Jahrgang 30 / Heft 2–3 / 2016

Zeitschrift für

# Pädagogische Psychologie

German Journal of Educational Psychology



## Themenheft/Special Issue

Learning through Communication

hogrefe

# Inhalt

Zeitschrift für

# Pädagogische Psychologie



hogrefe

Original Article

# How Dialogic Settings Influence Evidence Use in Adolescent Students

Elisabeth Mayweg-Paus[1] & Fabrizio Macagno[2]

[1] Department of Psychology and Sport Studies, University of Münster, Germany
[2] Faculdade de Ciências Sociais e Humanas, Universidade Nova de Lisboa, Portugal

**Abstract:** This study examines how evidence is used differently in argumentative discourse compared to individual arguments. Applying a 1 × 2 crossover study design, 37 secondary school students were asked either to discuss a social issue with their partner before individually writing an essay outlining their opinion or, vice versa, first to discuss and then to write. As background information, they were provided with pieces of evidence with different levels of quality. Dialogs and essays were analyzed regarding (a) the type of evidence and (b) the way evidence was used. Results showed that in their essays students referred more often to the pieces of evidence provided to them (shared evidence). In contrast, they used evidence more often to address the opposing viewpoint in dialogs by incorporating it in a more elaborated (clearer) line of reasoning. Findings suggest that dialogues are a more effective tool than individual writing production, and the study provides first hints regarding how to design curricula that will encourage students to use evidence in a more sophisticated way in their argumentation.

**Keywords:** argumentative dialogue, individual argument, argumentative function, argumentative structure, evidence use

**Welchen Einfluss haben dialogische Lernsituationen auf den Gebrauch von Evidenz bei adoleszenten Schülerinnen und Schülern?**

**Zusammenfassung:** Diese Studie untersucht wie sich argumentativer Diskurs und individuelles Argumentieren in Bezug auf den Gebrauch von Evidenz unterscheiden. In einem 1 × 2 Cross-over Design diskutierten 37 Mittelstufenschülerinnen und –Schüler ein gesellschaftliches Thema mit ihrem Partner, entweder bevor oder nachdem sie einen kurzen Aufsatz zu ihrer eigenen Meinung verfassten. Als Hintergrundinformationen erhielten sie eine Sammlung qualitativ unterschiedlicher Evidenzen zu dem Themenbereich. Die Dialoge und Aufsätze wurden untersucht in Hinblick auf a) die Art der Evidenz und b) auf welche Weise diese genutzt wurde. Die Ergebnisse zeigen, dass sich die Schülerinnen und Schüler in den Aufsätzen häufiger auf die ihnen gemeinsam vorliegenden Evidenzen beziehen (geteilte Evidenz). In den Dialogen nutzen sie Evidenz hingegen häufiger, um den gegenteiligen Standpunkt zu adressieren und zeigen dabei eine klarere Argumentationslinie. Die Ergebnisse weisen auf eine höhere Effizienz der Dialoge im Vergleich zum individuellen Schreiben hin. Gleichzeitig gibt die Studie erste Hinweise für die Gestaltung von Curricula, die Schülerinnen und Schüler dazu anregen Evidenz in ihrer Argumentation einzusetzen.

**Schlüsselwörter:** argumentativer Diskurs, Evidenz, argumentative Funktion

## Introduction

In recent decades, argumentation has come to be viewed as a central component of education (Driver, Newton & Osborne, 2000; Duschl & Osborne, 2002; Kuhn, 1993; Rapanta, Garcia-Mila & Gilabert, 2013). According to this view, one of education's most important goals is to provide learners with the capabilities to assess the available information, to select the most relevant and adequate evidence and to apply this in their judgments and decisions (Osborne, Eduran, Simon & Monk, 2001). The use of evidence for argumentative purposes (namely, to support, assess, question or refute a claim) is particularly informative, because it reveals a student's epistemological background (Sandoval & Millwood, 2005). The interpretation, assessment and justification of a piece of evidence, or rather the critical stance in relation to it, is mirrored by how a student coordinates and connects it with given claims and other concurring or conflicting evidence (Kuhn 1993; Kuhn, Amsel & O'Loughlin, 1988; Kuhn et al., 1995). In contrast, a student's failure to use evidence in a way that distinguishes it from a specific claim (Jiménez-Aleixandre, Rodríguez & Duschl, 2000; McNeill, 2011) and is functional in relation to that claim (McNeill, Lizotte, Krajcik & Marx, 2006) reveals problems in interpreting such evidence, in distinguishing it from an advocated viewpoint and in addressing it critically. Similarly, a student's epistemological commitment is mirrored by the way she uses such information to build an argument and defend or attack a viewpoint (Rapanta et al., 2013; Sandoval & Millwood, 2005). The quality of understanding is thus related to the way a student interprets evidence in order to use it argumentatively (Sandoval & Millwood, 2005). The dialectical assessment of a piece of information through arguments, questioning




DOI 10.1024/1010-0652/a000171

and attacks is related crucially to its understanding (Nussbaum & Edwards, 2011). The crucial issue is understanding how to improve students' use of evidence in their explicit reasoning in order to design strategies that will help them to adequately integrate evidence in their line of reasoning. Our study addresses this question by testing whether and how the argumentative mode (dialogic vs. monologic) affects the use of evidence from a functional and structural perspective.

## Dialogic argumentation

Dialogic argumentation is viewed as a process in which two or more people engage in a debate characterized by opposing claims (Kuhn & Udell, 2003) that they need to support with arguments and question critically (Walton, 1989). In an argumentative dialogue, a participant is subject to the interlocutor's scrutiny of her own position. This scrutiny and the efforts to challenge the opposing view can be assumed to impact directly on the quality of argumentation. In a dialogic argumentative setting, a student needs to be more critical regarding not only her own position but also the opposing one. She should achieve this by drawing on more evidence and elaborating it further in order to address the challenges to her own position in a relevant way (Walton & Macagno, 2007). In dialogues, students need to analyze the reasons for preferring one point of view or one piece of evidence over another. Therefore, they are encouraged to take a critical stance toward the presented evidence (Osborne et al., 2001). They should engage directly in understanding and assessing their interlocutor's reasons, thereby elaborating more and in greater depth on their own and the other's point of view.

Argumentative dialogues are considered to be an effective educational strategy for developing reasoning skills (Koballa, 1992; Kuhn, 1992; Kuhn & Crowell, 2011). Indeed, argumentative dialogues have been shown to have a positive effect on the ability to generate counterarguments and rebuttals (Kuhn, Goh, Iordanou & Shaenfield, 2008). Kuhn and Moore (2015) have shown that students who were taught within a 2-year dialogue-focused curriculum (see also Kuhn, 2015) tended to back up their claims with more evidence than students belonging to a nondialogic group. Moreover, the same intervention revealed that students engaging in a dialogic setting were more likely to draw on evidence from their own personal knowledge (instead of from a list provided to them) and to address and weaken an opposing claim instead of using evidence only to support their own view.

This study raises two crucial questions: (a) How can the use of evidence serve as an indicator of quality of argumentation? (b) How does the dialogic setting impact on the quality of evidence use? Although, as indicated above, many studies have investigated the use of evidence in educational contexts, it is not entirely clear *what* type of evidence they used and *how* they used it. To address these issues, we designed a study to cast light on the relationship between the various types of uses of evidence and their argumentative quality and to test the differences between a dialogic and a monological setting. To do this, we needed to draw some basic distinctions on both a content level (namely between various types of evidence) and a functional level (namely between various types of argumentative uses of evidence).

The first distinctions concern what type of evidence is taken into account. Some studies refer solely to scientific evidence (such as empirical data from experiments and surveys, see Iordanou & Constantinou, 2014, 2015; Sandoval, 2003; Sandoval & Millwood, 2005); others also include evidence drawn from single cases (nonscientific evidence, see Kuhn & Moore, 2015). Therefore, we decided to provide students with a broader range of five subcategories of evidence. To determine which evidence is used, we not only considered these distinct types but also analyzed where the evidence came from. In particular, we distinguished between evidence that we provided as input on cards and evidence stemming from students' personal experiences or knowledge. This enabled us to test not only the effect of the argumentative mode on the quality of evidence per se but also how it might trigger different sources of evidence.

The second distinction concerns how evidence is used for argumentative purposes. Our basic assumption is that a critical understanding of evidence can be revealed through its argumentative uses, or more specifically through a distinction between more and less elaborated (or sophisticated) uses to support a viewpoint or rebut an incompatible or different one. Hence, we distinguished the types of evidence use on the basis of the structure of an argument and the possible functions of a piece of evidence within that argument (Pollock, 1974; Toulmin, 1958; Walton, 2006). Such functions were then compared and assessed in terms of which uses were more or less sophisticated. This served as the basis for developing a specific coding scheme.

## Function and structure of evidence use

From an argumentative point of view, evidence serves two fundamental functions (or purposes): It can be used to support one's own viewpoint or to weaken a position that is incompatible or different from the one defended (the interlocutor's perspective). From a structural point of view, evidence can support or weaken a position in different ways. These can be distinguished by drawing on Toulmin's

 

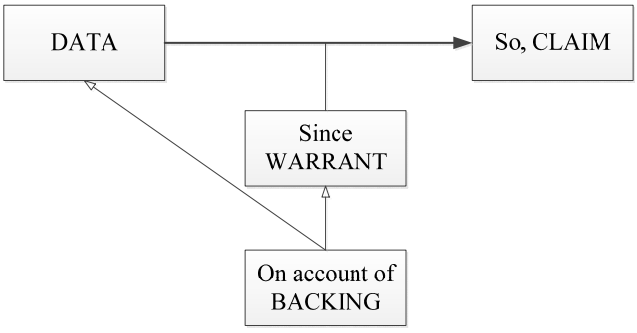


**Figure 1.** Toulmin's structure of argument analysis.

(1958) argument structure. According to Toulmin, an argument can be represented as an interconnected set of (a) a claim (C), (b) data (D), (c) a warrant connecting claim and data ("since W") and (d) backings (B) substantiating the warrants. The structure of Toulmin's model as applied in our research is depicted in Figure 1.

This reading of Toulmin's scheme aims to distinguish between two distinct argumentative uses of evidence that relate to the structure of an argument. The *data* are considered as the premise (or the premises) that are linked directly to the claim, providing the factual and relevant reason to draw the conclusion (the claim). The *backing* instead, is conceived as the support that can be given to an element grounding the claim. This is usually the warrant, but can also be the data. In this sense, the backing is not essential for the logical and material support of the claim. It reinforces the elements supporting it.

In this interpretation of Toulmin's model, the conclusion is grounded on premises (data and warrant) that are an interpretation of the available information. This casts light on which elements thereof are relevant to the conclusion. For example, in order to support the claim that "smoking should be banned," an arguer may use the available evidence that "major studies indicate a relation between smoking cigarettes and lung disease." However, this piece of information needs to be interpreted in order to be used to support the intended conclusion, thereby casting light on which aspect thereof is relevant in this case. Therefore, the arguer will point out that because smoking affects health (or more specifically lung health), it should be forbidden. This data, namely the interpretation used as a premise, allows the triggering of a commonly shared warrant from negative consequences (whatever affects health should be forbidden). The piece of evidence is then used to back up the data. In this view, evidence is used in a relevant argumentative way to support the desired conclusion after being interpreted accordingly. Otherwise, it can be used simply as a direct support of a claim, but no relation of relevance to the warrant is brought to light. For example, the arguer may simply say that "smoking should be banned, because major studies indicate a relation between smoking cigarettes and lung health." This uninterpreted piece of evidence leaves the interlocutor with the burden of retrieving the intended relation to the conclusion, but also the possibility of rejecting the argument directly (a "relation" means that no evidence of an actual cause has been found). Whereas a piece of data can be supported by several backings (thus saving the argument from one of the possible criticisms of the evidence cited), when evidence is simply cited as a data, the whole argument is exposed to default if such evidence is rebutted by a conflicting interpretation (which applies even more particularly for neutral pieces of evidence). Uninterpreted evidence thus makes the process of argument reconstruction much more complex and less clear, and grants the interlocutor the possibility of interpreting it strategically (Macagno, 2008, 2012).

For these reasons, we distinguished between two levels of evidence use differentiated according to their argumentative relevance:

1. First-order evidence: Evidence used to support or challenge a claim directly in an argument (functioning as Toulmin's data).
2. Second-order evidence: Evidence used to support or challenge the validity of how data (first-order evidence) is used to support a claim in an argument (functioning as Toulmin's backing).

Apart from the dimensions of validity and acceptability, an argument can be evaluated by considering the two dialectical criteria of clarity (see Aristotle, trans. 1995) and effectiveness (considered as diminishing the possibilities of possible attacks) (Macagno & Walton, 2014). Hence, the arguments leaving less room (and less of a burden) of interpretation to the interlocutor should be considered as both clearer and more effective. Based on this, we can claim that the most sophisticated use of evidence corresponds to its more elaborated use as a backing, which presupposes its interpretation and its consequent reasoned and critical use in a way that is significant for the conclusion (Sandoval & Millwood, 2005).

Evidence use can be divided into four categories that combine the argumentative purpose of the use of a piece of evidence (functional level) with its structural role (structural level). A claim can be supported with first- or second-order evidence (evidence used as data or as a backing). Similarly, a position can be attacked with first- or second-order evidence, namely, by using evidence as a backing to support a contrary conclusion (counter-claim), or as an underminer, attacking the foundations of the argument in support of the incompatible view (Macagno, Mayweg-Paus & Kuhn, 2014; Mayweg-Paus, Macagno & Kuhn, 2015). An underminer can weaken either the interpretation of the

evidence used by the interlocutor (namely the data) or the evidence itself. Whereas the first type of refutation can be considered as corresponding to the direct support of a claim, underminers mirror the interpretation and the assessment of the pieces of evidence. For this reason, whereas first-order evidence employed both to support a claim and attack an argument reveals a lower quality of evidence use, first-order evidence used as a backing or an underminer can be considered as a sign of higher quality use.

## Rationale of this study

To identify the specific effect of dialogic interaction on the use of evidence in argumentation, we compared adolescents' arguments on a social issue in a peer-peer argumentative dialogue and in an individual essay. In both cases, students engaged in an argumentative activity (because they needed to provide reasons supporting a claim) derived from an explicit alternative or doubt (in dialogues) or an implicit one (in essays). All participants received a set of cards with short pieces of evidence of different types, and they were told that they could use these if they wanted to. The number of pieces of evidence supporting a position, the opposing position, and a neutral position on the topic were equal.

The aim of this study was twofold: From a content point of view, our goal was to assess the differences between the two argumentative modes relative to the following dimensions: (a) the use of the different types of evidence provided in the cards, and (b) the source of the evidence that the students used in general (whether it was drawn from the cards or from their personal experience or knowledge). From a functional and structural point of view, we intended to test the impact of the dialogic setting on the critical evaluation of evidence and points of view. Therefore, we developed a coding system designed to capture how evidence was used from a functional-structural point of view and what counted as a more elaborated or sophisticated use (Clark, Sampson, Weinberger & Erkens, 2007).

# Method

## Participants and design

Thirty-seven students (54.1 % female) from two biology classes (10th graders) at a public high school in New Jersey participated in the study. Their mean age was 15.62 years ($SD$ = .92). A total of 48.6 % of the participants were Hispanic, 27 % were Black, 18.9 % were White, 2.7 % were Asian, and 2.7 % were American Indian. The two classes were comparable regarding their grade point average (GPA) on a 4-point scale (Class A: $M$ = 2.27, $SD$ = .61, Class B; $M$ = 2.1, $SD$ = .82), $F(1, 35)$ = .522, $p$ = .40, *ns*, as well as on their biology performance (in percentage) in particular (Class A: $M$ = 74.15 %, $SD$ = 13.88 %, Class B: $M$ = 77.52 % $SD$ = 8.62 %), $F(1, 35)$ = .757, $p$ = .30, *ns*. Note that students needed to achieve at least 50 % to pass the biology class. The experimental manipulation was performed in a 1 × 2 crossover design with students writing an individual essay either before or after engaging in mutual discussion with a peer partner.

## Materials and procedure

The topic chosen was "*banning cigarette sales in the US*." Students are likely to have different views on this topic, and, at the same time, they can be expected to be open to considering other possible perspectives on it and to engaging in deeper discussions. Students were given a set of cards containing 15 distinct pieces of evidence selected according to five categories depending on the type of source on which they were based: anecdotal-single case, popularity, authority/expert source, descriptive statistics and laboratory evidence. These five categories were ranked along a dimension ranging from low to high quality. The definition of quality was based on two criteria. The first was purely theoretical, grounded on relevant works in argumentation (Walton, 2006). The second was statistical: We asked a group of 10 experts (academic staff at a German and a Portuguese University (6 male, $M$ = 33.80 years old, $M$ = 7.50 years working in academic context) to rank the five types according to how helpful they would be in an academic discussion. The experts' ranking (from the strongest to the weakest) can be summarized as follows:

1. Laboratory evidence (80 % of the sample. 10 %: statistics; 10 %: authority);
2. Descriptive statistics (80 % of the sample; 10 %: authority; 10 %: anecdotal);
3. Authority (80 % of the sample; 20 %: laboratory evidence);
4. Popularity (70 % of the sample; 30 % anecdotal);
5. Anecdotal (60 % of the sample; 30 %: popularity; 10 %: statistics).

For each category, students received one piece of evidence in favor of banning cigarette sales, one against and one that was neutral. Six pieces of evidence were taken from a list already used in the work of Kuhn and Moore (2015). Four other pieces were slightly modified versions of evidence used by a random sample of 20 laypeople interviewed in New York City's Grand Central Station. The

 

**Table 1.** Evidence list – Classification system.

| Evidence type | Purpose | Example |
|---|---|---|
| Anecdotal (single case)<br>*Reference to a specific case* | Con | Georg Harrison, a musician for the Beatles, was a smoker and died of lung cancer in the age of 58. |
| | Neutral | A man called John Stasser used to smoke, but he was not addicted to cigarettes. He simply enjoyed smoking. |
| | Pro | A woman named Helen Reichert lives in NYC; she is 108 years old and has been smoking half a pack of cigarettes every day for over 80 years. |
| Authority<br>*Reference to a source such as scientists, government, institutes* | Con | The government emphasizes that smoking is one of the factors enhancing the probability of getting heart attacks. |
| | Neutral | Dr. F. R. Moore, president of a major cancer center, claims that smoking, combined with other causes, may lead to lung cancer, but there are many other factors. |
| | Pro | Scientists have shown that a strong genetic influence of people living long and healthy lives regardless whether or not they smoke. |
| Popularity (popular opinion)<br>*Reference to the majority/ common sense* | Con | A survey showed a large majority of people support banning cigarettes sales. |
| | Neutral | A survey showed a large majority of people think smoking is an important issue. |
| | Pro | A survey showed a large majority of people think everybody has the right to decide whether they want to smoke. |
| Laboratory evidence (authority)<br>*Reference to a study by describing the design/findings* | Con | Several major studies indicate a relation between smoking cigarettes and lung disease. |
| | Neutral | Laboratory research shows that people use smoking to help them cope when facing highly stressful situations. |
| | Pro | Studies show the nicotine in cigarettes causes fast-acting chemical reactions in your brain that relieve anxiety and nervousness. |
| Statistics: descriptive (induction)<br>*Reference to numbers and data* | Con | Each year, an estimated 443,000 people die from smoking or exposure to secondhand smoke, and another 8.6 million have to live with a serious illness caused by smoking. |
| | Neutral | An estimated 17 million Americans try to quit smoking each year, and about 8 % of them succeed. |
| | Pro | Thousands of farmers in the U.S. make their living from farming tobacco leaves, and the tobacco industry contributes an average of $ 16.5 billion to the economy each year |

sample covered a wide range of professions (such as dance teachers, physicians, police officers, etc.) and age groups (mean age was 38.32 years, *SD* = 14.90; max: 65, min: 18 years). The remaining five pieces were developed on the basis of an Internet search. Table 1 gives an overview of the 15 pieces of evidence.

The intervention took place within a 90-min school class. First, students were provided with two questions asking for their view on the topic ("Should cigarette sales be banned in the U.S.?" and "How sure are you of your opinion?"). This information enabled us to match students to opposing-site pairs. Then, the cards containing the 15 different pieces of evidence were distributed to them. Students were asked to read the information on the cards carefully and rank the cards according to their strength. In the next step, students in Class A were instructed to write an essay outlining their view on the topic. Students in Class B were randomly assigned to opposing-site pairs, and were asked to discuss the topic verbally. In the last phase, students in Class B were instructed to write an essay and students in Class A were randomly assigned to opposing-site pairs and asked to discuss the topic verbally. All face-to-face dialogs were audiotaped.

## Dependent measures

### Coding of dialog quality

The first step in the coding procedure was to identify the relevant pieces of evidence. Therefore, the dialogs and the essays were first divided into idea units (Asterhan & Schwarz 2009; Jucks & Paus 2013). Idea units were further divided into on-task (addressing the argumentation task) and off-task units. Then, we classified the on-task units into evidence units (namely units that reported or referred to evidence) and nonevidence units. Interrater agreement (calculated on roughly 50 % of both dialogues and essays) was good (in dialogs: Krippendorff's $\alpha$ = .94; in essays:

**Table 2.** What kind of evidence was used?

| Kind of evidence | Example from use in dialog | Example from use in essay |
|---|---|---|
| From cards (α = .85 – .94) | "I am against banning because there are people who are like 100 years old, and sometimes even more, and they have smoked for 80 years and they are still alive." | "I think cigarette sales shouldn't be banned. There is a woman who was 108 years old and who has been smoking half a pack of cigarettes for over 80 years." |
| Personal (α = .80 (dialog), α = .79 (essay)) | "(...) my grandfather has lung cancer, so that's because I am like against, just because of smoking (...) | "I believe, cigarettes sales should be banned. I personally know a young boy who has a hole in his throat since birth. That hole was due to his mother's addiction before and while she was pregnant with him (...) |

Krippendorff's α = .94). In a next step, we coded the on-task units of both essays and dialogs by taking into account two distinct aspects of evidence use: the kind of evidence and the structure and function of evidence use.

### Type of evidence used

We took into account the five different evidence categories from the cards plus the category of personal evidence containing evidence not drawn from the cards but from personal experience or knowledge (personal experience = 95 % and personal knowledge from hearsay = 5 %). Interrater agreement on roughly 50 % of the dialogs and the essays was good for both the classification of the students' evidence in the five categories (Krippendorff's α = .85–94), and the detection of personal evidence in the dialogs (Krippendorff's α = .80) and in the essays (Krippendorff's α = .79). Table 2 presents an overview of these results.

### Structure and function of evidence use

We grounded our analysis on two interrelated dimensions of the use of a piece of evidence (Clark et al., 2007): its function and its structure. For the functional coding, we referred to a coding scheme developed by Kuhn and Moore (2015) that differentiates between evidence advanced to support a claim and evidence used to weaken a claim. Within these two generic goals of evidence use, a further distinction was the way a piece of evidence can be used to support or weaken a claim (first- and second-order evidence). We referred to this distinction as concerning the "structure" of evidence use, and we based our differentiations on Toulmin's (1958) argument pattern. The final coding scheme consisted of four distinct categories (see also Table 3 for examples and interrater reliabilities): (a) support premises, (b) support claim, (c) weaken claim and (d) weaken evidence. *Support premises* is the code for an evidence unit used as a backing in Toulmin's pattern, namely as second-order evidence. *Support claim* is the use of evidence to support a viewpoint directly, namely as first-order evidence. *Weaken claim* represents the code for the second-order evidence units used to counter an argument, namely to attack the conclusion directly or the data on which it is based. *Weaken evidence* is the code for first-order

**Table 3.** Our coding scheme.

| Category | Description | Example |
|---|---|---|
| Support claim (α = .82 (essay), α = .79 (dialog)) | Evidence (italics) is used to support directly the generic viewpoint, but it is not related to an argument. | "My first statement is *a survey that showed a large majority of people support banning cigarettes sales. And a Beatles musician George Harrison was a smoker and died of lung cancer at 58.* I strongly believe cigarettes sales should be banned." |
| Support premises (α = .75 (essay), α = .73 (dialog)) | Evidence is used to back up the premises of an argument (bold), giving strength to it. It is related indirectly to the student's claim, because it supports a line of reasoning. | "There are many reasons to why it should be banned, not only because its bad for you but it can be addicting to people, like an estimated 17 million Americans try to quit smoking each year and about 8 % of the succeed. |
| Weaken claim (α = .81 (essay), α = .82 (dialog)) | Evidence is used to weaken the opposing claim by providing a counter-reason not to accept it. | "A lot of people believe that **smoking can lead to a short life and causes various sicknesses,** but not always. Scientist have shown a strong genetic influence of people living a long healthy life." |
| Weaken evidence (α = .85 (essay), α = .79 (dialog)) | Evidence is used to directly weaken the evidence that supports the opposing view either directly or indirectly (underlined). | Although Helen Reichert is 108 and has been smoking 80 years that doesn't mean nothing. Not every one person is alike. Just like she lived to be 108, *George Harrison only lived to be 58.* |

 

evidence used for attacking directly the interlocutor's evidence used as a backing. In this latter case, the interlocutor's argument is undermined by countering the evidential grounds on which it stands.

## Results

To take random effects into account and increase the generalizability of results, we employed linear mixed-effect models. These provide several advantages (see Barr, Levy, Scheepers & Tily 2013; or Baayen, Davidson & Bates 2008, for details). To determine the potential effects of the argumentative mode (dialog and essay) on the use of evidence and on the function and structure of evidence use in the following analyses, we classified the argumentative mode, the class and the intercept as the fixed effects and the subjects and the dyads as random effects. In particular, we used the factor mode to compare aspects of evidence use of students in both classes in either dialog or essay. We used the factor class to compare aspects of evidence use in both modes between the two classes, whereas we viewed the interaction between the two factors as illustrating an effect of the order of intervention.

The analysis showed that dialogs were longer (number of on-task units) ($M$ = 20.35, $SD$ = 12.30) than essays ($M$ = 10.76; $SD$ = 4.03), $t(71) = -4.01$, $p < .001$, $d = 1.05$. An average of 3.73 ($SD$ = 2.28) on-task units in the dialogs and 2.53 ($SD$ = 1.73) on-task units in the essays referred to evidence, with no difference between the two modes, $t(71) = -1.46$, $p = .15$, *ns*. The following analyses were based on the proportions of each category in the distinct three coding systems on the total number of on-task units referring to evidence. To make the data suitable for further analyses, we then performed a logit-transformation of these proportions and used these transformed values.

**Evidence Use.** In the essays, 81.9 % of the evidence was drawn from the cards and the remaining 18.1 % was personal evidence. Moreover, 26.6 % of the card evidence (representing 81.9 % of the total evidence) was anecdotal, 12.6 % was from authority, 1.9 % was from popularity, 9.0 % was laboratory and 31.9 % was descriptive statistics. In the dialogs, 67.8 % of the evidence was drawn from the evidence cards and the remaining 32.2 % was personal evidence. Here, 23.7 % of the card evidence was anecdotal, 12.5 % was from authority, 3.0 % was from popularity, 5.5 % was laboratory and 23.1 % was descriptive statistics (see also Figure 1).

We then examined whether students' performance in dialogs and essays differed with regard to the use of evidence from the cards. The analysis revealed that students used more evidence from the cards in the essays ($M$ = 2.83, $SD$ = 2.70) than in the dialogs ($M$ = 1.38, $SD$ = 3.10, fixed coefficient: -0.75), $t(71) = 2.30$, $p = .02$, $d = 0.49$. Moreover, there was no specific influence of either the class, $t(71) = 1.03$, $p = .31$, *ns*, or the order of intervention, $t(70) = -0.82$, $p = .41$, *ns*. Furthermore, the amount of evidence used in essays and dialogs did not differ for any of the five predefined subcategories ($p > .2$, *ns*.).

### The function and structure of evidence use

A total of 26.7 % of the evidence in the essays and 30.9 % in the dialogs was second-order evidence used to support the premises of an argument, whereas 60.3 % of the evidence in the essays and 29.5 % in the dialogs corresponded

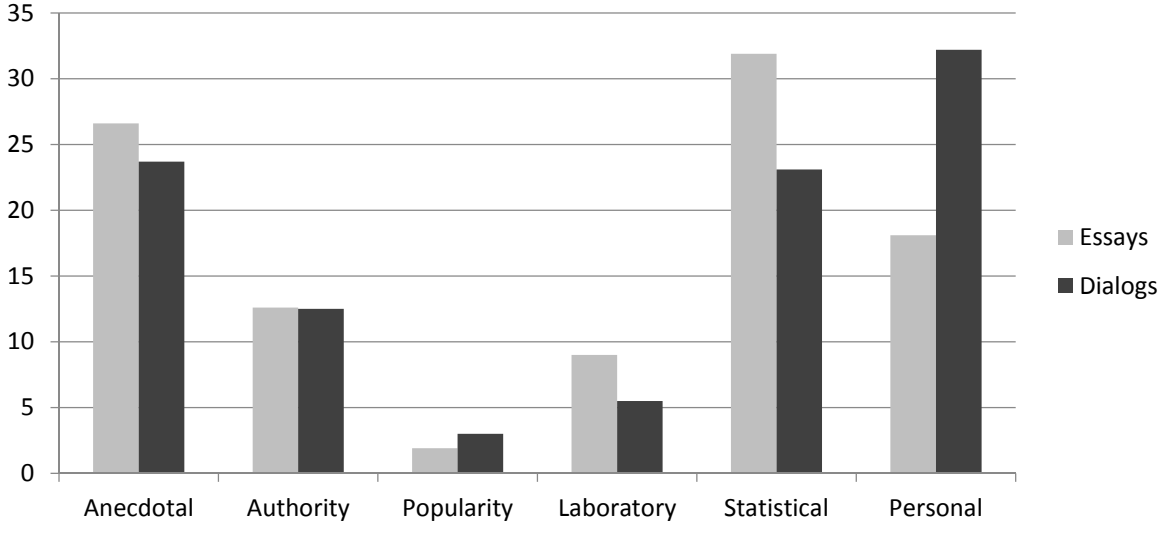


**Figure 2.** Means and standard deviations for the five types of evidence presented on the cards as well as personal evidence in essays and dialogs ("What kind of evidence was used?").

to first-order evidence used to support a claim directly. In addition, 12.2 % of the evidence in the essays and 27.2 % in the dialogs was used to weaken the other's claim, and 1.0 % in the essays and 12.4 % in the dialogs was used to weaken the other's evidence (see also Figure 2).

We then examined whether students' performance in essays and dialogs differed in terms of the purpose of evidence use as represented by the four coding categories. The analysis showed that students gave evidence to support the premises of their arguments to an almost equal extent in both dialogs and essays, $t(71) = 0.41$, $p = .68$, *ns*. Furthermore, there was no specific influence of the class, $t(71) = 0.54$, $p = .54$, *ns,* or the order of intervention, $t(70) = -1.27$, $p = .21$, *ns* (fixed coefficient: –1.19). However, in their essays, students used first-order evidence more often to support their claim directly ($M = 0.84$, $SD = 3.97$) than in the dialogs ($M = -1.74$ $SD = 3.28$), $t(71) = 2.86$, $p = .01$, $d = 0.71$, (fixed coefficient: –2.99). Again, there was no specific influence of either the class, $t(71) = 0.83$, $p = .41$, *ns*, or the order of intervention, $t(70) = -0.81$, $p = .42$, *ns*.

We further analyzed whether there were any differences in the students' use of evidence to attack the opposing site. Students used more first-order evidence to weaken the opposing claims in dialogs ($M = -2.07$, $SD = 2.75$) than in essays ($M = -3.48$, $SD = 2.74$), $t(71) = -2.66$, $p = .01$, $d = 0.51$, (fixed coefficient: –0.13). However, there was no specific influence of either class, $t(71) = -0.55$, $p = .59$, *ns*, or the order of intervention, $t(70) = 1.42$, $p = .16$, *ns*. A further analysis showed that students used second-order evidence more often to attack the evidence supporting the opposing claim in the dialogs ($M = -3.31$, $SD = 2.74$) than in the essays ($M = -4.49$, $SD = 0.64$, fixed coefficient: –1.14), $t(71) = -2.74$, $p = .01$, $d = 0.59$. As in the former analyses, there were again no differences between the classes, $t(71) = -1.48$, $p = .14$, *ns*, or the order of intervention, $t(70) = 1.29$, $p = .20$, *ns*.

**Table 4.** Example of an essay.

| Move | Contribution | Code |
|---|---|---|
| 1 | Each year, an estimated 443,000 people die from smoking or exposure to secondhand smoke and another 8.6 million have to live a serious illness caused by smoking. | Supp_claim |
| 2 | If I was invited to join a discussion about banning cigarette sales in the US, I would tell them that I believe cigarettes should be banned and that I am one hundred percent certain of my point of view. | Other |
| 3 | I personally know a young boy who has a hole in his throat since birth. That hole was due to his mother's cigarette addiction before and while she was pregnant with him. | Supp_ claim |
| 4 | Commercials on cigarettes give information on what cigarettes are made out of. | Supp_ claim |
| 5 | Studies show that the methane found in cigarettes is also found in cat urine and dog poop. | Supp_ claim |
| 6 | Studies also show the nicotine in cigarettes causes fast-acting chemical reactions in your brain that relieve anxiety and nervousness. | Supp_ claim |
| 7 | My grandmother can smoke up to a pack of cigarettes a day. | Supp_ claim |
| 8 | And reading these facts makes me even more nervous of her health in the future. | Other |

Table 4 shows a prototypical essay (D08_1A), coded with the categories of the coding scheme. The student's claim was put forward in Move 2, after supporting it di-

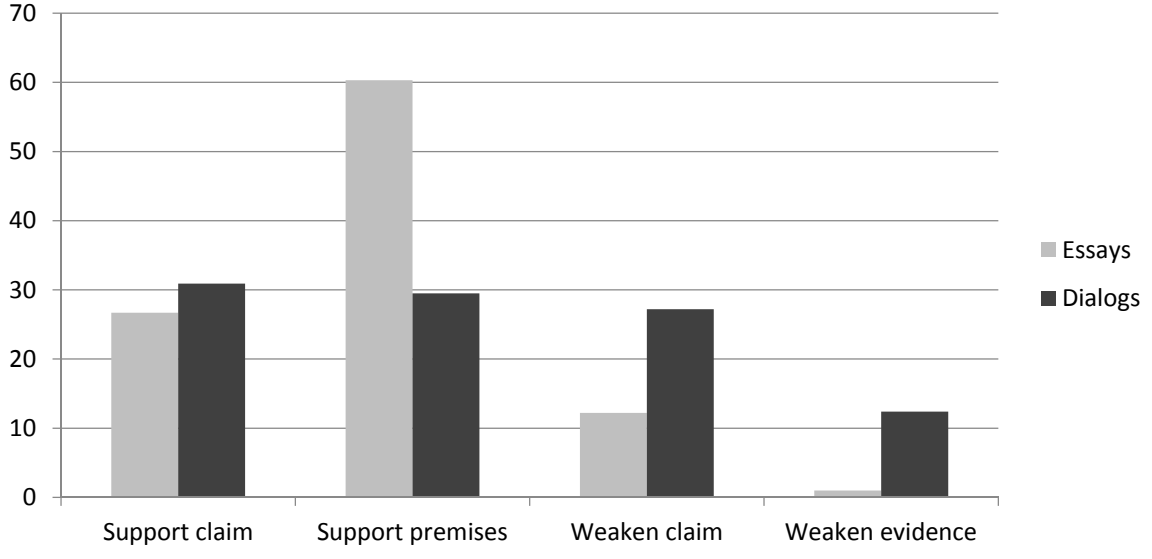


**Figure 3.** Means and standard deviations afor the four categories representing functional and structural aspects of evidence use ("How was evidence used?").

 

rectly in Move 1. Moves 3 to 7 were all intended to support the claim directly ("I believe cigarettes should be banned"), without interpreting it or providing a link between it and the various pieces of evidence. In particular, the student used statistical evidence in Move 1, personal evidence in Moves 3 and 7, scientific evidence in Move 6 and mixed type of evidence in Moves 4 and 5 (scientific evidence known through personal experience or hearsay). In all these moves, the student used evidence as first-order evidence without showing why such information could be relevant to the desired conclusion.

Table 5 reports how the coding scheme captured the different uses of evidence in a dialogue. After the claim was put forth and defended with arguments by A, B used Turn 2 to advance a contrary opinion backed by evidence. Its conclusion (smoking causes death) was then attacked in Turn 3 with an argument supported by second-order evidence. In Turns 4 and 5, the students used first-order evidence to support the claims without any interpretation. However, in Turns 6 to 10, they opened up a meta-dialogic discussion on the relevance of the problem of addiction, using second-order evidence pro and contra the relationship between cancer (and mortality) and smoking. Finally, in Turn 11, Student B advanced another argument leading to the choice of banning cigarette sales, which was then qualified (undermined) by Student A, who pointed out that the dangers resulting from addiction do not affect everyone.

**Table 5.** Example of a dialog.

| Turn | Dialog Partner | Contribution | Code |
|---|---|---|---|
| 1 | A | "Ok, I feel like it shouldn't be like, people should keep selling them because farmers live from tobacco, farming tobacco. So without tobacco they can't, they can't live without tobacco. They don't get money, they don't get paid. They don't have no money to pay their bills if they don't farm tobacco." | Other |
| 2 | B | "But at the same time people are dying because of that like last year, an estimated 443,000 people die from smoking or exposure to secondhand smoking." | Supp_ claim |
| 3 | A | "Yeah but scientists showed that also for people, uhm, with cancer, uhm, for help take away their stress and it gives them the desire to eat" | Weak_ claim |
| 4 | B | "Yeah, but how do you feel about people trying to quit smoking, like each year about 17 million people try to quit smoking, but only about 8 % of them succeed." | Supp_ claim |
| 5 | A | "But people think that everyone has the right to smoke or not." | Supp_ claim |
| 6 | A | "Yeah, but it doesn't matter if you smoke or not because there are people that are like 100 years old and they have been smoking for 80 years and they are still alive, it doesn't matter if you smoke or not. It depends on how you are ..." | Weak_ev |
| 7 | B | "But at the same time remember Georg Harrison, he used to be smoking and died of lung cancer at only 58." | Weak_ev |
| 8 | A | "Yeah but that wasn't only the smoking, he was eating ..." | Other |
| 9 | B | "Not at all." | Other |
| 10 | A | "But he got lung cancer because he didn't know though." | Other |
| 11 | B | "You never know but at the same time the government is still like what it does, it enhances the probability of getting heart attack, it is telling us that's danger and in my opinion it should be banned." | Weak_ claim |
| 12 | A | "Yeah, but it's like, there are people that don't get addicted to cigarettes they only enjoy smoking them. It's not everybody that is smoking cigarettes." | Weak_ claim |

## Discussion

The study reported here was designed to investigate how a dialogic setting influences middle school students' evidence use in argumentation. The crossover design, in which each participant's argumentative behavior in a dialogic and a monologic mode is compared directly within the person, allowed us to control for individual differences as well as effects of order. The experimental setting employed in this study helped to render the differences resulting exclusively from the argumentative setting more salient. The data were analyzed by selecting three fundamental criteria: the argumentative purpose of evidence use (function), the structure thereof (argumentative structure) and the content of the pieces of evidence.

From a functional point of view, we detected a fundamental difference between the two settings regarding the consideration of the opposing position. Students used more evidence to support their own claim than to weaken the opposing one – regardless of whether they were writing an essay (87 %) or discussing with a peer partner (60 %). However, whereas the essays included almost exclusively first-order evidence in support of the own view, the use of evidence in the dialogs was more balanced. Here, students tended to use second-order evidence to weaken the oppos-

ing view, which is in line with Walton's (1989) dual goals of argumentative discourse. These results replicate findings on more practiced students reported by Kuhn and Moore (2015). Apparently, although being taught dialogic argumentation as a 2-year-long intervention (compared with a nondialogic curriculum) clearly improved students' ability to use evidence to support their own claim, the proportions of evidence use addressing the opposing view remained almost the same between dialogs and essays. This shows how much the setting itself, namely the argumentative mode, influences the way evidence is used.

From a structural perspective, we found a crucial difference in the kind of argumentative use of evidence to support or weaken a viewpoint. Students used evidence in different ways: They cited it to support their claim directly (first-order evidence), or they interpreted it to back up their arguments or weaken the grounds of the opposing one (second-order evidence). In the essays, when supporting their view, students used first-order evidence more often, that is, a less elaborated argumentative strategy. In the dialogic setting, when referring to the opposing view in order to weaken it, students used second-order evidence more frequently to weaken the argument put forward by the interlocutor. However, this second finding has to be interpreted with caution, because the essays generally included less evidence weakening the opposing claim (this also holds for first-order evidence used to attack the opposing side's evidence directly). Nevertheless, these findings indicate that in dialogs, arguers engage more deeply and authentically in a critical analysis of the positions and the evidence at their disposal. Students seem to be more critical – by referring more frequently to the opposing view – and clearer – by showing a more elaborated line of reasoning. This finding brings to light another important issue. In the dialogic setting, students used second-order evidence more frequently, but they also challenged it more often. The use of second-order evidence seems to lead the interlocutor to the choice of trying to weaken or rebut it. As a matter of fact, this option is easier and more strategic than the alternative, namely supporting alternative data warranting a contrary position with a different backing (Macagno & Walton, 2014). The simple second-order evidence undercutter leaves the burden of persuasion with the speaker, and the attacker simply has to continue to rebut the backings without any need to provide an alternative line of argument.

From a content point of view, the students drew almost exclusively on evidence from the evidence cards in their essays (on average 82 % of references to evidence were of this type). In the dialogs, in contrast, results were very different. Participants were comparably more likely to draw on evidence from their own personal knowledge or experiences. Here, 68 % of evidence references came from the cards, whereas 32 % came from personal knowledge. In our view, this difference might be attributed to the students' (perceived) expectations as well as their prior experiences associated with the setting. Essay writing is a well-established educational method in the school context, and students are used to being evaluated on the basis of their performance. Therefore, they will tend to opt for the pieces of evidence provided by the teacher. In contrast, peer dialogs represent a much more natural setting. Not surprisingly, students will then tend to ground their line of argumentation much more on evidence taken from their personal experience. Moreover, the presence of the peer partner increases this effect. When evidence was taken from personal knowledge or experience in our study, the further reply of the partner was affected by this argumentative choice.[1] However, this interdependency of the dialog partners raises a crucial issue: The dynamics of every single dialog depend very strongly on the input of the particular interlocutors (such as prior knowledge on the topic or prior experience in argumentation). Thus, the students' behavior was not determined solely by the mode, but also by the specific structure of the dyad they were assigned to. In this study, the sample size of the within design as well as the statistical procedure chosen allowed us to control these potential effects. Nevertheless, future research should address this point by incorporating individual measures and investigating their impact on dialogic dynamics. In this context, it is necessary to consider the influence of other relevant aspects such as emotions and/or motivations.

No difference was found regarding the five qualitative categories from the cards. Students showed a strong tendency to use descriptive statistics as evidence in their arguments (32 % in essays and 23 % in dialogs). Then, in the argumentative setting, they referred preferably and more or less equally (27 % in essays and 24 % in dialogs) to anecdotal evidence (the weakest category according to the experts' ratings), whereas laboratory evidence (actually rated by experts to be the strongest evidence) was used only very rarely (9 % in essays and 6 % in dialogs).

These descriptives point strongly to the need to develop students' argumentative skills, particular those used for distinguishing, evaluating and using different types of evidence in a sophisticated way. The present findings can improve our understanding of how curricula based on the dialogic approach (Kuhn, Hemberger & Khait, 2014) can contribute to

[1] Our results confirm what Kuhn and Moore (2015) observed for students with 2 years of practice in dialogical argumentation. However, whereas the amount of shared evidence (or evidence from a list provided to the students) they reported in essays was almost the same as in our findings, it was much less in dialogs (around 20 %). This might be attributed to their higher experience with arguing in a dialogical setting.

 

the development of argumentation in the classroom. Whereas the dialogic setting itself exerts a positive influence on the functional and structural dimensions of evidence use, this does not apply directly to the understanding of evidence quality. Beyond further practice – which could already be proven to be of high relevance for the development of argument skills (Crowell & Kuhn, 2014) – additional support is required to help students use "strong" evidence in an elaborated way. One approach to help students develop a more complete and precise understanding of different types of evidence could be metacognitive prompting (see Jucks, Schulte-Löbbert & Bromme, 2007; Thiebach, Mayweg-Paus & Jucks, 2016).

Metacognitive prompts are questions or hints that encourage metacognitive activities during the learning process. Prompts are designed to overcome superficial processing (King, 1992). In particular, *reflection prompts* (Davis, 2003) target learners' reflection on their own understanding and learning, and provide them with possibilities regarding how to use different prompts. Prompts can be implemented in different ways. Within the new technologies, they are often used in computer-supported settings (Chen, Wei, Wu & Udden, 2009; Stadtler & Bromme, 2007). When using evidence for an argumentative purpose, students could be provided with reflection prompts asking them to think about how convincing the evidence they intend to use might be (e.g. how easily [or not] it could be rebutted by the opposing side). Additionally, one could make transcripts of former (successful) dialogs available that can be used for reflective activities. Enhancing such meta-level awareness has already been shown to be a very efficient approach for developing argumentative strategy use (see Kuhn et al., 2008).

From a theoretical point of view, this study focused mostly on the function and the structure of the use of evidence by developing and applying a coding scheme in which these two dimensions are combined. We addressed the problem of the content of the pieces of evidence separately by classifying these pieces of evidence into classes ranked by experts and analyzing the possible differences in their use. A basic link that needs to be investigated is the relationship between the functional-structural dimension and the content. A piece of evidence can be used as a premise or as a backing to either support a viewpoint or attack an argument. However, is the specific piece of evidence adequate for its intended function? Sandoval and Millwood (2005) and McNeill (2011) have tackled this issue. The first authors introduced the coding categories of sufficiency and conceptual quality to capture whether sufficient relevant data are cited to justify a claim, and whether they can back it up reasonably. The second author proposed the criteria of sufficiency and appropriateness to analyze the type and quality of the relationship between evidence and claim. However, such coding schemes fail to distinguish the functional-structural from the content dimensions, and the criteria for assessing the appropriateness and the conceptual quality of evidence use are left unspecified. Drawing on advances in argumentation theory, it is possible to develop the functional-structural coding scheme to assess the relevance of a piece of evidence for the intended conclusion (Macagno, 2008; Macagno & Walton, 2014). Evidence can be evaluated as preferentially pro or contra a given conclusion, and its degree of support can be ranked as weak or strong. In this fashion, it would be possible to examine not only whether or not students have taken the conflicting view into account and have interpreted the piece of evidence, but also whether they have understood its meaning and its possible argumentative consequences.

## References


Aristotle. (1995). Topics. In J. Barnes (Ed.), *The works of Aristotle*. Princeton, NJ: Princeton University Press.

Asterhan, C. S. C., & Schwarz, B. B. (2009). Argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialog. *Cognitive Science, 33*, 374–400. doi: 10.1111/j.1551-6709.2009.01017.x

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278. doi: 10.1016/j.jml.2012.11.001

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412. doi: 10.1016/j.jml.2007.12.005

Chen, N.-S., Wei, C.-W., Wu, K.-T., & Udden, L. (2009). Effect of high level prompts and peer assessment on online learners' reflection levels. *Computers & Education, 52*, 283–291. doi: 10.1016/j.compedu.2008.08.07

Clark, D., Sampson, V., Weinberger, A., & Erkens, G. (2007). Analytic framework for assessing dialogic argumentation in online learning environments. *Educational Psychology Review, 19*, 343–374. doi: 10.1007/ s10648–007–9050–7

Crowell, A., & Kuhn, D. (2014). Developing dialogic argumentation skills: A three-year intervention study. *Journal of Cognition and Development*, *15* (2), 1–46. doi: 10.1080/15248372.2012.725187

Davis, E. A. (2003). Prompting middle school science students for productive reflection: Generic and directed prompts. *Journal of the Learning Sciences, 12* (1), 91–142. doi: 10.1207/s15327809jls1201_4

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84*, 287–312. doi: 10.1002/(SICI)1098–237X(200005)84:3<287::AID-SCE1>3.0.CO;2-A

Duschl, R. A., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education, 38* (1), 39–72. doi: 10.1080/03057260208560187

Iordanou, K., & Constantinou, C. P. (2014). Developing pre-service teachers' evidence-based argumentation skills on socio-scientific issues. *Learning and Instruction, 34*, 42–57.

Iordanou, K., & Constantinou, C. P. (2015). Supporting use of evidence in argumentation through practice in argumentation and reflection in the context of SOCRATES learning environment. *Science Education, 99*, 282–311. doi: 10.1002/sce.21152

Jimenez-Aleixandre, M. P., Rodriguez, A. B., & Duschl, R. A. (2000). Doing the lesson or doing science: Argument in high school genetics. *Science Education, 84*, 757–792.

Jucks, R., & Paus, E. (2013). Different words for the same concept: Learning collaboratively from multiple documents. *Cognition and Instruction, 31*(2), 1–28. doi: 10.1080/07370008.2013.769993

Jucks, R., Schulte-Löbbert, P., & Bromme, R. (2007). Supporting experts' written knowledge communication through reflective prompts on the use of specialist concepts. *Journal of Psychology, 215*, 237–247. doi: 10.1027/0044-3409.215.4.237

King, A. (1992). Comparison of self-questioning, summarizing, and notetaking-review as strategies for learning from lectures. *American Educational Research Journal, 29*, 303–323, 303–323. doi: 10.3102/00028312029002303

Koballa, T. R. (1992). Persuasion and attitude change in science education. *Journal of Research in Science Teaching, 29* (1), 63–80. doi: 10.1002/tea.3660290107

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education, 77*, 319–337.

Kuhn, D. (2015). Thinking together and alone. *Educational Researcher, 44*(1), 46–53. doi: 10.3102/0013189X15569530

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. Orlando, FL: Academic Press.

Kuhn, D., & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' thinking. *Psychological Science, 22*, 545–552. doi: 10.1177/0956797611402512

Kuhn, D., Garcia-Mila, M., Zohar, A., Andersen, C., White, S. H., Klahr, D., & Carver, S. M. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development, 60*(4), i. doi: 10.2307/1166059

Kuhn, D., Goh, W., Iordanou, K., & Shaenfield, D. (2008). Arguing on the computer: A microgenetic study of developing argument skills in a computer-supported environment. *Child Development, 79*, 1310–1328. doi: 10.1111/j.1467-8624.2008.01190.x

Kuhn, D., Hemberger, L., & Khait, V. (2014). *Argue with me*. New York, NY: Wessex.

Kuhn, D., & Moore, W. (2015). Argumentation as a core curriculum. *Learning: Research and Practice*, *1* (1), 66–78. doi: 10.1080/23735082.2015.994254.

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development, 74*, 1245–1260. doi: 10.1111/1467-8624.00605

Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review, 62*, 155–179. doi: 10.17763/haer.62.2.9r424r0113t670l1

Macagno, F. & Walton, D. (2014). *Emotive language in argumentation*. New York, NY: Cambridge University Press.

Macagno, F. (2008). Dialectical relevance and dialogical context in Walton's pragmatic theory. *Informal Logic, 28*, 102–128.

Macagno, F. (2012). Reconstructing and assessing the conditions of meaningfulness: An argumentative approach to presupposition. In H. Ribeiro (Ed.), *Inside arguments: Logic and the study of argumentation* (pp. 247–268). Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Macagno, F., Mayweg-Paus, E., & Kuhn, D. (2014). Argumentation theory in education studies: Coding and improving students' argumentative strategies. *Topoi: An International Review of Philosophy.* doi: 10.1007/s11245-014-9271-6.

Mayweg-Paus, E., Macagno, F., & Kuhn, D. (2015). Developing argumentation strategies in electronic dialogs: Is modeling effective? *Discourse Processes*. doi:10.1080/0163853X.2015.1040323

McNeill, K. L. (2011). Elementary students' views of explanation, argumentation, and evidence, and their abilities to construct arguments over the school year. *Journal of Research in Science Teaching, 48*, 793–823.

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The Journal of the Learning Sciences, 15*, 153–191. doi: 10.1207/s15327809jls1502_1

Nussbaum, E. M., & Edwards, O. V. (2011). Critical questions and argument stratagems: A framework for enhancing and analyzing students' reasoning practices. *Journal of the Learning Sciences, 20*, 443–488.

Osborne, J. F., Erduran, S., Simon, S., & Monk, M. (2001). Enhancing the quality of argument in school science. *School Science Review, 82* (301), 63–70.

Pollock, J. L. (1974). *Knowledge and justification*. Princeton, NJ: Princeton University Press.

Rapanta, C., Garcia-Mila, M., & Gilabert, S. (2013). What is meant by argumentative competence? An integrative review of methods of analysis and assessment in education. *Review of Educational Research, 83*, 483–520. doi: 10.3102/0034654313487606

Sandoval, W. A. (2003, July). The inquiry paradox: Why doing science doesn't necessarily change ideas about science. In Proceedings of the sixth international computer-based learning in science conference (pp. 825–834).

Sandoval, W., & Millwood, K. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction, 23* (1), 23–55. doi: 10.1207/s1532690xci2301_2

Stadtler, M., & Bromme, R. (2007). Dealing with multiple documents on the WWW: The role of metacognition in the formation of documents models. *International Journal of Computer Supported Collaborative Learning, 2*, 191–210. doi: 10.1007/s11412-007-9015-3

Thiebach, M., Mayweg-Paus, E., & Jucks, R. (2016). Better to agree or disagree? - The role of critical questioning and elaboration in argumentative discourse. *Zeitschrift für Pädagogische Psychologie, 30*(2–3), 133–149.

Toulmin, S. (1958). *The uses of argument*. Cambridge, UK: Cambridge University Press. doi: 10.1017/S0031819100037220

Walton, D. N. (1989). Dialogue theory for critical thinking. *Argumentation, 3*, 169–184. doi: 10.1007/bf00128147

Walton, D. N. (2006). *Fundamentals of critical argumentation*. New York, NY: Cambridge University Press.

Walton, D. N. & Macagno, F. (2007). The fallaciousness of threats: Character and ad baculum. *Argumentation, 21*(1), 63–81.

Manuscript submitted: 13.3.2015
Manuscript accepted: 2.9.2015
Conflict of interest: No

**Elisabeth Mayweg-Paus**
Westfälische Wilhelms-Universität Münster
Institute of Psychology for Education
Fliednerstraße 21
48149 Münster
Germany
e.mayweg@uni-muenster.de

 