

INVESTIGATING THE RELATION BETWEEN PRE-SERVICE TEACHERS' ARGUMENTATIVE WRITING AND KNOWLEDGE INTEGRATION WHEN LEARNING WITH MULTIPLE TEXTS

Thomas Lehmann

University of Bremen, Germany

Abstract

Recent research shows that learning tasks which require students to consult multiple sources and write an argumentative essay promote the cognitive integration of knowledge and ideas presented in the source documents. This study examines whether this holds true in pre-service teacher education, which asks for an integration of knowledge and ideas from conceptually distinct and mainly separately taught knowledge domains (i.e., content knowledge (CK), pedagogical knowledge (PK), and pedagogical content knowledge (PCK)). Forty-six pre-service teachers received domain-specific textbook excerpts, each representing either CK, PK, or PCK, and an argument writing task. The students read the texts and wrote argumentative essays on how the content provided in the source documents is related as regards teaching. Using structural and analytic methods to investigate the argumentative essays it was found that students' argumentative writing was related to knowledge integration across domains. The generation and inclusion of warrants and the developed connectivity and causality in students' essays appeared particularly important for their knowledge integration. The findings are discussed with regard to how certain instructional scaffolds can support pre-service teachers' argumentative reasoning/writing to enhance knowledge integration across domains. Some directions for future research are drawn.

Keywords: argumentative writing; knowledge integration; writing-to-learn; multiple-documents comprehension; pre-service teacher education

Correspondence concerning this paper should be addressed to:

University of Bremen, Faculty 12: Pedagogy and Educational Sciences. Address: Universitäts-Boulevard 11/13, 28359 Bremen, Germany, Phone: +49 421 218 69130, Fax: +49 421 218 98 69130. E-mail: tlehmann@uni-bremen.de

Introduction

Students are used to engage in learning tasks that require to “consult, contrast, and integrate information from multiple sources in written form” (Mateos, Martín, Cuevas, Villalón, Martínez, & González-Lamas, 2018, p. 119). However, studies on integrating knowledge when trying to adopt information from multiple sources suggest that students struggle with implementing effective integration processes (e.g., Bigot & Rouet, 2007). This is a considerable issue for pre-service teacher education for two reasons: First, the core domains of teachers’ professional knowledge (i.e., content knowledge (CK), pedagogical knowledge (PK), and pedagogical content knowledge (PCK) are taught in a fragmented manner for the most part in initial teacher education (Ball, 2000). Second, much of students’ learning in initial teacher education is dominated by reading texts. These texts are rather domain-specific and do not provide an integrated representation of the different disciplines relevant to teaching. Hence, it is not that surprising that pre-service teachers’ CK, PK, and PCK is hardly integrated (Darling-Hammond, 2006). This circumstance correlates with knowledge compartmentalization, which leads to inert knowledge (Renkl, Mandl, & Gruber, 1996), and thus, is problematic in regards to the knowledge’s applicability. As a result, researchers’ interest in the literacy skills needed to construct an integrated understanding of teaching and instructional strategies to foster students’ knowledge integration has increased in recent years (Lehmann, 2020a).

Some of the studies (e.g., Lehmann, Pirnay-Dummer, & Schmidt-Borcherding, 2020; Lehmann, Rott, & Schmidt-Borcherding, 2019; Wäschle, Lehmann, Brauch, & Nückles, 2015) focus on self-regulated learning with multiple domain-specific sources such as textbooks, journal articles, lecture notes, etc. in (quasi-) experimental *writing-to-learn* settings. In line with prior research on multiple-documents comprehension (e.g., Bråten & Strømsø, 2009; Wiley & Voss, 1996, 1999), it was found that argumentative writing contributes to pre-service teachers’ self-regulated knowledge integration when learning with multiple texts in terms of generating relations between different knowledge domains and merging domain-specific information into a common mental model (Lehmann et al., 2019). However, it remains largely open which argumentative reasoning processes promote the integration of knowledge that originates from diverse domains and is represented in different learning sources. Thus, the goal of the present study was to gain a deeper understanding of the relation between

argumentative writing and the integration of CK, PK, and PCK. In particular, the study aimed to explore pre-service teachers' argumentative writing with regard to their knowledge integration when reading several domain-specific texts and writing an argumentative essay on how the texts' contents relate to each other in reference to teaching.

Knowledge integration through argumentative reasoning and writing

The concept of knowledge integration has evolved within constructivist approaches to learning. According to scholars in educational science and cognitive psychology (e.g., Clark & Linn, 2013; Linn, 2000; Schneider 2012), knowledge integration is perceived as a dynamic mental process of interrelating originally unconnected pieces and structures of knowledge, both pre-existing and newly acquired. It is a form of learning which can promote a professional knowledge base (Lehmann, 2020b; Baumert & Kunter, 2006) and inhibit *knowledge-in-pieces* (diSessa, 1993, 2002; Wagner, 2006) - the latter being more likely to be inert, and hence, inapplicable (Renkl et al., 1996). This is important as both pre- and in-service teachers who provide knowledge in pieces tend to malprioritize their knowledge and make unfavorable decisions for their students' learning (Clift, Ghatala, Naus, & Poole, 1990; Glogger-Frey, Ampatziadis, Ohst, & Renkl, 2018).

Several studies support that argumentation in terms of argumentative reasoning and communication through dialogue or writing benefits learning in general - and presumably knowledge integration in particular. For example, it was found that explaining ideas to oneself (Chi, DeLeeuw, Chiu, & Lavancher, 1994) or to other persons (Webb, Troper, & Fall, 1995) promotes comprehension. Furthermore, research showed that argumentative communication helps students to recognize the incompleteness of their understanding (Keil, Rozenblit, & Mills, 2004). A more recent study of Miller and colleagues (2014) indicates that even the mere anticipation of an argumentative discussion with a disagreeing person is enough to foster an increased cognitive engagement with and an improved conceptual understanding of the learning content. Schwarz (2009) attributes the positive effects of argumentation on learning to the necessity to acquire and reason about new information and a multiplicity of ideas to validate personal theories and speculations, to understand alternative positions, and to formulate objections and/or counter objections. The social constructivist perspective assumes that an argumentation's unique structure of linking premises, conclusion, conditions,

rebuttals etc. to each other affects human thinking and subsequently improves and extends the organization of knowledge (ibid.).

Wiley and Voss (1996, 1999) conducted several experiments to examine the influence of argument writing tasks on knowledge integration in terms of (a) relating pieces of information and ideas for the purposes of justification, and (b) developing a deep integrated understanding of multiple learning sources. Their results showed that argument writing tasks provoke a higher degree of transformation, integration, and causality in students' essays compared to tasks that instructed to write a narrative, a summary, a description, or an explanation.

Bråten and Strømsø (2009) also investigated how to enhance a well-integrated understanding of multiple study texts by comparing an argument writing task, a summary writing task, and a task that asked to write a general overview for global comprehension. The authors tested their participants' ability to draw inferences across the content provided in different study texts (on the topic climate change) and found that argument and summary writing are more effective in provoking a deep-level integrated understanding of multiple study texts after controlling for gender, age, word decoding, and prior knowledge than writing a general overview. In contrast to the findings of Wiley and Voss (1996, 1999), the results of Bråten and Strømsø's study did not indicate argument writing tasks to be superior to summary tasks. However, this result might be explained by the fact that all participants were asked to pay attention to the causes and effects provided in the texts, and to elaborate on the most relevant information given in the documents irrespective of the experimental condition.

Gil, Bråten, Vidal-Abarca, and Strømsø (2010a, 2010b) continued Bråten and Strømsø's (2009) investigation in order to contrast the effect of summary writing with argument writing on students' knowledge integration. In their first study they found that personal epistemology about the topic of texts might be the decisive factor for the effectiveness of argument writing tasks (Gil et al., 2010a). However, their data provided only limited evidence for the hypothesis that argument tasks are more beneficial for students with sophisticated epistemological beliefs. The authors admit that their participants presumably lacked experience in argumentative writing due to the educational system they went through. Hence, Gil et al. (2010b) considered prior topic knowledge as a covariate in a subsequent study. This study indicated that high prior knowledge allows to take advantage of argumentative writing whereas low-knowledge learners seem to be overstrained.

In a recent study, Lehmann et al. (2019) took a first step to adapt argument writing tasks and task-supplemental prompts to foster pre-service teachers' knowledge integration across domains when learning with multiple texts on CK, PK, and PCK. Three textbook excerpts served as learning sources representing the CK-, PK-, and PCK-domain of teachers' professional knowledge. As regards content, the texts addressed logics and mathematical proving (CK-text), an information processing model, learning strategies, and general consequences for classroom practice (PK-text), and individual and social difficulties with mathematical proving as well as a process model for teaching mathematical reasoning (PCK-text). In a laboratory experiment Lehmann and colleagues instructed participants in an argument writing condition to read the three domain-specific documents, understand them as a whole, and write an argumentative essay about how the contents relate to each other. The essays were compared to those of participants who received an unspecific writing task for global comprehension. Based on a category-driven content analysis, results showed that the argument writing task promoted knowledge integration as measured by (a) generating and providing integrative elaborations, and (b) making more switches between domain-specific information in pre-service teachers' essays.

Taking these results together, one can conclude that argument writing tasks are an effective instructional strategy to enhance knowledge integration in general, and pre-service teachers' integration of CK, PK, and PCK in particular. For argumentative writing in general, research additionally suggests that explicit argument instruction and prompting specific goal setting enhance writing performance, the awareness for different views on a particular issue, developing argument knowledge and strategies, and applying knowledge in other topics (*e.g.*, Ferretti, MacArthur, & Dowdy, 2000; Page-Voth & Graham, 1999; Varghese & Abraham, 1998; Yeh, 1998a). However, most of these studies were conducted with students in school and at an undergraduate level. Moreover, it remains unclear which reasoning processes underlying argumentative writing cause the improvements in students' across-domain knowledge integration (*e.g.*, pre-service teachers' CK, PK, and PCK integration). Thus, there is a need for more research that aims at a deeper understanding of the relation between argumentative writing and knowledge integration. To address this desideratum the present study integrates theoretical and empirical models on argumentative reasoning and writing as well as different methodical approaches to assessment.

Argumentative reasoning and writing

For several decades research on argumentative reasoning perceived *reasoning* as a strictly logical cognitive activity, and the understanding of an *argument* was limited to the meaning of a logical argument (*cf.* Hornikx & Hahn, 2012). Today, these concepts are characterized in a broader sense in cognitive psychology and educational science. The concept of individuals' reasoning includes that uncertain information and relevance relations between pieces of information, which lie outside of (traditional) logic, can be central for argumentative reasoning just as logical inferences. This understanding involves that people reason and construct arguments without deduction, too, and that this does not necessarily make them unskillful reasoners. In accordance with this percept, the idea that bringing an argumentation forward is a linguistic act in which something needs to be clarified by making it (largely) undisputed was adopted for the present study. This includes that it is irrelevant whether argumentation is realized verbally or through writing and whether the contentious is something already manifest or only vague and/or potentially controversial. However, a person must realize that the issue is actually arguable, otherwise he/she will not engage in argumentation. This is due to the basic principle of argumentation, which is to clarify the issue under reference to the (rather) indisputable. Against this understanding, argumentative reasoning is considered "a flow of propositions within a discourse of reasoned argumentation" (Anderson et al., 2001, p. 2). Apart from the expanded conceptualization of argumentative reasoning and argumentation which goes beyond traditional logic, this percept is also in line with Piaget's (1928) notion that "reasoning is an argument which we have with ourselves, and which reproduces internally the features of a real argument" (p. 204).

A major concern of research on argumentative reasoning was to grasp and evaluate the intra-individual reasoning processes which originate argumentation. To this end, Reznitskaya, Kuo, Glina, and Anderson (2009) propose a synthesis of the Argument Schema Theory (AST) and normative models of argumentation. AST integrates social-constructivist conceptions of learning and a structuralist percept of knowledge organization and storage. According to AST, an argument schema involves a generalized mental representation of knowledge about and skills of argumentation. Drawing from normative models of argumentation (*e.g.*, Toulmin, 1958), different field-invariant structural elements were identified and complemented to constitute the components of a well-developed argument

schema. These components (also called “Toulmin elements” or “elements”) are claims, reasons, backings, qualifiers, warrants, counter-claims, and rebuttals. Table 1 provides descriptions of the six elements and the counter-claim, which opposes an argument. These (or similar) descriptions have been used for the purpose of coding in previous studies (*e.g.*, Jiménez-Aleixandre, Rodríguez, & Duschl, 2000; Kelly, Druker, & Chen, 1998; Lehmann & Schmidt-Borcherding, 2018; Simon, 2008).

Table 1. Description of structural argument elements (Lehmann & Schmidt-Borcherding, 2018)

Structural argument element	Description
Claim	Student states his/her position on an issue or question by way of a conclusion or hypothesis that suggests the relatedness of at least two variables (cause-effect) or thematic aspects (content-wise relatedness). The claim is presented as true.
Reason	Student uses data/facts to justify a claim or provide support for it. This includes empirical data (<i>e.g.</i> , data from observations), hypothetical data (<i>e.g.</i> , data from thought experiments, examples), and facts under reference to authoritative sources (<i>e.g.</i> , by citing). The reason always gives answer to the question, why someone poses a specific claim.
Warrant	Statement of a student that establishes the connection between a claim and the reason for the claim. A warrant shows that the move from reason to claim is valid. It answers the question, how claim and reason are connected or to what extent a reason justifies a claim. A warrant refers to two argument elements namely a claim and its reason.
Qualifier	Student includes a statement that limits the validity of a claim without or with specifying a condition or certain conditions under which the claim holds true. The qualifier restricts the range of an argument.
Backing	Student includes further data/facts, information, rules, theories, or assumptions to support either a given reason, a warrant, or a qualifier in order to strengthen the argumentative effect of the according component. In contrast to a warrant, backings solely refer to a single supporting argument element.
Rebuttal	Student presents an objection to an aforementioned supporting argument element (<i>i.e.</i> , reason, warrant, and backing). The rebuttal has a similar function as the counter-claim in relation to a claim. A rebuttal can be unspecific (“that’s not true”) or the opposite is formulated (<i>e.g.</i> , by negation of the supporting argument element).
Counter-claim	Student states counter-position or an alternative position. Counter-claims can contradict a claim both explicitly (by denoting a claim as untrue) and implicitly (by stating an alternative position without explicit reference to the claim). A counter-claim does not necessarily assert exactly the opposite of a claim.

Furthermore, AST assumes that argumentative reasoning depends on a set of epistemological beliefs which shape an “explanatory framework” (Mishra & Brewer, 2003) for the argument schema. Hence, analyzing argumentative reasoning is never truly field-independent, even if only the field-invariant structural elements are used for the coding of arguments. This is in line with Zohar

and Nemet's (2002) modification of Toulmin's framework, which suggests that an evaluation of the quality of (written) arguments asks for an examination of structure and content (*see also* Garcia-Mila & Andersen, 2007).

Based on this theoretical background, many scholars apply a *structural approach* to score students' arguments with regard to which and to what extent different structural components are represented in the verbal and/or written formulations in order to evaluate argumentative reasoning performance (*e.g.*, Jiménez-Aleixandre et al., 2000; Kelly et al., 1998; Simon, 2008). It is commonly accepted that "stronger arguments contain more of these different components than weaker arguments" (Sampson & Clark, 2006, p. 656). Still, not every "good" argument will necessarily consist of all of these elements. Moreover, it is important to note that the structural approach for measuring argumentative reasoning performance is more than simply analyzing the occurrence of different elements of an argument. Attributing whether and which field-invariant elements are included in an argumentative speech or text demands due consideration of both content and context, that is, the field-dependent question of what counts as an appropriate claim, warrant, backing, etc. (Sampson & Clark, 2006; Zohar & Nemet, 2002).

Sandoval's *analytical approach* for the investigation of argumentative reasoning involves to account for the conceptual and the epistemic quality of an argument (Sandoval, 2003; Sandoval & Millwood, 2005). The framework suggests that sophisticated argumentative reasoning requires a conceptual understanding of relevant theories and their application to clarify an issue, as well as an epistemic understanding of what characterizes high-quality arguments. Following the framework, one needs to estimate arguments with regard to different dimensions. This involves to evaluate how well a student has formulated causal claims and warranted these claims using available data to measure the *conceptual* quality of his/her argumentative reasoning performance. Then, the *epistemological* quality is assessed based on how well the student has cited relevant data to warrant his/her claim within a specific theoretical scope, formulated a coherent causal explanation for a phenomenon, and included proper rhetorical references when referring to data. Sandoval (2003) argues that it is possible to determine whether an individual can compose an argument which clarifies a particular issue by referring to these dimensions.

Other researchers include further macro-level features in their approach to measure students' argumentative reasoning/writing performance. For example, the

holistic approach of Hidi, Berndorff, and Ainley (2002) considers (a) writing in a mature style, (b) reasoning persuasively, (c) observing the conventions of written language, and (d) including different argument elements, that is, position or thesis statement, supportive statements and evidence, treatment of the opposing view, and conclusion. Reznitskaya et al. (2009) also applied a holistic approach in their study but they did not want to evaluate word choice, spelling, and/or punctuation, and thus writing performance. To focus exclusively on reasoning performance, their holistic scoring concentrated exclusively on (a) the provision of argument components, (b) the discussion of opposing view, (c) the structure and focus of the text, and (d) organization signals and the clarity of argumentation. Reznitskaya and colleagues compared the analytic and the holistic approach and found that, although both methods have adequate reliability, the analytic approach is more sensitive to and leads to more detailed information about differences in students' argumentative reasoning performance.

Contrary to Reznitskaya et al. (2009), the studies of Yeh (1998a, 1998b) indicate that the epistemic and conceptual quality and further macro-level features (*i.e.*, language use and compliance with language conventions) are crucial for the purpose of assessing argumentative writing performance. For example, the results of Yeh's (1998b) "development study" support the validity of a three-factorial scoring scheme, which involved the quality dimensions development (including organization, focus, and clarity), adherence to conventions regarding correctness, and voice (*i.e.*, use of language for rhetorical effects). However, considering Reznitskaya et al.'s (2009) findings it can be argued that the details included in Yeh's (1998a, 1998b) development scoring rubric are useful for facilitating an *analytical* assessment of students' argumentative reasoning performance, as well. That is, to be sensitive to and generate detailed information about students' reasoning researchers should consider to what degree an essay comprises (a) a definite, well-qualified claim or proposal, (b) strong, developed, well-organized supporting arguments, (c) a response to major objections and alternatives, (d) definitions, elaborations, and illustrations of key terms, ideas and connections to avoid misinterpretation, (e) connecting words or ideas that explicate how statements build on each other through, and (f) clear, concise, and consistent wording. Nevertheless, one must be aware that students' adherence to conventions, and their voice (*i.e.*, appropriate, sophisticated, audience-centered, vivid language filled with conviction) can influence such an analytical approach.

Objectives and hypotheses of the study

Against this background, the present study aimed to investigate the relation between argumentative writing and knowledge integration using different methodical approaches. More specifically, it sought to examine which specific components and quality dimensions of pre-service teachers' argumentative essays predict their knowledge integration across the core domains of their professional knowledge (*i.e.*, CK, PK, and PCK) when learning from multiple texts. Two different methodical approaches were adapted to examine pre-service teachers' argumentative essays, that is, the structural approach and the analytical approach. The holistic approach was not used in this study due to its lack in information richness as compared to an analytic scoring (*cf.* Reznitskaya et al., 2009).

Relating to the structural argument writing performance, the study examined whether students' generation and provision of certain structural argument elements can predict their knowledge integration across domains. The *argument structure hypothesis* expects that students' knowledge integration in argumentative writing is particularly dependent from (a) *claiming* (*i.e.*, stating a position by way of a conclusion or hypothesis that suggests the relatedness of at least two variables) and (b) *warranting* (*i.e.*, establishing the connection between a claim and a corresponding reason to explicate how they are connected or to what extent the reason justifies the claim). Moreover, the study investigated different macro-level features (*i.e.*, quality dimensions) of written argumentations and estimated the degree to which these dimensions are relevant to learners' knowledge integration. The *quality dimension hypothesis* assumes that (a) *connections and causality*, (b) *plausibility*, (c) *structure and coherence*, and (d) *focus and clarity* in pre-service teacher's argumentative essays are the quality dimensions related to knowledge integration across domains.

As regards knowledge integration, the present study followed the definitions of Schneider (2012) and Linn (2000) which describe knowledge integration as a form of learning that involves interrelating originally unconnected pieces and structures of knowledge. Accordingly, the study focused students' *integrative elaboration* of information and knowledge entities that pertain to different knowledge domains (*i.e.*, CK, PK, and PCK) and were represented in separate source material. This appears reasonable because (a) elaboration is a cognitive learning strategy which targets deep learning by way of establishing relationships, connecting the content with pre-knowledge and personal

experiences, and generating examples, comparing different concepts, and examining the conclusiveness of arguments (Lehmann et al., 2019; Pintrich, Smith, Garcia, & McKeachie, 1993; Wäschle et al., 2015), and (b) much of students' learning in initial teacher education is dominated by reading texts, which are rather domain-specific and do not provide an integrated representation of different disciplines relevant to teaching.

Method

Participants

Forty-six pre-service primary school mathematics teachers participated in this study. The participants had an average age of $M=24.6$ years ($SD=4.67$). They were enrolled in a Master of Education (M. Ed.) teacher education program, and studied in average for a total of $M=6.65$ semesters ($SD=1.49$) inclusive of their Bachelor's studies. The vast majority (93%) of participants were female, which is typical for pre-service primary school teachers in Germany. All participants were native speakers of German. They already had some practical work experience through student teaching internships (average duration: $M=4.48$ months, $SD=4.47$). They were recruited in courses on learning and instruction where they could voluntarily sign up to participate in the study for course credit towards the end of the lecture period.

Materials

The study used material that has already been tested and used successfully in the laboratory study of Lehmann et al. (2019). This included three domain-specific textbook excerpts on CK, PK, and PCK, respectively, an argument writing task, knowledge tests, and the coding approach to capture participants' integrative elaborations across domains. The study texts, which were handed to the participants as domain-specific learning material, were slightly adapted excerpts taken from German textbooks (*see* Table 2). According to lecturers from several universities, these textbooks are frequently used in German teacher education within the CK, PK, and PCK domain. Each document essentially allowed to make connections between information that is presented in one or both other texts. The argument writing task asked the participants to read the texts carefully and to understand them as a whole. It further instructed the participants to "write an

argument about how the contents of the various documents are related in regard to teaching”.

Table 2. Description of documents taken from teacher education textbooks and used as learning material

Professional knowledge domain	Content description	Source	Words
CK	Logic and mathematical proving	Grieser, D. (2015). <i>Analysis I</i> . Wiesbaden: Springer.	1707
PK	Learning relevant cognitive and metacognitive processes and some consequences for classroom practice	Renkl, A. (2015). <i>Wissenserwerb</i> [Knowledge acquisition]. In E. Wild & J. Möller (Eds.), <i>Pädagogische Psychologie</i> [Educational psychology] (pp. 3-24). Berlin: Springer.	1950
PCK	Pupils' individual and social difficulties in mathematical reasoning and proving, and a process model for teaching mathematical reasoning	Brunner, E. (2014). <i>Mathematisches Argumentieren, Begründen und Beweisen. Grundlagen, Befunde und Konzepte</i> [Mathematical reasoning, justifying, and proving. Basics, findings, and concepts]. Berlin: Springer Spektrum.	2183

Note: Three text documents provided the content to be studied and cross-related in an argumentative essay. All documents were slightly adapted to increase the internal validity of the study.

Measures

“Within-domain” Knowledge Acquisition. To measure participants' CK, PK, and PCK acquisition within each domain, three separate knowledge tests were used as pre-post-measures. Each test was developed exclusively based on the study texts that served as learning material. The CK test consisted of a total of 13 items (nine multiple-choice questions with four response options (one correct), and four short answer question with open text response). The PK test also consisted of seven multiple-choice and four short answer questions. The PCK test comprised seven multiple-choice and five short answer questions. Open text responses were dichotomously scored either correct or incorrect. The test items measured both factual knowledge and conceptual understanding. More information about the reliability of the tests and item examples can be found at Lehmann et al. (2019).

Structural Argumentative Writing Assessment. Toulmin's argument pattern (Toulmin, 1958) was taken as a theoretical starting point to assess the structural argument quality of participants' essays. This methodical decision is theoretically underpinned in that stronger arguments comprise more of the different Toulmin elements than weaker arguments (Sampson & Clark, 2006;

Simon, 2008). For the development of a coding scheme, more recent empirical work, which implemented structural argument analyses following the Toulmin-framework (Jiménez-Aleixandre et al., 2000; Kelly et al., 1998; Simon, 2008), was taken into consideration, as well. Based on this work, descriptions for the structural argument elements (*i.e.*, claim, reason, warrant, backing, qualifier, rebuttal, and counter-claim; *see* Table 1) were enunciated and used as coding categories. As mentioned earlier, coding the statements included in an argument as certain field-invariant elements requires to factor both content and context, that is, the field-dependent question of what counts as an appropriate claim, warrant, backing, etc. (Sampson & Clark, 2006; Toulmin, 1958; Zohar & Nemet, 2002). The author and a graduate student research assistant independently coded all essays with close to satisfying interrater agreement (Cohen's $\kappa=.68$). Coding differences were solved through discussion.

Analytic Argumentative Writing Assessment. The analytic scoring aimed to assess different quality dimensions of students' essays. To this end, a scoring scheme with eight categories, which were derived from both theory and methodical procedures described by Hidi et al. (2002), Reznitskaya et al., (2009), Sandoval (2003), and Yeh (1998a, 1998b), was developed while making it consistent to the theoretical background depicted above. Hence, the scoring of the categories involved considering the conceptual understanding of relevant theories and their application to clarify an issue, and the epistemic understanding of what characterizes high-quality arguments. The categories were (a) connections and causality, (b) plausibility (content validity), (c) structure and coherence, (d) variation of different argument elements, (e) explicit differentiation of argument elements, (f) focus and clarity, (g) voice (use of language for rhetorical effects, and (h) adherence to language conventions. All categories were independently scored by two raters on a 5-point rating scale (from 0 to 4 points) according to the criterion descriptions (*see* Appendix A for the full scoring scheme). The raters were the same as for the structural analysis. Interrater reliabilities using intraclass correlation coefficients (ICCs) were estimated. The resulting ICCs were between .81 and .88 suggesting good agreement between raters (*see* Appendix A for all ICCs). Hence, the two raters did not seriously reduce statistical power for the following analyses. The final score for each quality dimension was calculated by averaging the scores assigned by the two raters.

Across-domain Knowledge Integration. To score participants' integration of CK, PK, and PCK, the procedure of Magliano, Trabasso, and Graesser (1999)

was adapted. This procedure has been successfully applied in multiple-documents comprehension research (e.g., Gil et al., 2010a, 2010b; Wiley & Voss, 1999), and with regard to pre-service teachers' knowledge integration across domains (Lehmann et al., 2019). First, all essays were parsed into units of ideas. Then, those idea units in which a student integrated knowledge and/or ideas from two or all three of the domain-specific text sources (and potentially prior knowledge) were identified. Such idea units were coded as an *integrative elaboration* and considered eligible as across-domain knowledge integration measure. Two independent raters coded the whole set of the data independently with good interrater reliability (Cohen's $\kappa=.79$). The raters were not the same as for the structural and analytic analyses, and they were familiar with the source material. Coding differences were solved through discussion.

Design and procedure

The study followed a single group correlational design. The participants were recruited in general pedagogy courses on learning and instruction where they could voluntarily sign up to participate in the study. Then, they received a link to an online survey website via e-mail. First, the website welcomed the participants and asked if they in fact are pre-service *mathematics* teachers because the study involved mathematical CK and PCK. After affirmation, the participants completed a demographic data survey and a code-generating item (which allowed to match the data and essays). Next, they completed the knowledge test (pre-test). Finally, they received the argument writing task and the three texts as separate downloadable pdf-files. The participants were informed to process the argument writing task within the next two weeks, to document how much time they spent on processing the task, and to submit their argumentative essay online (as Word- or pdf-file) through a file upload tool. They received an automatic reminder three days before the submission date (76 % submitted their essay prior to the reminder). The upload tool erased all file attributes of the essays, thus, allowing to collect the data anonymously. Together with their file upload, the participants answered the code-generating item and an item on the time they spent on processing the task. In average, the participants reported that it took them 81 minutes ($SD=14.16$) to complete their essays (inclusive of reading the source texts). Finally, participants processed the knowledge post-test.

Data analysis

The data from the participants was collected and coded as described above and then analyzed applying univariate-descriptive, bivariate-correlational, and multivariate approaches. The descriptive analysis aimed at providing a clear picture of the different measures of the study. Participants' knowledge acquisition (pre-post) was tested for significance using a multivariate analysis of variance and subsequent analyses of variance for each knowledge domain (*i.e.*, CK, PK, and PCK). The correlational analysis focused on the relation between each of the different argument elements and knowledge integration, as well as on the relation between the different argument quality dimensions and knowledge integration - the latter as measured by the frequency of integrative elaborations included in the essays. A multiple regression analysis with backward elimination was performed to simultaneously test the hypotheses of the study. *P*-values <.05 were considered significant. All statistical analyses were performed using SPSS version 25.

Results

Initial data analysis

As an initial data analysis, descriptive statistical values were calculated for all measures including knowledge gains (delta-values) in CK, PK, and PCK (*see* Table 3). As can be seen in Table 3, participants increased their CK, PK, and PCK. A multivariate analysis of variance with repeated-measures for domain-specific knowledge acquisition using the pre- and post-test results indicates that the knowledge gain was significant, Wilks' $\lambda=.159$, $F(3,43)=75.61$, $p<.001$, $\eta_p^2=.841$. Subsequent repeated-measures ANOVAs showed a significant increase of knowledge in all three domains (CK: $F(1,45)=52.67$, $p<.001$, $\eta_p^2=.539$; PK: $F(1,45)=206.1$, $p<.001$, $\eta_p^2=.821$; PCK: $F(1,45)=40.65$, $p<.001$, $\eta_p^2=.475$).

Table 3. Means, standard deviations, minima, and maxima for all variables

Measure	Variable	M	SD	Min	Max
CK	Pre-test	2.94	2.25	0	10
	Post-test	5.67	2.13	0	10
	Δ CK	3.05	2.15	0	8
PK	Pre-test	1.63	1.76	0	6
	Post-test	7.12	1.82	3	11
	Δ PK	5.59	2.58	0	10

Table 3. Means, standard deviations, minima, and maxima for all variables - *continued*

Measure	Variable	M	SD	Min	Max	
PCK	Pre-test	2.41	1.31	0	6	
	Post-test	4.52	1.87	0	9	
	Δ PCK	2.34	1.99	0	7	
Knowledge integration Structural argument elements	Integrative Elaboration	6.48	4.41	0	18	
	Claim	5.02	2.68	0	14	
	Reason	4.47	3.44	0	19	
	Warrant	1.53	1.69	0	7	
	Qualifier	1.22	1.15	0	4	
	Backing	.18	.49	0	2	
	Rebuttal	.04	.21	0	1	
	Counter-claim	.02	.15	0	1	
	Argument quality dimensions	Connections and causality	2.48	.90	0	4
		Plausibility (content validity)	2.51	.83	0	4
Structure and coherence		2.48	.78	1	4	
Variation of different argument elements		2.41	.79	0	4	
Explicit differentiation of argument elements		2.45	.77	0	4	
Focus and clarity		2.41	.90	0	4	
Voice (use of language for rhetorical effects)		2.53	1.00	0	4	
Adherence to language conventions		3.12	.73	1	4	

Table 4 presents the zero-order correlations between different argument elements and quality dimensions, respectively, and the provision of integrative elaborations in participants' written arguments. The correlational results show that many of the variables correlated significantly (55 out of 120 correlations). The provision of integrative elaborations in students' essays was positively related to six out of the 15 argumentative variables. Three of those referred to the inclusion of certain argument elements, that is, (a) claim, $r=.42$, (b) reason, $r=.39$, and (c) warrant, $r=.51$. The other three correlating variables referred to different quality dimensions, that is, (a) connections and causality, $r=.48$, (b) variation of different argument elements, $r=.42$, and (c) explicit differentiation of argument elements, $r=.31$.

These results indicate that various categories, both structural and conceptual/epistemological, were related to each other. Taking the amount of inter-correlations among the variables into account, a multiple regression analysis with backward elimination was conducted to test the hypotheses of this study simultaneously and adjust for potentially confounding effects in the bivariate analyses.

Table 4. Zero-order correlations (Pearson's r) between integrative elaborations, structural argument elements and quality dimensions

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 Integrative Elaboration	—															
2 Claim	.42 ^{***}	—														
3 Reason	.39 ^{**}	.81 ^{**}	—													
4 Warrant	.51 ^{**}	.41 ^{**}	.46 ^{**}	—												
5 Qualifier	.26	.34 [*]	.29	.50 [*]	—											
6 Backing	.29	.39 [*]	.16	.30 [*]	.33 [*]	—										
7 Rebuttal	-.02	.11	.11	.22	.10	-.06	—									
8 Counter-claim	-.15	.08	.13	.06	.05	-.08	-.03	—								
9 Connections and causality	.48 ^{**}	.55 ^{**}	.50 ^{**}	.41 ^{**}	.41 [*]	.17	.09	.13	—							
10 Plausibility (content validity)	.26	.36 [*]	.38 ^{**}	.11	.19	-.11	.09	.07	.55 ^{**}	—						
11 Structure and coherence	.25	.21	.19	.17	.28	-.10	.21	.09	.53 ^{**}	.69 ^{**}	—					
12 Variation of different argument elements	.42 ^{**}	.49 ^{**}	.47 ^{**}	.33 [*]	.41 ^{**}	.11	.22	.17	.66 [*]	.68 ^{**}	.72 ^{**}	—				
13 Explicit differentiation of argument elements	.31 [*]	.34 [*]	.28	.14	.27	-.02	.22	.02	.54 ^{**}	.61 ^{**}	.77 ^{**}	.74 ^{**}	—			
14 Focus and clarity	.27	.24	.21	.25	.43 ^{**}	.08	.11	.15	.69 ^{**}	.61 ^{**}	.77 ^{**}	.74 ^{**}	.59 ^{**}	—		
15 Voice (use of language for rhetorical effects)	.28	.31 [*]	.26	.14	.52 ^{**}	.12	.16	.06	.61 ^{**}	.57 ^{**}	.77 ^{**}	.79 ^{**}	.74 ^{**}	.77 ^{**}	—	
16 Adherence to language conventions	-.01	-.05	-.07	-.02	.16	.08	-.02	.27	.24	.35 [*]	.54 ^{**}	.44 ^{**}	.36 [*]	.59 ^{**}	.51 ^{**}	—

Note: ^{*} p<.05; ^{**} p<.01

Hypotheses testing

Table 5 displays a summary of the full and the final regression model which resulted from the multiple regression analysis with backward elimination entering those argument elements and quality dimensions as predictors that proved significant in the correlational data analysis. According to the F-test results, the regression of the full model was significant, $F(6,45)=53.74, p=.005$. It provided a substantial goodness of fit, adjusted- $R^2=.27$. Hence, about 27 percent of the variance in students' integrative elaborations could be explained based on the model with all entered predictors. Although the overall regression of the full model with all candidate predictors seemed to provide sufficient explanatory power with satisfying accuracy, many of them were not statistically significant.

Using the backward eliminations method, the final regression model was reached in step 5, $F(2,45)=11.42, p<.001$ (see Table 5), with 'reason' being removed after step 1, 'differentiation of argument elements' after step 2, 'claim' after step 3, and 'variation of different elements' after step 4. A sixth step, in which participants' prior domain-specific knowledge was added as predictor candidates, was carried out. However, participants' pre-knowledge in CK, PK, and PCK remained insignificant predictors.

Table 5. Summary of multiple regression analysis for argument variables predicting knowledge integration

	Predictors	Raw <i>B</i>	<i>SE</i>	Stand. β	<i>t</i>	<i>p</i>
Full model	Claim	.224	.380	.135	.589	.559
	Reason	-.057	.295	-.044	-.194	.847
	Warrant	.936	.396	.354	2.366	.023
	Connections and causality	.934	.914	.191	1.022	.313
	Variation of different argument elements	.572	1.234	.103	.464	.646
	Explicit differentiation of argument elements	.279	1.121	.049	.249	.805
Final model	Warrant	.984	.357	.372	2.757	.009
	Connections and causality	1.613	.661	.329	2.440	.019

The final model yielded an adjusted- R^2 of .32, thus, suggesting a good model fit. Accordingly, 32 percent of the variance in students' integrative elaborations could be explained based on the final model. Compared to the full model, the adjusted- R^2 of the final model increased by .05. Regarding the question, which of the candidate predictor variables would be significant for

explaining knowledge integration across domains, the results showed that students' inclusion of warrants ($\beta=.37$), and the connections and causality scores of their essays ($\beta=.33$) were significant positive predictors with medium effect sizes.

The result of a Durbin-Watson test for the regression analysis did not indicate auto-correlation ($d = 2.29$). Variance inflation factor (VIF) coefficients were between 1.20 and 3.24. Thus, the requirements for an interpretation of the above-reported results were met.

Discussion

There has been increasing interest in pre-service teachers' knowledge integration among educational and psychological researchers and teacher educators in recent years (*see* Lehmann, 2020a). Lehmann (2020b) has argued that when (pre-service) teachers are able to interrelate pieces and structures of their CK, PK, and PCK, and merge them into a more coherent, common knowledge base, they will provide less inert knowledge and gain more professional competence. That is, they will be able to design more effective lessons and learning tasks for their students, and be more effective in classroom practice (*see also* Clift et al., 1990; Darling-Hammond, 2006; Glogger-Frey et al., 2018; Wäschle et al., 2015). A particular line of research within this context has focused on instructional strategies (as opposed to curricular strategies) to foster knowledge integration in learning environments that involve multiple domains (*e.g.*, Lehmann et al., 2020, 2019; Wäschle et al., 2015). The studies of Lehmann and colleagues (2020, 2019) took first steps to adapt argument writing tasks for pre-service teachers' knowledge integration. A major finding was that argument writing tasks can be an effective mean to foster pre-service teachers' knowledge integration in writing-to-learn settings that include multiple study texts representing CK, PK, and PCK. However, it remained unclear which reasoning processes underlying argumentative writing caused the improvements in students' across-domain knowledge integration (*e.g.*, pre-service teachers' CK, PK, and PCK integration). Yet, this would allow to develop learner-specific instructional assistance. Thus, the present study aimed at gaining a deeper understanding of argumentative writing in relation to knowledge integration. Specifically, the goal of this study was to examine different aspects of pre-service teachers' argumentative writing with

regard to their role for knowledge integration when learning from multiple domain-specific texts. Based on previous findings on argumentative reasoning and writing, two hypotheses were tested: (1) the argument structure hypothesis expected students' knowledge integration across domains in argumentative writing to be dependent from (x_1) claiming and (x_2) warranting. (2) The quality dimension hypothesis assumed that an essay's (a) connections and causality, (b) plausibility, (c) structure and coherence, and (d) focus and clarity are related to knowledge integration across domains.

Before testing these hypotheses, participants' knowledge acquisition within each domain was examined. Results showed that the processing of the argument writing task (which required to engage with three domain-specific textual learning sources and relate the topics addressed as regards teaching) led to significant improvements in declarative and conceptual knowledge within the CK-, PK-, and PCK-domain. This finding supports the efficacy of "arguing to learn" (Schwarz, 2009) in terms of constructing specific knowledge and developing understanding. Thus, the study contributes to the research body showing that although students often fail to develop sophisticated written arguments, the writing of argumentative essays is conducive to students' learning of important scientific ideas and construction of professional knowledge (e.g., Lehmann et al., 2019; Sandoval & Millwood, 2007).

With regard to the argument structure hypothesis, the results of bivariate correlation analyses were supportive in terms of the identified relations between (x_1) claiming, as well as (x_2) warranting, on the one hand, and (y) knowledge integration (as measured by integrative elaborations across domains), on the other. In addition, the correlational results extended the argument structure hypothesis in that reasons were also found to be related to students' knowledge integration. Together these results indicate that the more claims, reasons, and warrants pre-service teachers generate and include in an argumentative essay about how the topics of domain-specific study texts relate to each other as regards teaching, the more effective they are in integrating CK, PK, and PCK. Concerning the quality dimension hypothesis, the correlational results provided only limited support, that is, connections and causality in argumentative essays was the only quality dimension related to students' knowledge integration across domains. Other than expected, plausibility, structure and coherence, and focus and clarity were not significantly correlated with students' knowledge integration across domains.

Due to the many significant bivariate correlations among the different argument variables (both structural and conceptual/epistemological), the hypotheses were further tested simultaneously by way of a multiple regression analysis with backward elimination. This enabled adjusting for potentially confounding effects in the bivariate zero-order analyses. Significant regressions were identified for the first model, and for the final model after four predictor variables were eliminated. The remaining predictors in the final model were ‘warrant’ and ‘connections and causality’, both with a medium-sized relation to the criterion. Thus, the results provided partial support for the argument structure and the quality dimension hypothesis in that warrants (as a structural argument component) and connections and causality (as a quality dimension of written arguments) predicted pre-service teachers’ knowledge integration across the core domains of their professional knowledge (*i.e.*, CK, PK, and PCK) when learning from multiple texts.

With regard the provision of warrants, the study suggests that becoming aware of unwarranted (proposed) relations between CK, PK, and PCK knowledge entities and generating warrants to link claims and reasons is central to pre-service teachers’ argumentative knowledge integration. This is in line with prior studies on the cognitive foundation of the competencies involved in argumentation. For example, Garcia-Mila and Andersen (2007) propose that an argument “is a justified assertion whose validity is provided by the coherence of the justification [...]. Within this coherence, ‘warrants’ play a central role in the justification by connecting data with claims” (p. 31). Note that students can incorporate warrants in various ways such as including references to empirical evidence and hypothetical or theoretical ideas (Kelly et al., 1998). However, claims are often only justified by students through providing warrants when their ideas about something (*e.g.*, effective teaching) are challenged. Sometimes even challenging their ideas is not enough (Jiménez-Aleixandre et al., 2000; Sandoval & Millwood, 2007). Hence, additional instructional scaffolds such as explicit task criteria (which enables students to carefully interpret the demands of the task and judge their own performance) and/or prompts (which provide brief assistance by drawing students’ focus to aspects or processes that facilitate task completion and learning) appear to be valid means to promote pre-service teachers’ integration of CK, PK, and PCK. One can thus recommend that both task criteria and prompts are designed towards assisting pre-service teachers’ generation and provision of warrants. That is, these instructional scaffolds should aim at helping pre-service teachers to reason about

the connections between claims and reasons, and enunciate them in their written texts. While an according task criterium might only be beneficial for students with more developed self-regulated learning and argumentation skills (Butler & Cartier, 2004), prompts in the form guiding questions can be realized to stimulate specific learning processes (Lehmann et al., 2019) such as reasoning about how claims and reasons are connected or to what extent a reason justifies a claim. Such prompts could complement integration prompts in the form of guiding questions, which are directed towards knowledge integration (and not towards argumentation).

Moreover, both these means, task criteria and prompts, could at the same time contribute to improve students' argumentative essays in terms of the quality dimension 'connections and causality' (which was found to be another significant predictor of pre-service teachers' knowledge integration). This interpretation is justified by the function of warrants, that is, they serve as coherence-maker and indicate whether a conclusion can be justified given the data (*cf.* Garcia-Mila & Andersen, 2007; Toulmin, 1958). Thus, another instructional strategy, which is worth to be considered in future research, is to imbed argument writing tasks in collaborative (online) learning tasks/environments. This appears appropriate since collaborative writing comes with a complexity that requires not only continuous dialogue and cooperation to come to an agreement about various aspects of the process and the product (*e.g.*, the contents, rhetorical and formal aspects, roles and responsibilities, etc.) but particularly because "the resulting dialogue and the resolution of tensions shared by different voices are the main mechanisms to create new meaning and to promote learning" (Corcelles & Castelló, 2015, p. 158; *see also* Keil et al., 2004; Miller et al., 2014).

Of course, the present study has several limitations that require consideration. First, the main limitation is a methodical. Although the study can be considered externally valid due to the use of authentic learning tasks and material, it followed a single group correlational design to understand the predictors of knowledge integration across domains. Thus, the findings are based on a non-experimental design which does not allow any causal conclusions, and hence, limits internal validity. A second constraint is that the sample consisted exclusively of pre-service elementary mathematics teachers. This limits the generalizability of the results. Therefore, it is recommended to extend this study to secondary pre-service teachers and pre-service teachers with different subject-related foci (*e.g.*, language, sciences, art, sports, etc.). For instance, students who are more often

required to write essays (*e.g.*, students in the humanities) might apply different strategies in self-regulated *writing-to-learn* settings. According studies could yield different results, hence, leading to domain-specific instructional scaffolds to support argumentative writing for knowledge integration. Third, other potentially important moderator or mediator variables, such as reading comprehension and reasoning skills, argument writing experience, motivation, the use of cognitive and metacognitive learning strategies were not examined. Only the extents of domain-specific knowledge in CK, PK, and PCK were entered in the regression analysis in addition to the argument variables. Yet, they remained insignificant predictors. This is somewhat surprising considering that Gil et al (2010b) found “that only students with high prior knowledge were able to take advantage of instructions to construct arguments while reading, whereas low-knowledge readers seemed to be more hindered than helped by such task instructions” (p. 157). A greater focus on students’ internal preconditions of successful argumentative reasoning/ writing for knowledge integration could produce interesting findings that contribute to the present understanding. Finally, it should be noted that knowledge integration was measured by students’ generation and provision of integrative elaborations in their argumentative essays. Admittedly, different methodological approaches can lead to different findings (*see* Lehmann et al., 2020). Hence, future studies might apply different methods to assess and analyze the cognitive integration of domain-specific knowledge and ideas.

Acknowledgment

The author thanks his student research assistants Imke-Silja Hintze, Therese Papperitz and Annika Beihofer for their help with the preparation of the data and different practical aspects of the study.

References

- Anderson, R. C., Nguyen-Jahiel, K., McNurlen, B., Archodidou, A., Kim, S., Reznitskaya, A., et al. (2001). The snowball phenomenon: Spread of ways of talking and ways of thinking across groups of children. *Cognition and Instruction*, 19(1), 1–46. https://doi.org/10.1207/S1532690XCII901_1

- Ball, D. L. (2000). Bridging practices. Intertwining content and pedagogy in teaching and learning to teach. *Journal of Teacher Education*, 51(3), 241–247. <https://doi.org/10.1177/0022487100051003013>
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften [Keyword: Professional competencies of teachers]. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520. <https://doi.org/10.1007/s11618-006-0165-2>
- Bigot, L. L., & Rouet, J.-F. (2007). The impact of presentation format, task assignment, and prior knowledge on students' comprehension of multiple online documents. *Journal of Literacy Research*, 39(4), 445–470. <https://doi.org/10.1080/10862960701675317>
- Bråten, I., & Strømsø, H. I. (2009). Effects of task instruction and personal epistemology on the understanding of multiple texts about climate change. *Discourse Processes*, 47(1), 1–37.
- Brunner, E. (2014). *Mathematisches Argumentieren, Begründen und Beweisen. Grundlagen, Befunde und Konzepte [Mathematical reasoning, justifying, and proving. Basics, findings, and concepts]*. Berlin: Springer Spektrum.
- Butler, D. L., & Cartier, S. C. (2004). Promoting effective task interpretation as an important work habit: A key to successful teaching and learning. *Teachers College Record*, 106(9), 1729–1758.
- Chi, M. T. H., DeLeeuw, N., Chiu, M., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18(3), 439–477. https://doi.org/10.1207/s15516709cog1803_3
- Clark, D. B., & Linn, M. C. (2013). The knowledge integration perspective: Connections across research and education. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 520–538). New York, NY: Routledge.
- Clift, R. T., Ghatala, E. S., Naus, M. M., & Poole, J. (1990). Exploring teachers' knowledge of strategic study activity. *The Journal of Experimental Education*, 58(4), 253–263.
- Corcelles, M. & Castelló, M. (2015). Learning philosophical thinking through collaborative writing in secondary education. *Journal of Writing Research*, 7(1), 157–200. <https://doi.org/10.17239/jowr-2015.07.01.09>
- Darling-Hammond, L. (2006). Constructing 21st-Century teacher education. *Journal of Teacher Education*, 57(3), 300–314.

- diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction, 10*(2-3), 105–225.
- diSessa, A. A. (2002) Why ‘conceptual ecology’ is a good idea. In M. Limon & L. Mason (Eds.), *Reconsidering conceptual change. Issues in theory and practice* (pp. 29–60). Dordrecht: Kluwer Academic Publishers.
- Ferretti, R. P., MacArthur, C. A., & Dowdy, N. S. (2000). The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normally achieving peers. *Journal of Educational Psychology, 92*(4), 694–702. <https://doi.org/10.1037/0022-0663.92.4.694>
- Garcia-Mila, M., & Andersen, C. (2007). Cognitive foundations of learning argumentation. In S. Erduran & M. P. Jiménez-Aleixandre (Eds.), *Argumentation in Science Education* (pp. 29–45). Dordrecht: Springer. https://doi.org/10.1007/978-1-4020-6670-2_2
- Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, H. I. (2010a). Understanding and integrating multiple science texts: Summary tasks are sometimes better than argument tasks. *Reading Psychology, 31*(1), 30–68. <https://doi.org/10.1080/02702710902733600>
- Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, H. I. (2010b). Summary versus argument tasks when working with multiple documents: Which is better for whom? *Contemporary Educational Psychology, 35*(3), 157–173. <https://doi.org/10.1016/j.cedpsych.2009.11.002>
- Glogger-Frey, I., Ampatziadis, Y., Ohst, A., & Renkl, A. (2018). Future teachers’ knowledge about learning strategies: Misconcepts and knowledge-in-pieces. *Thinking Skills and Creativity, 28*, 41–55.
- Grieser, D. (2015). *Analysis I*. Wiesbaden: Springer.
- Hidi, S., Berndorff, D., & Ainley, M. (2002). Children’s argument writing, interest and self-efficacy: an intervention study. *Learning and Instruction, 12*(4), 429–446. [https://doi.org/10.1016/S0959-4752\(01\)00009-3](https://doi.org/10.1016/S0959-4752(01)00009-3)
- Hornikx, J. & Hahn, U. (2012). Reasoning and argumentation: Towards an integrated psychology of argumentation. *Thinking & Reasoning, 18*(3), 225–243. <https://doi.org/10.1080/13546783.2012.674715>
- Jiménez-Aleixandre, M. P., Rodríguez, A. B., & Duschl, R. A. (2000). “Doing the lesson” or “doing science”: Argument in high school genetics. *Science Education, 84*(6), 314–342.
- Keil, F. C., Rozenblit, L., & Mills, C. M. (2004). What lies beneath? Understanding the limits of understanding. In D. T. Levin (Ed.), *Thinking*

- and Seeing: Visual Metacognition in Adults and Children* (pp. 227–249). Cambridge, MA: MIT Press.
- Kelly, G. J., Druker, S., & Chen, C. (1998). Students' reasoning about electricity: Combining performance assessments with argumentation analysis. *International Journal of Science Education*, 20(7), 849–871. <https://doi.org/10.1080/0950069980200707>
- Lehmann, T. (Ed.) (2020a). *International Perspectives on Knowledge Integration: Theory, Research, and Good Practice in Pre-service Teacher and Higher Education*. Leiden: Brill | Sense. <https://doi.org/10.1163/9789004429499>
- Lehmann, T. (2020b). What is Knowledge Integration of Multiple Domains and How Does it Relate to Teachers' Professional Competence? In T. Lehmann (Ed.), *International Perspectives on Knowledge Integration: Theory, Research, and Good Practice in Pre-service Teacher and Higher Education* (pp. 9–29). Leiden: Brill | Sense. https://doi.org/10.1163/9789004429499_002
- Lehmann, T., & Schmidt-Borcherding, F. (2018). Pre-service Teachers' Provision of Reasons and Warrants Relates to Their Integration of Multiple Knowledge Domains in Argumentative Writing. *Proceedings of the 2018 annual meeting of the American Educational Research Association*. New York, NY: AERA Paper Repository.
- Lehmann, T., Pirnay-Dummer, P., & Schmidt-Borcherding, F. (2020). Fostering integrated mental models of different professional knowledge domains: Instructional approaches and model-based analyses. *Educational Technology Research and Development*, 68(3), 905–927. <https://doi.org/10.1007/s11423-019-09704-0>
- Lehmann, T., Rott, B., & Schmidt-Borcherding, F. (2019). Promoting pre-service teachers' integration of professional knowledge: effects of writing tasks and prompts on learning from multiple documents. *Instructional Science*, 41(1), 99–126. <https://doi.org/10.1007/s11251-018-9472-2>
- Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processing during comprehension. *Journal of Educational Psychology*, 91, 615–629. <https://doi.org/10.1037/0022-0663.91.4.615>
- Mateos, M., Martín, E., Cuevas, I., Villalón, R., Martínez, I., & González-Lamas, J. (2018). Improving written argumentative synthesis by teaching the

- integration of conflicting information from multiple sources. *Cognition and Instruction*, 36(2), 119–138.
- Miller, B. W., Anderson, R. C., Morris, J., Lin, T. J., Jadallah, M., & Sun, J. (2014). The effects of reading to prepare for argumentative discussion on cognitive engagement and conceptual growth. *Learning and Instruction*, 33, 67–80. <https://doi.org/10.1016/j.learninstruc.2014.04.003>
- Mishra, P., & Brewer, W. F. (2003). Theories as a form of mental representation and their role in the recall of text information. *Contemporary Educational Psychology*, 28(3), 277–303. [https://doi.org/10.1016/S0361-476X\(02\)00040-1](https://doi.org/10.1016/S0361-476X(02)00040-1)
- Page-Voth, V., & Graham, S. (1999). Effects of goal setting and strategy use on the writing performance and self-efficacy of students with writing and learning problems. *Journal of Educational Psychology*, 91(2), 230–240. <https://doi.org/10.1037/0022-0663.91.2.230>
- Piaget, J. (1928). *Judgement and Reasoning in the Child*. London, UK: Routledge.
- Pintrich, P. R., Smith, D., Garcia, T., & McKeachie, W. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement*, 53(3), 801–813. <https://doi.org/10.1177/0013164493053003024>
- Renkl, A. (2015). *Wissenserwerb [Knowledge acquisition]*. In E. Wild & J. Möller (Eds.), *Pädagogische Psychologie [Educational psychology]* (pp. 3–24). Berlin: Springer.
- Renkl, A., Mandl, H., & Gruber, H. (1996). Inert knowledge: Analyses and remedies. *Educational Psychologist*, 31(2), 115–121.
- Reznitskaya, A., Kuo, L., Glina, M., & Anderson, R. C. (2009). Measuring argumentative reasoning: What's behind the numbers? *Learning and Individual Differences*, 19(2), 219–224.
- Sampson, V. D., & Clark, D. B. (2006). Assessment of argument in science education: A critical review of the literature. *Proceedings of the 7th International Conference on the Learning Sciences* (pp. 655–661). Bloomington, IN: International Society of the Learning Sciences.
- Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *Journal of the Learning Sciences*, 12(1), 5–51. https://doi.org/10.1207/S15327809JLS1201_2

- Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction*, 23(1), 23–55. https://doi.org/10.1207/s1532690xci2301_2
- Sandoval, W. A., & Millwood, K. A. (2007). What can Argumentation tell us about epistemology? In S. Erduran & M. P. Jiménez-Aleixandre (Eds.), *Argumentation in Science Education* (pp. 71–90). Dordrecht: Springer. https://doi.org/10.1007/978-1-4020-6670-2_4
- Schneider, M. (2012). Knowledge integration. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 1684–1686). New York, NY: Springer. https://doi.org/10.1007/978-1-4419-1428-6_807
- Schwarz, B. B. (2009). Argumentation and learning. In N. Muller-Mirza & A.-N. Perret-Clermont (Eds.), *Argumentation and Education. Theoretical Foundation and Practices* (pp. 91–126). Dordrecht: Springer.
- Simon, S. (2008). Using Toulmin's argument pattern in the evaluation of argumentation in school science. *International Journal of Research & Method in Education*, 31(3), 277–289.
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge, UK: Cambridge University Press.
- Varghese, S. A., & Abraham, S. A. (1998). Undergraduates arguing a case. *Journal of Second Language Writing*, 7(3), 287–306.
- Wagner, J. F. (2006). Transfer in pieces. *Cognition and Instruction*, 20(1), 1–71. https://doi.org/10.1207/s1532690xci2401_1
- Wäschle, K., Lehmann, T., Brauch, N., & Nückles, M. (2015). Prompted journal writing supports preservice history teachers in drawing on multiple knowledge domains for designing learning tasks. *Peabody Journal of Education*, 90(4), 546–559.
- Webb, N. M., Troper, J. D., & Fall, R. (1995). Constructive activity and learning in collaborative small groups. *Journal of Educational Psychology*, 87(3), 406–423. <https://doi.org/10.1037/0022-0663.87.3.406>
- Wiley, J., & Voss, J. F. (1996). The effects of 'playing historian' on learning in history. *Applied Cognitive Psychology*, 10(7), 63–72.
- Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory from text. *Journal of Educational Psychology*, 91(2), 301–311. <https://doi.org/10.1037/0022-0663.91.2.301>

- Yeh, S. S. (1998a). Empowering education: Teaching argumentative writing to cultural minority middle-school students. *Research in the Teaching of English*, 33(1), 49–83.
- Yeh, S. S. (1998b). Validation of a scheme for assessing argumentative writing of middle school students. *Assessing Writing*, 5(1), 123–150. [https://doi.org/10.1016/S1075-2935\(99\)80009-9](https://doi.org/10.1016/S1075-2935(99)80009-9)
- Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, 39(1), 35–62. <https://doi.org/10.1002/tea.10008>

Received June 5, 2020
Revision September 1, 2020
Accepted October 29, 2020

Appendix

Appendix A

Scoring scheme for assessing different quality dimensions of students' argumentative essays

Quality dimension	Score	Criterion description	ICC
Connections and causality	4	The essay exhibits definite, well-developed connections and explicates content-related causality. Statements are presented in relation to other statements. Potential links are realized and further discussed. No or only very few statements remain unconnected.	.86
	3	The essay explicates connections and make repeated attempts to indicate causality. Most statements are presented in relation to other statements. Potential links are realized. Only few statements remain unconnected.	
	2	The essay makes little attempt to generate connections and display causality, involves solely undeveloped connections, and lacks causality. Different statements are sporadically presented in relation to other statements and potential links are overlooked.	
	1	The essay hardly involves consideration to generate connections and display causality and clearly lacks both connections and causality. Different statements are largely presented in an isolated manner and remain unrelated to other statements.	
	0	The essay makes no attempt to generate connections and display causality at all. It comprises no connections and causality and/or does not provide a sufficient basis for a reliable rating.	
Plausibility (content validity)	4	The essay is plausible and shows no lack in content validity. The statements included are correct and scientifically reasonable throughout the essay.	.85
	3	The essay appears rather plausible and valid. Most of the statements included in the essay are correct and scientifically reasonable.	
	2	The essay lacks plausibility. Multiple statements included in the essay are incorrect and many of them are unscientific or scientifically weak.	
	1	The essay clearly lacks plausibility. The essay involves many incorrect and/or scientifically questionable or irrational statements.	
	0	The essay is not plausible at all and/or does not provide a sufficient basis for a reliable rating.	
Structure and coherence	4	The essay is clearly structured and coherent, that is, sentences and paragraphs build on each other through linking words or ideas. Structural inconsistencies or deficiencies in coherence occur very rarely (or never).	.81
	3	The essay appears rather structured and coherent, that is, most statements build on each other explicitly and implicitly. However, few structural inconsistencies or deficiencies in coherence are also included.	

Scoring scheme for assessing different quality dimensions of students' argumentative essays - *continued*

Quality dimension	Score	Criterion description	ICC
Variation of different argument elements	2	The essay appears rather unstructured and incoherent. Many sentences and paragraphs do not build on each other. If they do, this happens rather implicitly. Structural inconsistencies or deficiencies in coherence occur more frequently.	.82
	1	The essay is unstructured and incoherent. The sentences and paragraphs do not build on each other, or only in exceptional cases. There are several structural inconsistencies or deficiencies in coherence.	
	0	The essay clearly lacks structure and coherence and/or does not provide a sufficient basis for a reliable rating.	
	4	The essay contains more than four of the structural elements of an argument (claim, reasons, warrant, qualifier, backing, rebuttal, counter-claim).	
	3	The essay contains four of the structural elements of an argument (claim, reasons, warrant, qualifier, backing, rebuttal, counter-claim).	
	2	The essay contains three of the structural elements of an argument (claim, reasons, warrant, qualifier, backing, rebuttal, counter-claim).	
	1	The essay contains two of the structural elements of an argument (claim, reasons, warrant, qualifier, backing, rebuttal, counter-claim).	
Explicit differentiation of argument elements	0	The essay clearly lacks using a variety of argument elements and/or does not provide a sufficient basis for a reliable rating.	.83
	4	Different structural argument elements are very often (or always) explicitly distinguished, precisely formulated, and thus, clearly recognizable.	
	3	Different structural argument elements are rather frequently explicitly distinguished, precisely formulated, and thus, clearly recognizable.	
	2	Different structural argument elements are sporadically explicitly distinguished, precisely formulated, and thus, clearly recognizable.	
	1	Different structural argument elements are very rarely (or never) explicitly distinguished, precisely formulated, and thus, clearly recognizable.	
Focus and clarity	0	The essay clearly lacks an explicit differentiation of argument elements and/or does not provide a sufficient basis for a reliable rating.	.86
	4	The statements and paragraphs are clear and focused throughout the essay. A reasonably developed basic idea and common thread is explicitly formulated. There are no missing definitions, elaborations, and illustrations.	
	3	Most statements and paragraphs are clear and focused. A reasonably developed basic idea and common thread is recognizable. Only occasionally, clarity could be improved through further definition, elaboration, and illustration.	
	2	Several statements and paragraphs are rather unclear or unfocused. The essay lacks to explicitly display a reasonably developed basic idea and common thread but includes them implicitly. The clarity could be improved through further definition, elaboration, and illustration.	

Scoring scheme for assessing different quality dimensions of students' argumentative essays - *continued*

Quality dimension	Score	Criterion description	ICC
Voice (use of language for rhetorical effects)	1	Many statements and paragraphs are unclear and unfocused. The essay lacks to explicitly display or implicitly include a reasonably developed basic idea and common thread. The reader struggles to not get lost in various aspects that adhere to a common line of thought.	.88
	0	The essay completely lacks focus and clarity and/or does not provide a sufficient basis for a reliable rating.	
	4	The essay maintains the argumentative text type throughout or to a very large extent. The use of language is sophisticated and characterized by very good comprehensibility and appropriate complexity. The terminology meets scientific requirements.	
	3	The essay maintains the argumentative text type to a rather large extent. The use of language is mature and characterized by good comprehensibility and rather appropriate complexity with only few exceptions. The terminology tends to meet scientific requirements.	
	2	The essay does rather not maintain the argumentative text type. The use of language is moderately immature and lacks comprehensibility or complexity. The terminology is rather unscientific.	
Adherence to language conventions	1	The essay does not maintain the argumentative text type. The use of language is immature and clearly lacks both comprehensibility and complexity. The terminology is unscientific.	.85
	0	The essay clearly lacks an appropriate use of language and/or does not provide a sufficient basis for a reliable rating.	
	4	The essay is virtually error-free and involves, if any, only very few errors in grammar, punctuation, and spelling.	
	3	The essay involves infrequent errors in grammar, punctuation, and spelling. Errors do not affect the understanding of the text.	
	2	The essay involves occasional errors in grammar, punctuation, and spelling. Sometimes the errors hinder the understanding of the text.	
	1	The essay involves many errors in grammar, punctuation, and spelling, which confuse the reader of the text.	
	0	The essay involves continual errors in grammar, punctuation, and spelling. It clearly lacks an adherence to language conventions and/or does not provide a sufficient basis for a reliable rating.	