environments as compared with traditional ones: A meta-analysis. *Journal of Educational Computing Research, 36*(1), 1-14.
*Sampson, V., & Clark, D. (2009). The impact of collaboration on the outcomes of scientific argumentation. *Science Education, 93*(3), 448-484.
*She, H.-C., & Lee, C. Q. (2008). SCCR digital learning system for scientific conceptual change and scientific reasoning. *Computers & Education, 51*(2), 724-742.
*Sprod, T. (1998). "I can change your opinion on that": Social constructivist whole class discussions and their effect on scientific reasoning. *Research in Science Education, 28*(4), 463-480.
*Stark, R., Puhl, T., & Krause, U.-M. (2009). Improving scientific argumentation skills by a problem-based learning environment: effects of an elaboration tool and relevance of student characteristics. *Evaluation & Research in Education, 22*(1), 51-68.
Wilson, D. B. (2005). *Meta-analysis macros for SAS, SPSS, and Stata*. Retrieved, June, 16, 2013, from http://mason.gmu.edu/~dwilsonb/ma.html
Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review*, *20*(1), 99-149.
Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review, 27*(2), 172-223.
*Zion, M., Michalsky, T., & Mevarech, Z. R. (2005). The effects of meta-cognitive instruction embedded within an asynchronous learning network on scientific inquiry skills. Research Report. *International Journal of Science Education, 27*(8), 957-983.