The Role of Argumentation for Domain-Specific Knowledge Gains in Computer-Supported Collaborative Learning: A Meta-Analysis

Christof Wecker, Frank Fischer, Ludwig-Maximilians-Universität München, Leopoldstr. 13, 80802 München, Germany
Email: christof.wecker@psy.lmu.de, frank.fischer@psy.lmu.de

Abstract: The meta-analysis reported in this paper investigated the role of the quality of argumentation for domain-specific knowledge gains in computer-supported collaborative learning settings. Given the scarcity of primary studies that report correlations between these two variables, a meta-regression approach was used that uses interventions' effects on argumentation to predict their effects on domain-specific knowledge. Effect sizes for 17 comparisons extracted from 12 studies were included in the analysis using a random-effects model. On average, the interventions have a small to moderate effect on argumentation. With respect to the relation of their effects on argumentation to their effects on domain-specific knowledge, no unequivocal picture emerges. These findings call into question the broadly shared theoretical assumption that argumentation can be a mechanism that mediates the effects of interventions on domain-specific knowledge. A set of recommendations for strengthening future research on the topic is presented.

Introduction

It is a broadly shared conviction of CSCL researchers that argumentation can be a powerful mechanism to foster domain-specific knowledge (Andriessen, 2006; Osborne, 2010). Based on this assumption, a broad variety of tools and interventions to foster argumentation has been developed, ranging from direct instruction about characteristics of good argumentation (e. g. Choresh, Mevarech & Frank, 2009; Nussbaum, Sinatra & Poliquin, 2008; Yeh, K. H. & She, 2010) to argumentation maps (e. g. Janssen, Erkens, Kirschner & Kanselaar, 2010; Munneke, van Amelsvoort & Andriessen, 2003; van Drie, van Boxtel, Jaspers & Kanselaar, 2005; van Drie, van Boxtel, Erkens & Kanselaar, 2005; Schwarz, Neumann, Gil, J. & Ilya, 2003) and from different kinds of discussion seeds (Clark, D'Angelo & Menekse, 2009) to collaboration scripts (Kollar, Fischer & Slotta, 2007; Stegmann, Weinberger & Fischer, 2007; Weinberger, Stegmann & Fischer, 2010). These tools have been used for a while now in order to induce interactions among learners that are characterized by high argumentative quality and to study the role of argumentation for the learning of domain-specific content. Although some of these studies employ a qualitative methodology and thereby provide detailed accounts of the mechanisms involved in learning through argumentation, also a considerable number of quantitative studies on the topic have been conducted.

Therefore, it seems to be the right time now for integrating these findings using quantitative methods of research synthesis in order to evaluate the evidence pertinent to the widespread convictions about the role of argumentation for learning. On closer inspection, however, many of the studies focus only on some of the variables that are relevant for the issue. For instance, sometimes only effects of an intervention on the quality of argumentation during interaction in a collaborative learning phase (e. g. Munneke et al., 2003) or on the acquisition of argumentation skills as measured by a posttest are measured and reported. Among the studies that measure both the quality of argumentation during collaboration and learning outcomes concerning domain-specific knowledge, only a small proportion either reports correlations between argumentation during learning and domain-specific knowledge acquisition (e. g. Choresh et al., 2009) or employs similar analytical strategies that allow for an assessment of the association between these two kinds of variables (e. g. Asterhan, 2008). Often, only correlations within some of the several experimental conditions are reported along with the information that no significant correlation was found in the other conditions (e. g. van Drie et al.,2005a).

This practice makes the conventional approach of integrating indicators of association between the two variables of interest unfeasible. As a sufficient number of the studies in CSCL research report effects of interventions on both measures of argumentation and measures of domain-specific knowledge, a different analytical approach is used in this meta-analysis: If the assumption about the role of argumentation for domain-specific knowledge acquisition described above is true, interventions that produce strong positive effects on argumentation should also produce strong positive effects on domain-specific knowledge, and interventions that produce small or even negative effects on argumentation should produce small or negative effects on domain-specific knowledge. Therefore, the technique of meta-regression is used in a somewhat unusual way to predict interventions' effects on domain-specific knowledge on the basis of their effects on argumentation in order to test the validity and generality of the importance of argumentation for domain-specific learning. This approach presupposes substantial variation in effect sizes with respect to both variables. Given the diversity of argumentation interventions used in primary studies, it seems likely that effect sizes are actually quite heterogeneous. To test whether this prerequisite is fulfilled, in addition to the covariation of effects on argumentation and domain-specific knowledge also the magnitude and variation of effects of the interventions on these variables are scrutinized.

Research Questions

This meta-analysis addressed the following research questions:

(1) What is the mean effect of argumentation interventions on argumentation, and are the effects of the single argumentation interventions homogeneous? It is expected that on average the argumentation interventions have a positive effect on argumentation. As the kinds of argumentation interventions investigated in CSCL research are quite different, systematic variation of their effects on argumentation is expected.

(2) What is the mean effect of argumentation interventions on domain-specific knowledge, and are the effects of the single argumentation interventions homogeneous? According to the assumption that argumentation mediates intervention effects on domain-specific knowledge, on average the argumentation interventions should have a positive effect on domain-specific knowledge. As in the case of effects on argumentation, systematic variation of effects on domain-specific knowledge is expected due to the heterogeneity of argumentation interventions studied in CSCL research.

(3) How are argumentation interventions' effects on argumentation related to their effects on domain-specific knowledge? According to the theoretical assumptions about the role of argumentation for domain-specific knowledge acquisition, a moderate positive relation between effects on argumentation and effects on domain-specific knowledge is to be expected.

Method

Selection of Studies

Criteria for Inclusion

The purpose of this meta-analysis is to provide generalized information about the relation of the quality of argumentation during learning to domain-specific knowledge in a setting of computer-supported collaborative learning. Therefore, the criteria for the inclusion of a study were the following:

(1a) The quality of argumentation in collaboration during the learning phase or – as a proxy indicator for this variable – the learners' acquired argumentation skills after the learning phase was measured.

(1b) The individual learners' domain-specific knowledge about the content discussed during the learning phase was measured.

(2) The study was conducted in a computer-supported collaborative learning setting.

(3) At least one effect of the intervention under investigation on the quality of argumentation or on argumentation skills on the one hand and at least one effect on individual domain-specific knowledge on the other hand as well as their associated variances can be determined based on the information provided.

Search Strategies

The approach to locate the studies that fulfill these criteria comprised several complementary strategies. First, the bibliographic databases ERIC and PsycINFO were searched. The search terms "argument" (truncated), "learning" (truncated) and "CSCL" (or a conjunction of several synonyms of "collaboration" (truncated) and "computer") were used conjunctively. These search terms were not limited to specific fields to warrant high recall at the cost of lower precision. This yielded 370 publications many of which were either unrelated to the topic or did not report empirical studies. Based on a thorough review of the abstracts of these publications, each publication for which it could not be ruled out definitively that they might fulfill the criteria for inclusion were selected for inspection. Second, relevant studies mentioned in the reference lists of studies and reviews retrieved were also included in the set of studies for integration. Finally, the digital versions of the proceedings of recent CSCL and ICLS conferences were searched electronically for occurrences of "argument" to yield further studies on the topic that have not been published in journals yet.

Sample of Studies

The sample included in this meta-analysis comprises 12 studies. The integration of effect sizes ís based on data from more than 1400 persons.

Coding of Variables

Two types of variables coded from the primary studies were included in this meta-analysis: Each study's effect on argumentation and each study's effect on domain-specific knowledge. The types of intervention used in the primary studies were also coded. As the focus of this paper is not on the effects specific or diverse kinds of interventions on the association between argumentation and domain-specific knowledge gains, characteristics of interventions and similar study features are not used in the analyses presented here.

Argumentation

All variables in a study that were either indicators of specific aspects of the quality of argumentation in collaboration during the learning phase or indicators of the learners' acquired argumentation skills after the learning phase were selected for coding. The types of variables included comprised indicators of the number of (specific types or components of) arguments as well as their argumentative or content-related quality on a micro level of analysis and the number and quality of specific argumentative speech acts on a higher level of analysis. If a study contained several types of measures of argumentation, the set of most proximal ones were selected. For example, if both indicators of argumentation skill in a posttest and argumentative quality of contributions during the learning phase were reported, only the latter were included in the analysis. If indicators of both argumentative strength and content-related accuracy of contributions during the learning phase were provided, only the former were included. The rationale behind this approach was to use the best estimates available of the effect of an intervention on the argumentative quality of contributions during the learning phase, which according to the theoretical assumptions should mediate an intervention's effect on domain-specific knowledge.

For each of these variables, an effect size and its variance were computed from the descriptive statistics (means, standard deviations and sample sizes of subgroups) provided in the study. If part of this information was insufficient, either any kind of measure of effect size – if reported in the studies – was extracted directly and transformed, if necessary, or the effect size was computed from the values of inferential test statistics.

Domain-specific Knowledge

All variables in a study that quantified individually attributable domain-specific knowledge due to the learning experience were coded. Individual attribution means that (in one case) conceptual change diagnosed for individual participants on the basis of discourse data from an online discussion was included, whereas the domain-specific quality of essays written by groups could not be used as an indicator of individual domain-specific knowledge. In the majority of cases this kind of data was extracted from posttests. Effect sizes were determined as in the case of argumentation.

Calculation and Statistical Analysis of Effect Sizes

Effect Size Metric

As most of the studies report main effects from between-subjects experimental designs, the unbiased estimator d of standardized mean differences suggested by Hedges and Olkin (1985, p. 81) was used for integrating the study effects. If pretest data were available, posttest effect sizes were corrected by subtracting the corresponding pretest effect size from the posttest effect size.

Statistically Dependent Effect Sizes

If for a study separate effect sizes could be determined for different subsamples (such as the different values of a second factor in a 2x2 design that is not relevant in the present context), the data from all subsamples were included as independent effect sizes (cf. Borenstein et al., 2009, p. 218 f.).

Other studies (typically ones that use a one-factorial design) compared several treatment groups to a common control group. In this case, effect sizes from the same study are correlated due to the values from the common control group that enters the effect sizes of all experimental groups (Borenstein et al., 2009, p. 239-241). One option to deal with this kind of dependency is to compute one common average effect size for the different intervention groups and use only this single effect size for further integration. As the purpose of the present meta-analysis is to determine the association between argumentation and domain-specific knowledge gains on the study level and averaging across intervention types would blur relevant variation and covariation, the effect sizes of the single comparisons from this type of study are used separately in the integration. We regard the partial violation of assumptions of the statistical methods employed for integration as less severe than the loss of important information we would have to suffer otherwise.

Finally, most studies report findings for a whole set of indicators of the quality of argumentation, and some also report findings for several scales of domain-specific knowledge or for multiple measurement time-points using the same instruments. In these cases, the effect size indices from each comparison were aggregated to yield one effect size for argumentation and one effect size for domain-specific knowledge per comparison. As most studies do not report correlations between the different dependent measures, we ran all analyses reported in this paper with three different global assumptions about the covariation among dependent measures ($r = 0$; $r = .3$ and $r = 1$). There were no differences between the three variants in any of the analyses that would lead to a different answer to any of our research questions. Therefore we consistently report only the results from the analyses assuming independence of multiple measurements.

These decisions resulted in the integration of 17 pairs of separate effect sizes for argumentation and for domain-specific knowledge. They were calculated on the basis of 55 original effect sizes extracted from the primary studies. For 11 comparisons we could extract effect sizes for argumentative quality of contributions, whereas for 6 comparisons we had to rely on the more distal posttest measures of argumentation skill.

Method of Integration

The primary studies included in this meta-analysis investigated the effects of a very broad array of argumentation interventions that ranged from instruction about good argumentation to the use of different kinds of discussion seeds and from argumentation maps to collaboration scripts. Therefore no unique common effect size can be assumed for these studies. Accordingly, the random-effects approach for integrating the findings was employed. The between-studies variance component was estimated using the method described by Raudenbush (1994, p. 310 f.).

Results

Research Question 1: Effects of the Argumentation Interventions on Argumentation

The individual effect sizes for argumentation and their 90 % confidence intervals as well as their summary effect are presented in figure 1.

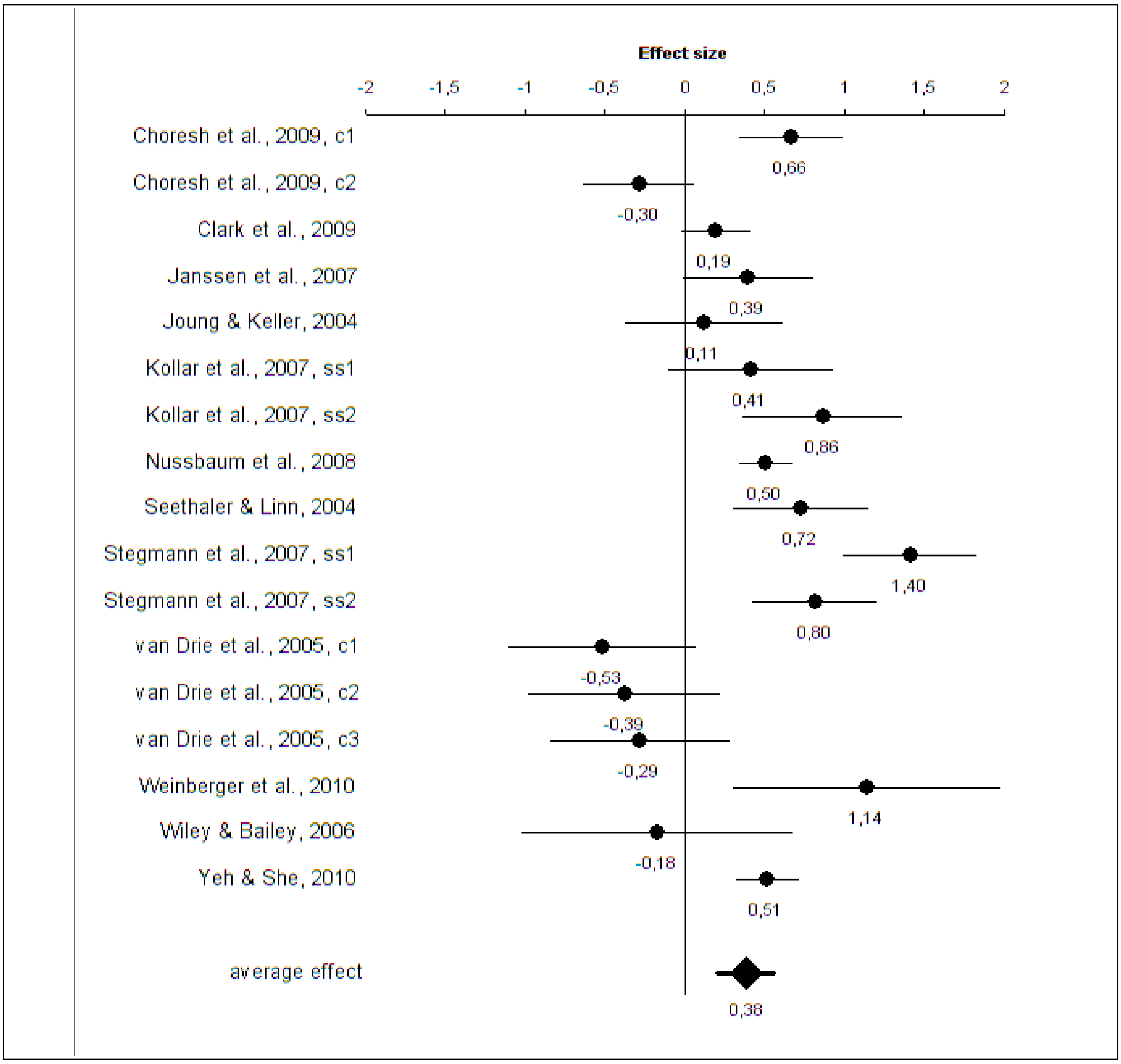

Figure 1. Forest plot of the interventions' effects on argumentation.
The error bars indicate 90 % confidence intervals. "c" along with a number refers to multiple comparisons with a common control group, "ss" along with a number refers to independent sub-samples.

A small to moderate average effect size of 0.38 was estimated, $SE = 0.11$; $CI_{90\%} = [0.20; 0.56]$; $p < .01$, one-tailed. The test for the homogeneity of the effects of the interventions on argumentation indicated variation in the true effect sizes, $Q_T(df = 16) = 58.99$; $p < .01$; $I^2 = .73$.

We further conducted a moderator analysis comparing the group of effect sizes based on measures for argumentative quality of contributions ($d = 0.32$; $SE = 0.15$; $CI_{90\%} = [0.07; 0.57]$; $p = .02$, one-tailed.) with the group based on the more distal posttest measures of argumentation skill ($d = 0.46$; $SE = 0.16$; $CI_{90\%} = [-0.19; 0.73]$; $p < .01$, one-tailed.). Effects from these two groups were homogeneous, $Q_B(df = 1) = 0.40$; $p > .99$.

These findings provide support for the validity of our approach: First, because there is substantial variation in effects on argumentation, which is required for detecting the association of interest, it is possible to investigate the role of argumentation for domain-specific knowledge gains on the level of studies. Furthermore, the inclusion of measures of post-treatment argumentation skill as more distal indicators of argumentative quality of contributions seems warranted given the homogeneity of effect sizes of both types. However, because the focus of the present analysis is not on the effects of different kinds of intervention on argumentation, no further attempt to explain the variance in effects on argumentation is made in the following.

Research Question 2: Effects of the Argumentation Interventions on Domain-Specific Knowledge

On average, argumentation interventions have no effect on domain-specific knowledge, $d = -0.04$; $SE = 0.14$; $CI_{90\%} = [-0.28; 0.19]$; $p = .62$, one-tailed. However, the test for the homogeneity of the effects of the interventions on domain-specific knowledge indicates variation in the true effect sizes, $Q_T(df = 16) = 129.46$; $p < .01$; $I^2 = .88$. To study the role of argumentation for domain-specific knowledge gains, this variation is further analyzed as a function of the interventions' effects on argumentation in the following.

It should be noted, however, that in one particular study large negative effects on domain-specific knowledge were observed (van Drie et al., 2005a). These effects were more than two standard deviations lower than the average effect size of the other studies in the sample, which means that they fulfil the frequently applied criterion for the exclusion of outliers. If the three effect sizes are excluded from the analysis, the estimate of the average effect size on domain-specific knowledge changes to $d = 0.17$; $SE = 0.13$; $CI_{90\%} = [-0.05; 0.39]$; $p = .10$, one-tailed. But also without these three effect sizes the test for the homogeneity of the effects still indicates variation in the true effect sizes, $Q_T(df = 16) = 84.74$; $p < .01$; $I^2 = .85$.

Research Question 3: Association between the Effects of the Argumentation Interventions on Argumentation and their Effects on Domain-Specific Knowledge

A two-dimensional forest plot (which is a variant of a scatterplot with some additional information) of the individual effects on argumentation and domain-specific knowledge is displayed in figure 2.

To determine the association between the interventions' effects on argumentation and on domain-specific knowledge, a meta-regression with the effects on domain-specific knowledge as the criterion and the effects on argumentation as the single predictor was conducted. A positive estimate for the unstandardized regression coefficient was obtained, $b = 0.58$; $SE = 0.31$; $CI_{90\%} = [0.06; 1.10]$; $\beta = .44$; $p = .03$, one-tailed. The higher the effects of the interventions investigated in the primary studies on argumentation, the higher were also their effects on domain-specific knowledge.

If the three comparisons from the van Drie et al. (2005a) study are excluded as outliers on the domain-specific knowledge dimension, however, an association of argumentation and domain-specific knowledge can no longer be detected, $b = -0.12$; $SE = 0.30$; $CI_{90\%} = [-0.61; 0.37]$; $\beta = -.12$; $p = .34$, one-tailed.

Discussion

The results of this meta-analysis provide evidence that on average argumentation interventions developed in CSCL research are successful with respect to their most proximal goal of enhancing argumentation. As expected, these interventions do not share one common effect size. This is an advantage for the goal of the present integration to investigate the role of argumentation for domain-specific knowledge acquisition on the study level, given the scarcity of correlations on the person level reported in primary studies. However, differential effects of different types of interventions on argumentation should also be investigated meta-analytically in the future.

This broad variety of effects on argumentation is accompanied by an average zero effect of these interventions on domain-specific knowledge on the basis of, again, varying true effects. Even if the van Drie et al. (2005a) study is removed as an outlier, no significant effect on domain-specific knowledge emerges. Given the broad consensus among CSCL researchers about the role of argumentation for learning, it is quite remarkable that on average argumentation interventions do not produce beneficial effects on domain-specific knowledge acquisition.

The picture becomes even more obscure when considering the relation between argumentation and domain-specific knowledge: Only if the van Drie et al. (2005a) study is retained in the sample of studies are the effects of the interventions on argumentation positively related to their effects on domain-specific knowledge. If this study is removed because it qualifies as an outlier with respect to domain-specific knowledge, no such relation can be detected. On inspection, however, the kinds of intervention investigated in this study do not stand out from the ones used in other studies in the field in any obvious way.

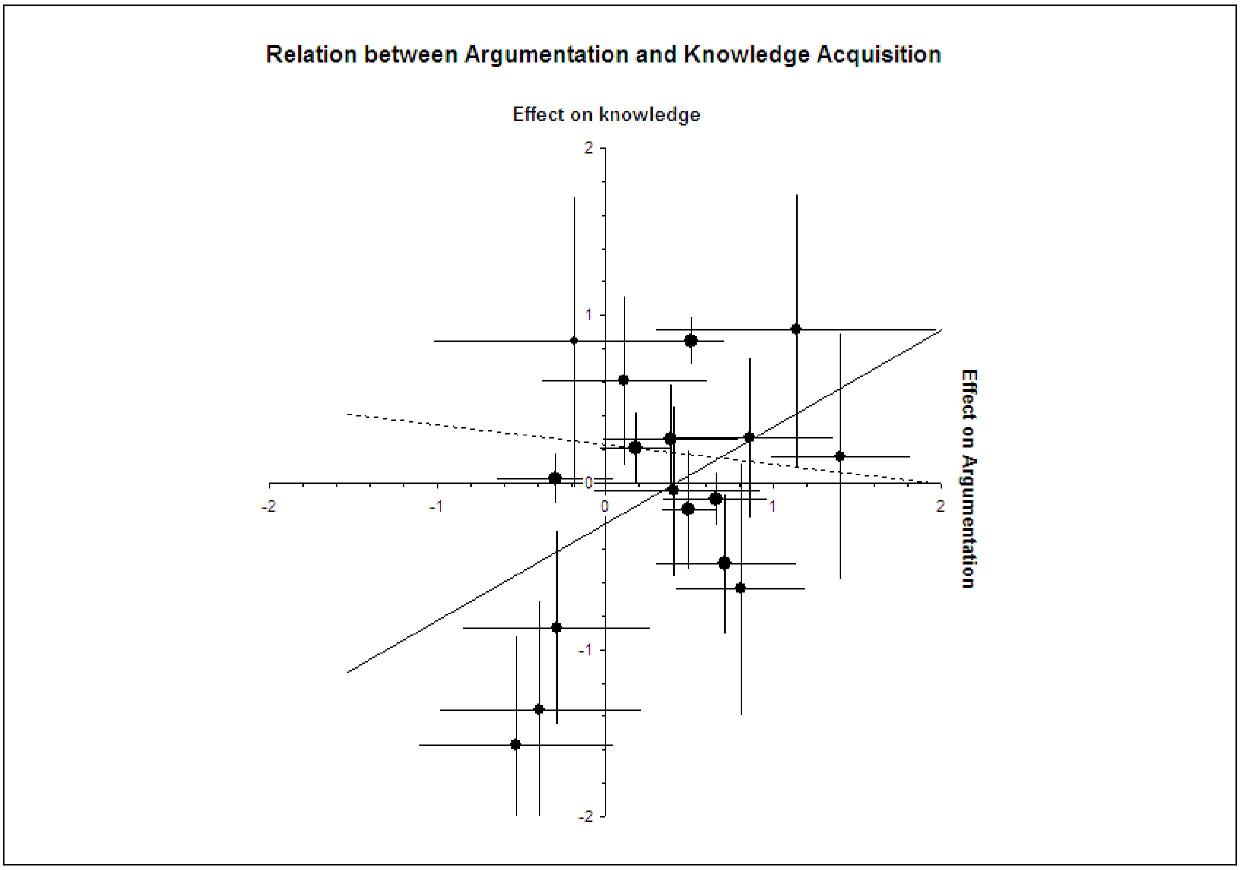

Figure 2. Two-dimensional forest plot of the interventions' effects on argumentation and individual domain-specific knowledge.

Effects on argumentation are displayed on the x-axis, effects on domain-specific knowledge on the y-axis. Each dot represents the ("two-dimensional") effect size for domain-specific knowledge and argumentation for one comparison from the primary studies. The size of each dot is proportional to the comparison's weight according to the random effects model. 90 % confidence intervals are displayed by the horizontal lines for effects on argumentation and by vertical lines for effects on domain-specific knowledge. The continuous diagonal line represents the equation of the meta-regression including all studies ($d_{knowl\_i} = -0.249 + 0.580 \cdot d_{arg\_i}$), whereas the dotted line represents the meta-regression after exclusion of the van Drie et al. (2005a) study ($d_{knowl\_i} = 0.227 - 0.120 \cdot d_{arg\_i}$).

Before discussing the consequences of these findings, potential limitations of this meta-analysis have to be taken into account. Its main shortcoming is its reliance on meta-regression to predict study-level or comparison-level effects on domain-specific knowledge on the basis of effects on argumentation instead of the more common approach of integrating within-study correlations between argumentation and domain-specific knowledge gains. However, apart from the fact that the latter is currently not feasible, the former approach constitutes a stringent test of the assumption that argumentation fosters domain-specific knowledge acquisition. While a correlation between effects of argumentation interventions on argumentation and effects on domain-specific knowledge does not imply that argumentation quality and domain-specific knowledge acquisition are correlated on the level of groups or individuals, the lack of a correlation on the level of experimental comparisons provides some indication against the assumption that argumentation fosters domain-specific knowledge acquisition, at least as they were measured in the studies included in the analysis.

What can the findings from this meta-analysis tell us? First of all, the body of research that has addressed the role of argumentation for domain-specific knowledge acquisition in CSCL using a quantitative methodology that allows for meta-analytic integration, is not abundant. Given this relative scarcity of evidence, more caution with respect to general statements about the effects of argumentation on domain-specific learning seems advisable (e. g. Osborne, 2010, p. 464 f.).

Furthermore, as this meta-analysis shows, the quantitative evidence amenable to integration currently does not provide the compelling picture one would expect, given the dominating view in the field. Rather it suggests reconsidering and even questioning cherished beliefs. It can be questioned, just to give an example, that arguments are "better", i. e. more functional for domain-specific knowledge acquisition, if more parts of them are mentioned explicitly, which is used as an indicator of argumentation quality in several coding schemes (cf. Clark, Sampson, Erkens & Weinberger, 2007) some of which were applied in the studies in the present meta-analysis. A more promising approach could be the analysis of argumentation quality based on a typology of argument schemes that are appropriate for different types of claims (cf. Walton, Reed & Macagno, 2008) because this approach links the formal quality of arguments closer to the content. Another assumption that might be called into question is what might be called the "collaborative enrichment" assumption that argumentation is the joint elaboration of content, which fosters individual knowledge acquisition (e. g.. Andriessen, 2006, p. 445). The mechanisms by which argumentation may sometimes provoke the dismissal of deeply entrenched views and thereby advance understanding, might be slightly more intricate.

Certainly the present meta-analysis does not force us to jettison the conviction that argumentation can lead to domain-specific learning. It should persuade us, however, to develop more accurate accounts of both argumentation quality and the mechanisms by which good arguments in a discussion among learners may influence the further progression of the discussion as well as the learners' cognitive processing and thereby lead to understanding of the content. Therefore, we suggest that future research should address the following points that are located on a theoretical level:

(1) As already indicated, in general, more thorough theoretical accounts of mechanisms that might explain beneficial effects of argumentation on domain-specific knowledge are needed.

(2) Based on such assumptions, a comprehensive taxonomy of any relevant aspects of the amount and quality of argumentation could be developed that allows comparing and integrating studies investigating the same aspects of argumentation. Such a taxonomy could differentiate, among other things, the occurrence or number of different types of arguments, their average or aggregated tenability and relevance, the breadth and depth of topics covered in the discussion (Munneke et al., 2003) or the occurrence of specific argumentative speech acts.

In terms of methodology, based on our experiences, we urge researchers in the field to consider the following points when reporting empirical research about the relation of argumentation and learning in CSCL:

(3) Fine-grained analyses of patterns of argumentation and their relation to learning as they are often found in qualitative studies of the same topic should be translated into quantitative indicators that allow for integration and thereby an assessment of the generality of the findings. Sample size is not so much of an issue in this respect, as evidenced by a study by Wiley and Bailey (2006) with only 8 participants: This study could be integrated in this meta-analysis because the analyses were quantified and all necessary information was reported. This recommendation has the goal to make the often more informative and more valid analyses of argumentation during collaboration typically found in qualitative studies (instead of the more distal acquisition of argumentation skills as measured by posttests) accessible to integration.

(4) Furthermore, it deserves mention that all statistics necessary to compute effect sizes need to be reported in publications. As a general rule, descriptive statistics (including standard deviations and subsample sizes) should be provided. Exact *p*-values and effect sizes should be presented (cf. APA, 2001, p. 138 f.), even in the case of insignificant effects, to allow for unbiased integration. If several measures for the same general variable are used in a primary study, they often need to be collapsed in a meta-analytical integration. This is strongly facilitated if the intercorrelations among the separated indicators are known. Therefore it is desirable that correlation matrices for sets of cognate indicators are presented, as they can often be found in research using multi-dimensional questionnaires. Furthermore, a more thorough practice of reporting indices of intercoder objectivity and reliability would allow for corrections for the attenuation of relationships between variables that is due to error variance in measurements. Most importantly, these indicators should not be averaged across variables, but reported separately for each individual variable used in the statistical analyses (cf. De Wewer, Schellens, Valcke & Van Keer, 2006, p. 11).

(5) As a complement to the current meta-regression approach it is still desirable to integrate findings about person-level covariation between argumentation and domain-specific knowledge acquisition. For this purpose, correlation coefficients are needed that quantify the association between argumentation during the learning phase and domain-specific knowledge as measured in post-tests. To avoid over-estimation of the importance of argumentation for learning, the calculation and presentation of partial correlations controlling for individual motivational and cognitive learning prerequisites such as general cognitive abilities or – as a minimum requirement – domain-specific prior knowledge on the basis of the whole sample rather than specific experimental conditions should become a common standard.

We believe that if these suggestions are picked up in future research, we will soon know more about the role of argumentation for computer-supported collaborative learning than we know now.

References

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

Andriessen, J. (2006). Arguing to learn. In R. K. Sawyer (Hrsg.), *Cambridge handbook of the Learning Sciences* (pp. 443-459). Cambridge: Cambridge University Press.

American Psychological Association (APA). (2003). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: American Psychological Association.

Asterhan, C. S. C. (2008). Processes of argumentation and explanation in conceptual change: Results from protocol analyses of peer-to-peer dialogue. In ISLS (Ed.), *International perspectives in the Learning Sciences: Cre8ing a learning world. Proceedings of the Eighth International Conference for the Learning Sciences - ICLS 2008* (vol. 1, pp. 60-67). ISLS.

Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: Wiley.

*Clark, D. B., D'Angelo, C. M. & Menekse, M. (2009). Initial structuring of online discussions to improve learning and argumentation: Incorporating students' own explanations as seed comments versus an augmented-preset approach to seeding discussions. *Journal of Science Education and Technology, 18*, 321-333.

Clark, D., Sampson, V., Erkens, G. & Weinberger, A. (2007). Analytic frameworks for assessing dialogic argumentation in online learning environments. *Educational Psychology Review, 19*, 343-374.

*Choresh, C., Mevarech, Z. R. & Frank, M. (2009). Developing argumentation ability as a way to promote technological literacy. *International Journal of Educational Research, 48*, 225-234.

De Wever, B., Schellens, T., Valcke, M. & Van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: a review. *Computers & Education, 46*, 6-28.

Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego: Academic Press.

*Janssen, J., Erkens, G., Kirschner, P. A. & Kanselaar, G. (2010). Effects of representational guidance during computer-supported collaborative learning. *Instructional Science, 38*, 59-88.

*Joung, S. & Keller, J. M. (2004). *The effects of high-structure cooperative versus low-structure collaborative design of [sic!] decision change, critical thinking, and interaction pattern during online debates*. Paper presented at the Association for Educational Communications and Technology, 27th, Chicago, IL, October 19-23, 2004.

*Kollar, I., Fischer, F. & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction, 17*(6), 708-721.

Munneke, L., van Amelsvoort, M. & Andriessen, J. (2003). The role of diagrams in collaborative argumentation-based learning. *International Journal of Educational Research, 39*, 113-131.

*Nussbaum, M. E., Sinatra, G. M. & Poliquin, A. (2008). Role of epistemic beliefs and scientific argumentation in science learning. *International Journal of Science Education, 30*(15), 1977-1999.

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science, 328*, 463-466.

Raudenbush, S. W. (1994). Random effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301-321). New York: Russell Sage.

Schwarz, B. B., Neumann, Y., Gil, J. & Ilya, M. (2003). Construction of Collective and Individual Knowledge in Argumentative Activity. *The Journal of the Learning Sciences, 12*(2), 219-256.

*Seethaler, S. & Linn, M. (2004). Genetically modified food in perspective: an inquiry-based curriculum to help middle school students make sense of tradeoffs. *International Journal of Science Education, 26*(14), 1765-1785.

*Stegmann, K., Weinberger, A. & Fischer, F. (2007). Facilitating argumentative knowledge construction with computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning, 4*(2), 421-447.

*van Drie, J., van Boxtel, C., Jaspers, J. & Kanselaar, G. (2005a). Effects of representational guidance on domain specific reasoning in CSCL. *Computers in Human Behavior, 21*, 572-602.

*van Drie, J., van Boxtel, C., Erkens, G. & Kanselaar, G. (2005b). Using representational tools to support historical reasoning in computer-supported collaborative learning. *Technology, Pedagogy and Education, 14*(1), 25-41.

Walton, D., Reed, C. & Macagno, F. (2008). *Argumentation schemes*. Cambridge: Cambridge University Press.

*Weinberger, A., Stegmann, K. & Fischer, F. (2010). Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human Behavior, 26*, 506-515.

*Wiley, J. & Bailey, J. (2006). Effects of Collaboration and Argumentation on Learning From Web Pages. In A. M. O'Donnell, C. E. Hmelo-Silver & G. Erkens (Eds.), *Collaborative Learning, Reasoning, and Technology* (pp. 296-321). Mahwah, NJ: Erlbaum.

*Yeh, K. H. & She, H. C. (2010). On-line synchronous scientific argumentation learning: Nurturing students' argumentation ability and conceptual change in science context. *Computers & Education, 55*, 586-602.