

AL: An Adaptive Learning Support System for Argumentation Skills

Christina Niklaus
Matthias Cetto
Siegfried Handschuh

Thiemo Wambsganss
Matthias Söllner
Jan Marco Leimeister

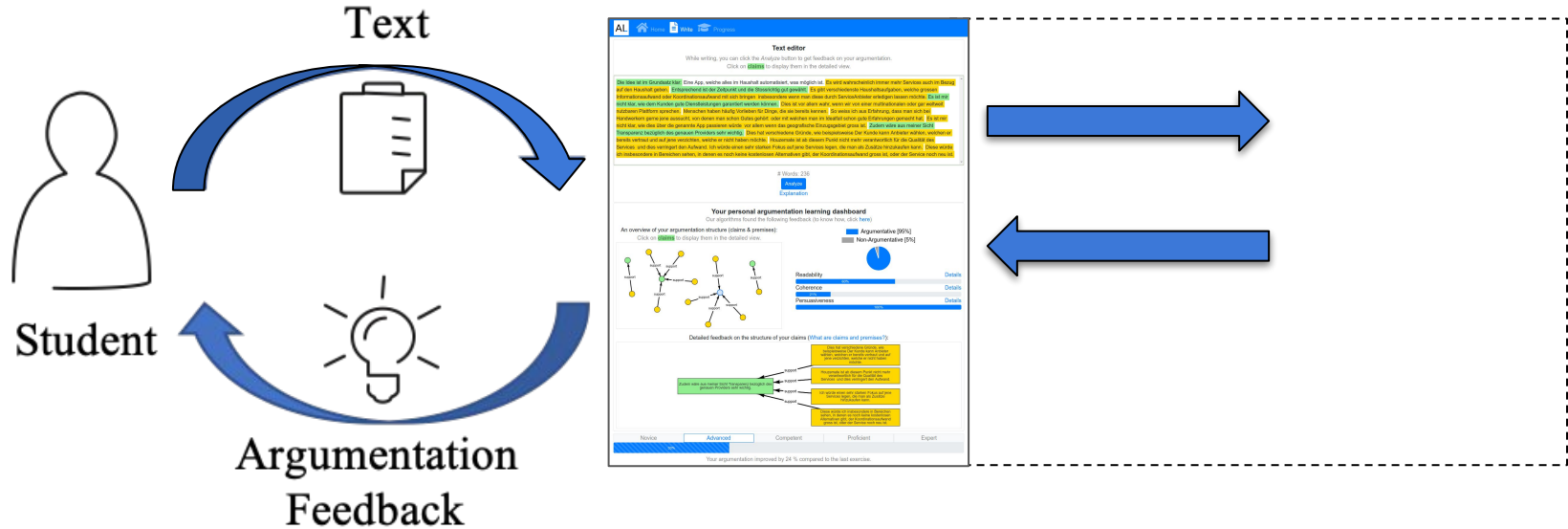
Motivation

- Nowadays, information is readily available.

-> People need to develop skills other than the replication of information.
- More important: **higher-order thinking skills**, e.g. the skill of arguing in a structured, reflective and well-formed way
- To develop such skills, it is of great importance for the individual student to receive continuous feedback throughout their learning journey (“formative feedback”).
- challenge: **How to provide formative feedback in large-scale lectures effectively?**

-> leverage Argument Mining techniques to design and built **AL, an adaptive learning tool that provides students with feedback on the argumentation structure** during the writing process

Workflow

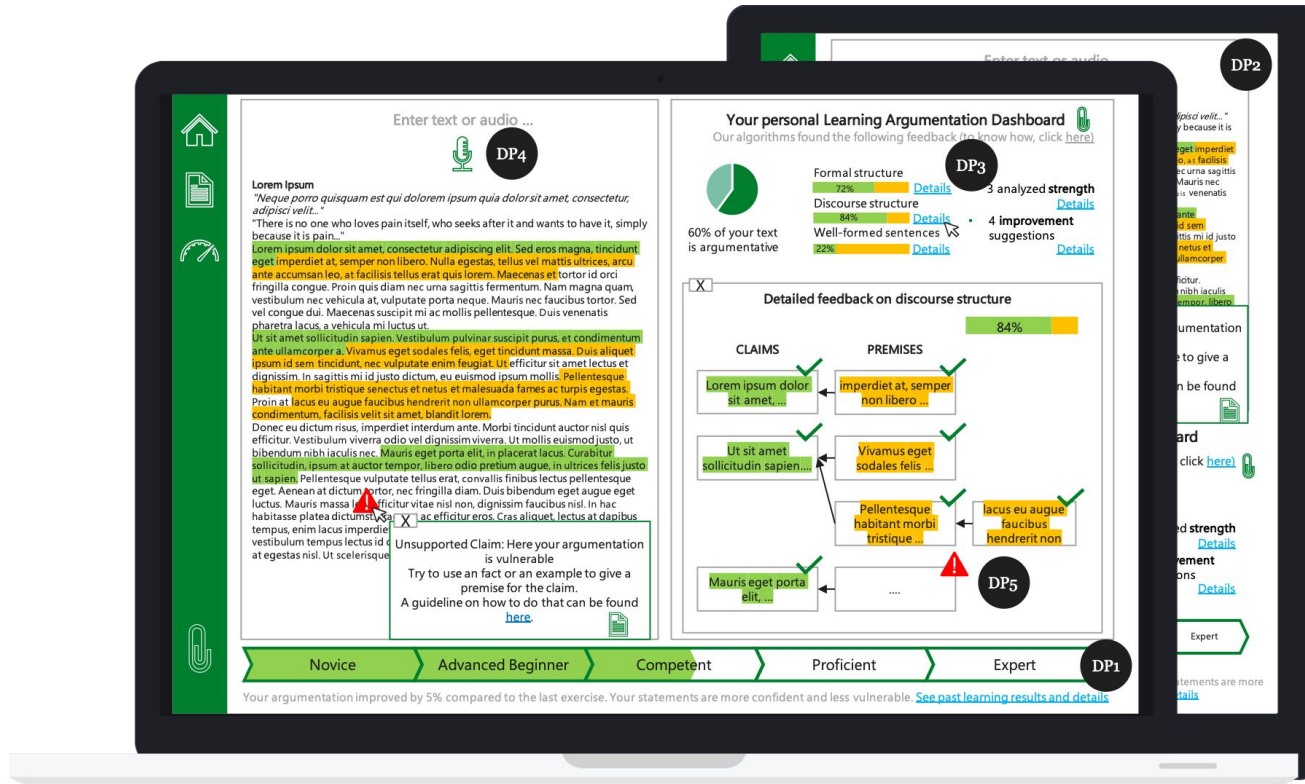


Goal: Improve the users' argumentation skills by providing **immediate, individual feedback** using a **ML-based analysis** of their argumentation structure in written texts.

Development Approach

- top-down:
 - literature analysis in the fields of educational technology and pedagogical theories
 - **30 interviews with students to derive user requirements** for a first design of AL:
 - experience with technology-based learning systems
 - perception of existing learning systems in use
 - importance of skills in university education
 - requirements for a system that supports learning meta cognition skills
 - requirements for a system that supports learning how to argue
- bottom-up:
 - **low-fidelity prototypes to test design hypotheses** with 49 users
 - 3 paper prototypes
 - 2 digital mockups
- result: **7 design principles** on how to build an adaptive argumentation feedback tool

Paper Prototype

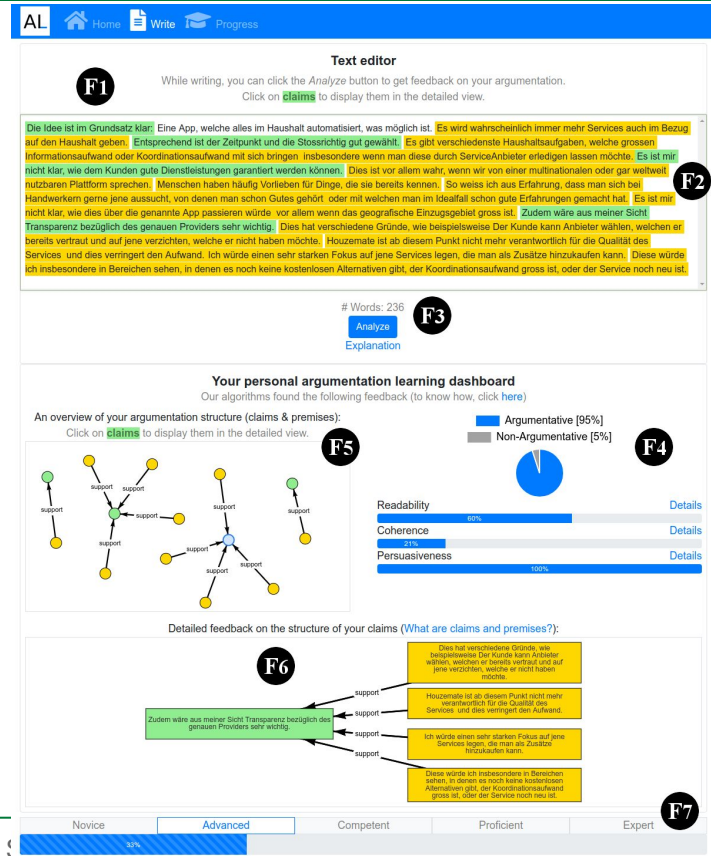


Design Principles

	Design Principle
1)	Provide the learning tool with a learning progress indicator in order for users to actively monitor their past and current learning development to convey a goal and purpose of learning for a long-term learning.
2)	Provide the learning tool as a web-based application with a responsive, lean and intuitive UX in order for users to intuitively and enjoyably use the tool.
3)	Provide the learning tool with a learning dashboard using gamification elements and a choice of different granularity levels in order for users to receive the right amount of needed feedback information.
4)	Provide the learning tool with a function that displays the theory of argumentation before arguing and incorporate it into the feedback in order to have an orientation in learning.
5)	Provide the learning tool with visual argumentation and discourse feedback on written or spoken information in order for users to apply argumentation and receive instant and individual feedback at any time and any place.
6)	Provide the learning tool with argumentation feedback along best practices, examples based on theory and/or <i>how-to-argue</i> guidelines and do not compare argumentation.
7)	Provide the learning tool with adaptive and individual feedback in order for users to receive useful and specific feedback on their given argumentation.

User Interaction of AL

- **F1:** Text editor for writing argumentative texts
- **F2:** Instant feedback on the given input, using highlighted text spans for identified claims and premises
- **F3:** Word Count
- **F4:** Three summarizing scores wrt. Readability, Coherence & Persuasiveness
- **F5:** Visual graph-based representation of the argumentation structure of the input text
- **F6:** Detailed view of the discourse of individual arguments
- **F7:** Learning progress bar that tracks the past and current learning development of the user

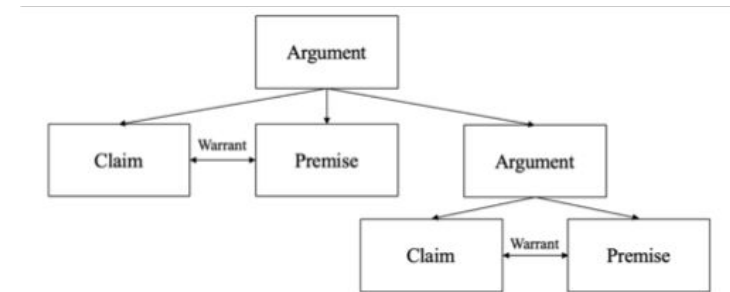


The screenshot displays the AL interface with the following components:

- Top Navigation Bar:** Includes 'AL', 'Home', 'Write', and 'Progress' buttons.
- Text editor (F1):** A text input area with a 'Text editor' label. Below the text, there is a feedback message: 'While writing, you can click the **Analyze** button to get feedback on your argumentation. Click on **claims** to display them in the detailed view.'
- Text Content:** A paragraph of German text about a household automation app. Key phrases are highlighted in yellow, indicating identified claims and premises. For example, 'Die Idee ist im Grundsatz klar' and 'Es wird wahrscheinlich immer mehr Services auch im Bezug auf den Haushalt geben'.
- Word Count (F3):** A section showing '# Words: 236' and buttons for 'Analyze' and 'Explanation'.
- Your personal argumentation learning dashboard (F4):** A section providing an overview of the argumentation structure. It includes a pie chart showing 'Argumentative [95%]' and 'Non-Argumentative [5%]'. Below the chart are three horizontal bar charts for 'Readability' (80%), 'Coherence' (75%), and 'Persuasiveness' (100%). Each chart has a 'Details' link.
- Argumentation Structure Graph (F5):** A visual graph showing the relationships between different parts of the text. Nodes are connected by lines labeled 'support' and 'rebuttal'.
- Detailed feedback on the structure of your claims (F6):** A section titled 'What are claims and premises?'. It shows a list of claims and premises with arrows indicating their relationships. For example, 'Dies hat verschiedene Gründe, wie beispielsweise der Kunde kann Anbieter wählen, welchen er bereits vertraut und auf jene verzichten, welche er nicht haben möchte' is supported by 'Houzzmate ist ab diesem Punkt nicht mehr verantwortlich für die Qualität des Services, und dies verringert den Aufwand'.
- Learning Progress Bar (F7):** A bar at the bottom showing the user's progress from 'Novice' to 'Expert'. The current level is 'Advanced' (50%).

Feedback Algorithm of AL

- build a **corpus of persuasive essays in German**
- train a model for
 - identifying **argument components**
 - identifying **argument relations**



Corpus of Persuasive Essays

- **1,000 student-generated peer reviews** written in **German**
- collected in business innovation lecture, where students develop and present a new business model for which they receive feedback reviews
- a student from the same course **elaborates on the strengths and weaknesses of a business model** and gives persuasive recommendations on what could be improved
- annotated by 3 native German speakers
 - components: *claims* and *premises*
 - relations: *support* and *attack*

Das Geschäftsmodell von Nutreno erscheint mir sehr gut durchdacht. Mir gefällt, dass der Autor sich auf bereits bestehende Technologien beruft, also z.B. dass in der DiabetesForschung bereits ähnliche verwendet werden. ... Dies trifft meiner Meinung nach genau das Ziel einer WellBeingLösung. ... Auch macht das RevenueModell in diesem Geschäftsmodell Sinn durch das Freemium Modell werden die Kunden zunächst neugierig gemacht und danach bei Gefallen dazu verleitet, auf das Premium Modell umzusteigen. Kritisch stehe ich noch den Sensorengegenüber. Kann man die so herstellen, dass sie bezahlbar sind und nicht die Kosten sprengen? Das PremiumModell sollte ja trotzdem bezahlbar sein. Weiter würde ich in der zweiten Version gross darauf hinweisen, dass die Daten sensibel behandelt werden und an niemanden weitergegeben werden. ...

Corpus Analysis

- Statistics:
 - **1000 documents** (feedback from students on business models)
 - 20K sentences
 - 20 sentences per document on average
 - 270 tokens per document on average

	%	Multi- π	Krippendorff's α	Krippendorff's α_U
Claim	0.7053	0.3423	0.3424	0.4379
Premise	0.7048	0.3738	0.3739	0.3812

Argument Component Identification

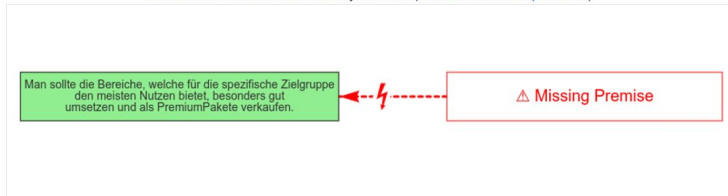
- **sentence-level multi-class classification task** (*claim, premise, non-argumentative*)
- stratified split of the data set into 80% training set and 20% test set
- several classifiers (SVM, Logistic Regression, Random Forest, Multinomial NB, Gaussian NB, Nearest Neighbor)
- best results: **SVM** (accuracy of 65.4%)
- visualization: highlighting components with different colors

Group	Feature	Description
Lexical	Unigrams	Binary and lemmatized unigrams of the component and its preceding tokens
	Dependency tuples	Lemmatized dependency tuples (2k most frequent)
Structural	Token statistics	Number of tokens of component, covering paragraph and covering sentence; number of tokens preceding and following the component in its sentence; ratio of component and sentence tokens
	Component position	Component is first or last in paragraph; component present in introduction or conclusion*; Relative position in paragraph; number of preceding and following components in paragraph
Indicators	Type indicators	Forward, backward, thesis or rebuttal indicators present in the component or its preceding tokens
	First-person indicators	"I", "me", "my", "mine", or "myself" present in component or its preceding tokens
Contextual	Type indicators in context	Forward, backward, thesis, or rebuttal indicators preceding or following the component in its paragraph
	Shared phrases*	Shared noun phrases or verb phrases with the introduction or conclusion (number and binary)
Syntactic	Subclauses	Number of subclauses in the covering sentence
	Depth of parse tree	Depth of the parse tree of the covering sentence
	Tense of main verb	Tense of the main verb of the component
	Modal verbs	Modal verbs present in the component
Probability	POS distribution	POS distribution of the component
	Type probability	Conditional probability of the component being a major claim, claim or premise, given its preceding tokens
Discourse	Discourse Triples	PDTB-discourse relations overlapping with the current component
Embedding	Combined word embeddings	Sum of the word vectors of each word of the component and its preceding tokens

Argument Relation Identification

- **binary classification task of argument component pairs** (support, non-support)
- best results: **SVM** (accuracy of 72.1%)
- visualization:
 - directed graph where a claim is connected with its supporting premises
 - unsupported claims are highlighted

Detailed feedback on the structure of your claims ([What are claims and premises?](#)):



Group	Feature	Description
Lexical	Unigrams	Binary lemmatized unigrams of the source and target components including preceding tokens (500 most frequent)
Syntactic	Part-of-speech Production rules	Binary POS features of source and target components Production rules extracted from the constituent parse tree (500 most frequent)
Structural	Token statistics Component statistics Position features	Number of tokens of source and target Number of components between source and target; number of components in covering paragraph Source and target present in same sentence; target present before source; source and target are first or last component in paragraph; pair present in introduction or conclusion*
Indicator	Indicator source/target Indicators between Indicators context	Indicator type present in source or target Indicator type present between source or target Indicator type follows or precedes source or target in the covering paragraph of the pair
Discourse	Discourse Triples	Binary discourse triples of source and target
PMI	Pointwise mutual information	Ratio of tokens positively or negatively associated with incoming or outgoing relations; Presence of words negatively or positively associated with incoming or outgoing relations
ShNo	Shared nouns	Shared nouns between source and target components (number and binary)

Summary Scores

- **Readability:** How readable is the text based on the Flesch Reading Ease score?
- **Coherence:** How large is the proportion of sentences that are connected via discourse markers?
- **Persuasiveness:** How large is the proportion of claims that are supported by premises as compared to unsupported claims?

Details for "Readability" ×

Your lines are easy to read. Nevertheless, try to formulate your statements even more clearly and to form short, concise sentences in order to improve the legibility and thus the comprehensibility of your text even further.

Close

Evaluation

- **RQ1: Do students perceive AL to be useful and easy to use, and would they continue to use it in the future?**
 - perceived usefulness
 - intention to use
 - ease of use

- **RQ2: How effective is AL with helping users to write more persuasive texts compared to the traditional discussion scripting approach?**
 - comparison of *formal quality of argumentation* and *perceived quality of argumentation* between AL (treatment group) and a discussion scripting approach (control group)

Discussion Scripting Approach

supports the writing process of users
with **input masks**

Reference-Tool

Please use the input masks with claims and premises to write an argumentative text.
If you click on the *Add* button, your argument will be inserted into the text.

Claim (Statement) F1

Der Zeitpunkt und die Stossrichtung sind gut gewählt.

supports

Premise (Evidence, Example) F2

Es wird wahrscheinlich immer mehr Services auch im Bezug auf den Haushalt geben.

Add F3

Text (Claims and Premises)

Die Idee ist im Grundsatz klar: Eine App, welche alles im Haushalt automatisiert, was möglich ist.
Der Zeitpunkt und die Stossrichtung sind gut gewählt. Es wird wahrscheinlich immer mehr Services auch im Bezug auf den Haushalt geben.

Words: 36

Experimental Setup

- *hypothesis*: **individual feedback on student's argumentation will help them to write more convincing texts**
- *study design*: laboratory experiment in which participants were asked to **write peer feedback based on a given essay** ("Does TV make students aggressive?")
- *participants*: **54 students** from our university
 - 24 participants in treatment group (AL)
 - 30 participants in control group (discussion scripting approach)

Study Design and Procedure

- **pre-test phase:**
 - personal innovativeness
 - feedback seeking
 - passive argumentation competency
- **writing phase:**
 - write a review about the argumentation of both parties (pro and con) concerning the weaknesses and strengths of their argumentation
- **post-test phase:**
 - perceived usefulness
 - intention to use
 - ease of use

Argumentation Quality

- **formal quality** of argumentation
 - annotator distinguished between
 - unsupported claims
 - supported claims
 - limited claims
 - supported and limited claims

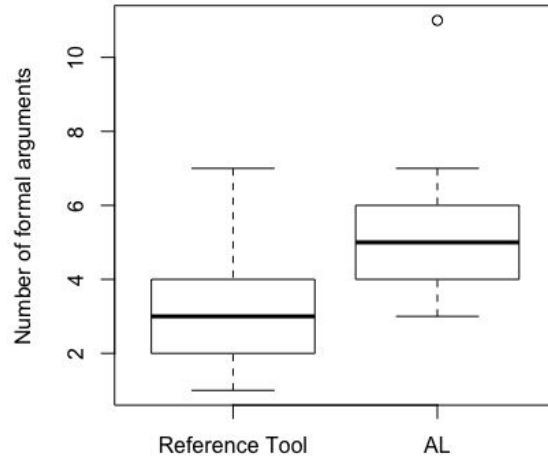
-> number of arguments
- **perceived quality** of argumentation:
 - subjectively judge how persuasive the argumentation is on a Likert scale from 1 to 5 points (2 annotators)

Results: Technology Acceptance

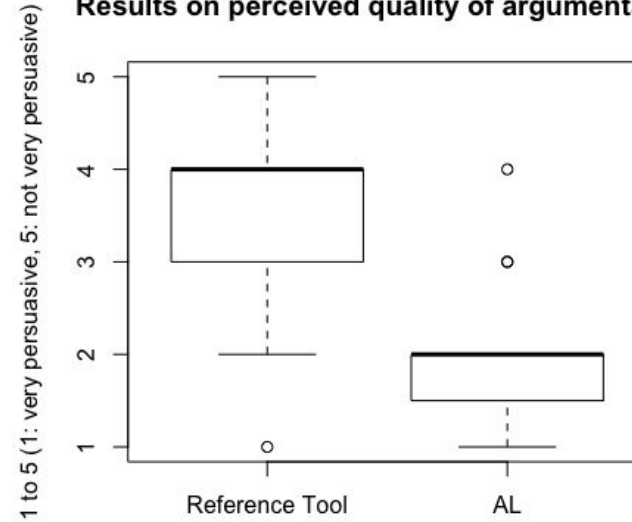
Group	Intention to use	Perceived usefulness	Perceived ease of use
Mean AL	2.33	2.52	2.17
Mean reference tool	3.5	3.28	2.84
SD AL	0.59	0.58	0.65
SD reference tool	1.14	1.12	1.08
p-value	<0.001	0.006	0.012

Results: Argumentation Quality

Results on formal quality of argumentation



Results on perceived quality of argumentation



Qualitative User Feedback

- + fast and direct feedback
- + graph-like visualization
- + summary scores
- accuracy

Group	Feature
On user interaction	<i>"Easy handling. Especially marked texts with colour and percentage values. Fast reaction of the tool. Motivated to write."</i>
On writing support	<i>"I was convinced by the mind map argumentation graphics. If you write longer texts, you can get lost quickly. With this tool you can see how the sentences and argumentation stand together."</i>
On visualization	<i>"It was positive that the tool presented which premises support which of my claims and which arguments hung, so to speak, freely in the air."</i>
On graphics and colour	<i>"I liked that the tool used colors to highlight the various elements and graphics to represent my text. Furthermore, the percentages of how my text was written were helpful and it is exciting to see how the tool judges my text."</i>
On speed of the tool	<i>"Very fast and instant feedback."</i>
Improvements on feedback accuracy	<i>"I'm not sure how well this algorithm really understands what I'm writing."</i>
Improvements on user on-boarding	<i>"Better introduction would be good, you had to try something before you knew how to do it. Maybe a short YouTube tutorial at the beginning, then you know more about what is important."</i>
Improvements on the summarizing scores	<i>"There are only three evaluation points (readability, coherence and persuasiveness) and it is unclear how relevant the individual factors are."</i>

Discussion and Future Work

- improve accuracy of feedback algorithm
- make summary scores more transparent and understandable for the users
- **longitudinal study**: field experiment to evaluate the long-term impact of individual and adaptive feedback

Conclusion

- Participants who used AL wrote **formally more argumentative texts**.
 - **Perceived persuasiveness** significantly **higher**.
 - **High perceived usefulness and intention to use**.
-
- -> AL helps students to **write more structured texts** and motivates them to write **more persuasive texts** in peer learning settings.

AL: An Adaptive Learning Support System for Argumentation Skills

Thiemo Wambsganss
 University of St.Gallen
 St.Gallen, Switzerland
 thiemo.wambsganss@unisg.ch

Christina Niklaus
 University of St.Gallen
 St.Gallen, Switzerland
 christina.niklaus@unisg.ch

Matthias Cetto
 University of St.Gallen
 St.Gallen, Switzerland
 matthias.cetto@unisg.ch

Matthias Söllner
 University of Kassel
 Kassel, Germany
 University of St.Gallen
 St.Gallen, Switzerland
 soellner@uni-kassel.de

Siegfried Handschuh
 University of St.Gallen
 St.Gallen, Switzerland
 University of Passau
 Passau, Germany
 siegfried.handschuh@unisg.ch

Jan Marco Leimeister
 University of St.Gallen
 St.Gallen, Switzerland
 University of Kassel
 Kassel, Germany
 janmarco.leimeister@unisg.ch

ABSTRACT

Recent advances in Natural Language Processing (NLP) bear the opportunity to analyze the argumentation quality of texts. This can be leveraged to provide students with individual and adaptive feedback in their personal learning journey. To test if individual feedback on students' argumentation will help them to write more convincing texts, we developed AL, an adaptive IT tool that provides students with feedback on the argumentation structure of a given text. We compared AL with 54 students to a proven argumentation support tool. We found students using AL wrote more convincing texts with better formal quality of argumentation compared to the ones using the traditional approach. The measured technology acceptance provided promising results to use this tool as a feedback application in different learning settings. The results suggest that learning applications based on NLP may have a beneficial use for developing better writing and reasoning for students in traditional learning settings.

