Short Article

Facilitating justification, disconfirmation, and transparency in diagnostic argumentation

Effects of automatic adaptive feedback in teacher education

Elisabeth Bauer[1] [iD], Michael Sailer[1] [iD], Jan Kiesewetter[2] [iD], Martin R. Fischer[2], Iryna Gurevych[3], and Frank Fischer[1] [iD]

[1]Ludwig-Maximilians-Universität München (LMU), Munich, Germany
[2]University Hospital, Ludwig-Maximilians-Universität München (LMU), Munich, Germany
[3]Technical University Darmstadt, Darmstadt, Germany

Abstract: Teachers need to learn complex skills in higher education, such as diagnostic argumentation. We suggest that relations between the argumentation facets justification, disconfirmation, and transparency are a relevant indicator for the quality of diagnostic argumentation. In an experimental study, we investigated whether automatic adaptive feedback – based on natural language processing – compared to static feedback facilitates relations between the argumentation facets in preservice teachers' diagnostic argumentation when learning with case-based simulations. A sample of N = 60 preservice teachers received adaptive or static feedback on their written explanations concerning simulated cases of pupils having behavioral or reading and writing problems. Using Epistemic Network Analysis, we analyzed learners' written explanations and found that adaptive feedback compared to static feedback facilitates relations between justification, disconfirmation, and transparency in preservice teachers' diagnostic argumentation. The results confirm that adaptivity is an important feature of effective feedback, which can be automated by methods of natural language processing.

Keywords: simulation, adaptive feedback, teacher education, diagnosing, natural language processing

Förderung von Begründung, Widerlegung und Transparenz im diagnostischen Argumentieren: Effekte automatisch adaptiven Feedbacks in der Lehrkräftebildung

Zusammenfassung: Lehrkräfte sind bereits während ihrer universitären Ausbildung mit dem Erlernen komplexer Fähigkeiten konfrontiert, wie etwa dem diagnostischen Argumentieren. Ein relevanter Indikator für die Qualität diagnostischer Argumentation ist das Zusammenspiel zwischen den Argumentationsfacetten Begründung, Widerlegung und Transparenz. In einer experimentellen Studie untersuchten wir die Effekte adaptiven Feedbacks, das mittels Natural Language Processing automatisiert wurde, im Vergleich zu statischem Feedback auf das Zusammenspiel zwischen den Argumentationsfacetten in der diagnostischen Argumentation von Lehramtsstudierenden beim Lernen mit fallbasierten Simulationen. Eine Stichprobe von N = 60 Lehramtsstudierenden erhielt entweder adaptives oder statisches Feedback auf ihre schriftlichen Erklärungen zu simulierten Fällen von Schülerinnen und Schülern mit Verhaltens- oder Lese- und Rechtschreibproblemen. Eine Epistemic Network Analysis der schriftlichen Erklärungen der Lernenden legt nahe, dass adaptives Feedback im Vergleich zu statischem Feedback das Zusammenspiel zwischen Begründung, Widerlegung und Transparenz in der diagnostischen Argumentation von Lehramtsstudierenden fördert. Die Ergebnisse unterstreichen die Relevanz von Adaptivität für effektives Feedback und dessen Automatisierung mittels Natural Language Processing.

Schlüsselwörter: Simulation, adaptives Feedback, Lehrerbildung, Diagnostizieren, Natural Language Processing

Introduction

Simulation-based learning offers future teachers opportunities to practice diagnostic skills, such as communicating with school psychologists about pupils, who might have significant learning difficulties (e.g., dyslexia). We characterize such communicative aspects of diagnostic skills as diagnostic argumentation, which we conceptualize through relations between the facets justification, disconfirmation, and transparency. In this study, we investigate how automatic adaptive feedback compared to static feedback facilitates relations between justification, disconfir-

mation, and transparency in preservice teachers' written diagnostic argumentation when learning with simulations.

Teachers' diagnostic skills

Teachers' diagnostic skills include assessing pupils' performance, progress, and learning prerequisites (e.g., Südkamp et al., 2018). Teachers also play an important role in initially identifying pupils who have clinically significant learning difficulties or behavioral problems, such as dyslexia or an attention deficit hyperactivity disorder (ADHD). In many educational systems, actual clinical diagnoses are made by clinical professionals, such as school psychologists, with whom teachers need to collaborate (Albritton et al., 2021). In such situations, teachers need to explain their diagnostic reasoning to achieve a joint understanding with a collaborating professional, which we refer to as *diagnostic argumentation.*

We characterize diagnostic argumentation by three complementary argumentation facets that build on two basic dimensions – an epistemic dimension (i.e., epistemic activities; e.g., evaluating evidence) and a content dimension (e.g., case-specific evidence, such as hyperactive behavior; Bauer et al., 2019; Bauer et al., 2022): *Justification* describes evaluating evidence (e.g., inattention, hyperactivity, etc.) as a basis for drawing diagnostic conclusions (see Toulmin, 1958). *Disconfirmation* of alternative explanations denotes explicating and discussing differential diagnoses (e.g., ADHD, emotional stress, etc.; see Lawson, 2003; Toulmin, 1958). *Transparency* concerning the applied methods constitutes of describing the sources of evidence and the processes of evidence generation (e.g., tests, observations, conversations, etc.; see Chinn & Rinehart, 2016).

Prior research suggests that preservice teachers seem to focus on justification and tend to omit disconfirmation and transparency in their diagnostic argumentation (Bauer et al., 2022). However, the information associated with the three facets is complementary: Justification in relation with disconfirmation indicates discussing two or more competing diagnoses in light of the evidence; justification in relation with transparency indicates explicating and potentially evaluating methods and sources used to generate specific evidence; disconfirmation in relation with transparency indicates generating evidence for drawing conclusions concerning one or more specific diagnoses (Bauer et al., 2019). Relating the different facets' complementary information strengthens the conclusiveness of diagnostic argumentation. Therefore, we consider relations between the complementary information of justification, disconfirmation, and transparency as a relevant indicator for the quality of diagnostic argumentation. Facilitating these as-

pects of diagnostic skills seems relevant for preparing future teachers for professional situations that require diagnostic argumentation.

Simulation-based learning and feedback

To facilitate diagnostic argumentation in higher education, meta-analytical evidence suggests using case-based *simulations* (Chernikova et al., 2020), which are approximations of practice that include simplified, yet valid representations of professional situations (Grossman et al., 2009). Simulation-based learning might benefit from elaborated feedback, which provides information on the appropriate task processing (Wisniewski et al., 2020).

In online learning environments (e.g., digital case-based simulations), elaborated feedback is often implemented as *static feedback*, for example by providing an expert solution that exemplifies optimal task processing. However, static feedback requires learners to compare their own solution themselves, which demands high levels of learners' cognitive capacities (Sweller et al., 2019) – especially with regard to demanding tasks, such as diagnostic argumentation. In contrast, *adaptive feedback* adjusts to a learner's task solution by directly addressing gaps compared to the expert solution (Bimba et al., 2017), such as missing argumentation facets in learners' diagnostic argumentation. By making relevant feedback information more visible and accessible to the learners (Machts et al., 2024), adaptive feedback increases the *salience* of the feedback information, which might facilitate learners' cognitive processing of the feedback (Sweller et al., 2019).

While providing an expert solution as static feedback is easily automated, automating adaptive feedback on learners' argumentative task solutions is challenging. Yet, recent technological advancements in the field of artificial intelligence-based natural language processing (NLP) – namely artificial neural networks that process the context within language instead of only recognizing words – facilitate automating the analysis of written arguments and thus, adaptive feedback on argumentative task solutions (e.g., Wambsganss et al., 2021). However, the effects of *NLP-based automatic adaptive feedback* compared to *static feedback* on the relations between justification, disconfirmation, and transparency in preservice teachers' diagnostic argumentation are yet to be investigated.

The present study

In this study, we reanalyzed data from a prior study (Sailer et al., 2023), which investigated effects of NLP-based automatic adaptive feedback compared to static feedback in

individual compared to collaborative learning settings on preservice teachers' diagnostic accuracy and the quality of justifications when learning with simulations. This prior study did not investigate relations between justification, disconfirmation, and transparency in preservice teachers' diagnostic argumentation. To analyze relations between argumentation facets, the present reanalysis employed the method of Epistemic Network Analysis (ENA; Shaffer, 2017). ENA analyzes relations between categories in data (e.g., Omarchevska et al., 2021), such as relations between argumentation facets within written explanations, as well as the relative emphasis on these categories.

In our reanalysis, we investigated the effects of adaptive compared to static feedback on the relations between justification, disconfirmation, and transparency and thus, on the quality of preservice teachers' diagnostic argumentation. We hypothesized that adaptive feedback compared to static feedback facilitates relations between justification, disconfirmation, and transparency, because adaptive feedback might increase the salience of feedback information addressing the argumentation facets and their relations.

Method

Participants and research design

For our reanalysis, we used a subsample of our prior study (Sailer et al., 2023) and excluded the learners of a collaborative learning condition to focus on the effects of adaptive compared to static feedback on *individual* learners' diagnostic argumentation. We reanalyzed data of $N = 60$ German preservice teachers (50 female, 10 male; age: $M = 22.47$, $SD = 3.45$ years; semester: $M = 4.62$, $SD = 3.13$, $Min = 1$, $Max = 14$), using a randomized experimental design with the between-subjects factor *type of feedback* and the two experimental groups *adaptive feedback* ($n = 30$) and *static feedback* ($n = 30$).

Learning environment, materials and tasks

Learners were asked to take on the role of a teacher and process six simulated pupil cases in a learning phase and two simulated pupil cases as post-test. All cases concerned pupils with various learning difficulties or behavioral problems that might indicate a clinical diagnosis in the range of ADHD or dyslexia. The cases were implemented on the case-based online platform CASUS (http://www.casus. net). In each case, learners could access different sources of evidence, such as conversation transcripts (see ESM 1, Supplement A; for all learning materials see: https://osf.io/

hn7wm/). To complete a case, preservice teachers wrote an explanation concerning their diagnostic reasoning.

Static and adaptive feedback

In the learning phase, learners in the *static feedback* condition (SFC) received case-specific expert solutions, which exemplified the epistemic and the content dimension of how experts would relate the complementary information of justification, disconfirmation, and transparency in their diagnostic argumentation (see ESM 1, Supplement B: Supplementary Figure 3).

In the *adaptive feedback* condition (AFC), learners' explanations were analyzed by an NLP-algorithm, which was trained using the Python-based web service NeuralWeb. The training data (i.e., written explanations on the same simulated cases of 118 preservice teachers) was manually coded regarding diagnostic entities (i.e., content dimension; e.g., hyperactivity) and epistemic activities (i.e., epistemic dimension; e.g., evaluating evidence). Thus, the algorithm could identify diagnostic entities and epistemic activities as correct, incorrect, or missing in new explanations written by learners in the present study (for details about the algorithm and the feedback system see Pfeiffer et al., 2019 and Schulz et al., 2019). Based on the automatic analysis, a suitable subset of around 40 case-specific feedback paragraphs were adaptively shown to the learner. Parts of the feedback addressed the epistemic activities and their relations (i.e., epistemic dimension) and other parts the diagnostic entities and their relations (i.e., content dimension; see ESM 1, Supplement B: Supplementary Figure 4). The adaptive feedback also offered highlighting diagnostic entities and activities found in a learner's submitted explanation.

Procedure

Participants spent $M = 156.50$ ($SD = 33.65$) minutes on the laboratory study. They watched a short video about navigating in CASUS and an 18-minute video input about pupils' learning difficulties and behavioral problems. Next, participants entered the learning phase with six simulated cases and received static or adaptive feedback. Participants took a short break after three cases. Finally, participants processed two post-test cases without receiving feedback.

Data sources and measurements

As data source, we used the written explanations from the two post-test cases (see ESM 1, Supplement A). We manu-

ally coded the explanations of all cases using a case-specific coding scheme for (a) justification, (b) disconfirmation, and (c) transparency. Two trained raters double coded 18 % of the data and achieved substantial agreement: For (a) *justification*, we coded the presence or absence of the six primary supporting pieces of evidence for the correct diagnosis (Cohen's κ = .90). For (b) *disconfirmation*, we coded the presence or absence of the six most relevant differential diagnoses (Cohen's κ = .94). For (c) *transparency*, we coded the presence or absence of informational sources of the six primary evidences for the correct diagnosis (Cohen's κ = .92).

Statistical analysis

To investigate the relations between justification, disconfirmation, and transparency in preservice teachers' diagnostic argumentation, we used an ENA (Shaffer, 2017). The ENA algorithm operationalized the argumentation facets' relations by accumulating and weighting co-occurrences of the argumentation facets, first within learners' written explanations, then per simulated case, and then per experimental group (SFC vs. AFC). The resulting mathematical model is depicted in two-dimensional network graphs. The graphs show the strength of relations (i.e., relative frequencies of co-occurrences) between justification, disconfirmation, and transparency per experimental group as a network. In addition, group means provide information about the relative focus on the facets and on their relations in the argumentation. By aligning two group means on one axis in the network space, systematic variance is shifted to one dimension, which enables statistical testing of group differences. To test group differences between SFC and AFC, we used an independent-samples *t*-test (α = .05).

Results

Using ENA, we investigated whether adaptive compared to static feedback facilitates relations between justification, disconfirmation, and transparency in preservice teachers' diagnostic argumentation in the two post-test cases. A randomization check indicated no significant a priori performance differences (see ESM 1, Supplement C) or differences in time on task in the learning phase (see ESM 1, Supplement D) between the two feedback conditions. Supplement E reports descriptive and inferential statistics of the individual argumentation facets.

Figure 1 presents the diagnostic argumentation networks of the SFC (Figure 1a) and of the AFC (Figure 1c). The thickness of the networks' colored edges reflects the relative strength of relations between two argumentation facets respectively. In both conditions, the relation between justification and disconfirmation (i.e., discussing differential diagnoses in light of relevant evidence) had the highest relative frequency. The relation between justification and transparency (i.e., explicating methods and sources used to generate relevant evidence) had the second highest relative frequency in both conditions. The comparison graph (Figure 1b) subtracts the two networks to highlight their differences and shows a group mean for each feedback condition (colored squares; dashed boxes are confidence intervals).

The group mean of the SFC (Figure 1b, red square) was located toward the edge representing the relation between justification and disconfirmation. However, along the edge connecting justification and disconfirmation as well as in comparison to the AFC group mean, the SFC group mean was located more toward justification. We interpret this finding such that overall, learners in the SFC showed a relatively strong focus on justification.

By comparison, the group mean of the AFC (Figure 1b, blue square) was located higher along the Y-axis, more to-
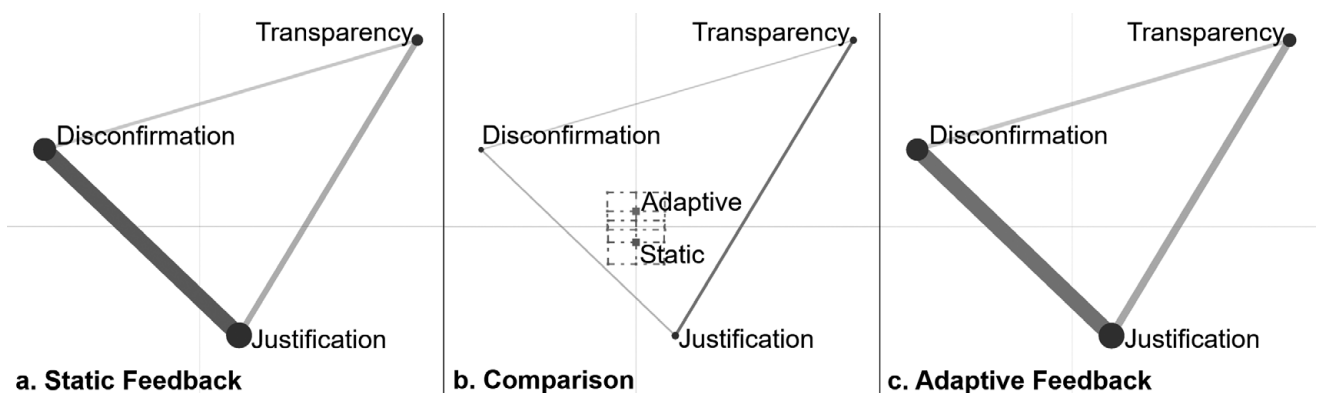

Figure 1. Diagnostic Argumentation Networks of the SFC (1a) and the AFC (1c) in the Post-test; The Comparison Plot (1b) Shows Differences Between the Two Networks, Group Means (Colored Squares), and Confidence Intervals (Dashed Boxes).

ward disconfirmation and transparency. This more central position suggests that learners in the AFC included relatively more relations between justification, disconfirmation, and transparency compared to learners in the SFC. Learners in the AFC especially focused more on the relation between justification and transparency (i.e., a stronger engagement in explicating methods and sources used to generate relevant evidence). However, as indicated by the comparison graph and the position of the group mean, learners receiving adaptive feedback also had a slightly stronger focus on the relation between justification and disconfirmation as well as the relation between disconfirmation and transparency.

The results of the t-test confirmed that the group means of the SFC ($M = -.12$, $SD = .65$) and the AFC ($M = .12$, $SD = .55$) were significantly different, $t(114.71) = -2.16$, $p = .03$, Cohen's $d = .40$. The findings support the hypothesis that adaptive compared to static feedback facilitates relations between justification, disconfirmation, and transparency in preservice teachers' diagnostic argumentation.

Discussion

In this study, we investigated whether NLP-based automatic adaptive feedback compared to static feedback facilitates relations between justification, disconfirmation, and transparency in preservice teachers' diagnostic argumentation when learning with simulations. We consider relations between the argumentation facets as a relevant indicator for the quality of diagnostic argumentation: Providing and relating the complementary information associated with the argumentation facets might facilitate achieving a joint understanding with collaborating professionals (Bauer et al., 2019; Bauer et al., 2022).

Using the method of ENA (Shaffer, 2017), we found support for the hypothesis that adaptive feedback compared to static feedback may foster relations between justification, disconfirmation, and transparency. Preservice teachers receiving static feedback had a relatively strong focus on justification. This stronger focus on justification has to be considered relative to the other argumentation facets, thus, indicating fewer relations between argumentation facets in general and the relation between justification and transparency in particular. In a prior study, we found that preservice teachers who received no feedback when learning with simulations focused on justification in their diagnostic argumentation as well (Bauer et al., 2022). The results of the present study therefore suggest that static feedback advanced preservice teachers' quality of diagnostic argumentation less than adaptive feedback. Adaptive feedback was superior in fostering preservice teachers' ar-

gumentation quality in terms of relating the complementary information associated with justification, disconfirmation, and transparency.

Adaptive feedback might have facilitated learners' cognitive processing of the feedback by increasing the salience of feedback information addressing the argumentation facets and their relations (Machts et al., 2024; Sweller et al., 2019). The adaptive feedback addressed whether the facets of diagnostic argumentation were identified as present or missing in learners' submitted explanations, whereas the static feedback exemplified the argumentation facets without directly addressing the learners' explanations. The adaptive feedback also offered to highlight the corresponding parts in learners' explanations (see ESM 1, Supplement B: Supplementary Figure 4), thus making connections between the feedback information and the learners' explanations more visible and accessible to the learners. The results emphasize that elaborated feedback is most effective when adaptively addressing learners' task processing and helping them to understand how to improve their performance (Bimba et al., 2017; Wisniewski et al., 2020). Using an artificial neural network algorithm for NLP proved to be feasible and effective for providing adaptive feedback on written explanations in real-time. However, training data and expertise in NLP are resource intense necessities that need to be considered. Yet, the NLP-based approach might be efficient for employing adaptive feedback on written task solutions at a large scale (e.g., for high numbers of students in teacher education programs).

Investigating the quality of diagnostic argumentation by the relations between justification, disconfirmation, and transparency is a relatively new approach, which requires further validation (e.g., comparing novices and experts). Our reanalysis was underpowered for detecting small effects, which suggests replication research. The employed NLP algorithms are specialized on the simulated cases used in this study, which limits generalizability of both the applicability of the algorithms and our study's results. Future research might investigate how to transfer the algorithms to other cases. Further research might also investigate the effects of adaptivity with regard to different feedback characteristics (e.g., salience through highlighting).

Conclusion

Artificial intelligence-based NLP proved to be an effective approach for automating adaptive feedback. Adaptive feedback compared to static feedback fosters the quality of preservice teachers' diagnostic argumentation, indicated by the relations between justification, disconfirmation, and transparency, when learning with simulated cases. Facilitating diagnostic argumentation in teacher education

programs by using case-based simulations with adaptive feedback might contribute to the further professionalization of future teachers.

Electronic supplementary material

The electronic supplementary material (ESM) is available with the online version of the article at https://doi.org/10.1024/1010-0652/a000363

ESM 1. Provides further information about the simulation and the post-test cases, the static feedback and automatic adaptive feedback, and the results of additional analyses (PDF).

References

Albritton, K., Chen, C.-I., Bauer, S. G., Johnson, A., & Mathews, R. E. (2021). Collaborating With School Psychologists: Moving Beyond Traditional Assessment Practices. *Young Exceptional Children, 24*(1), 28–38.

Bauer, E., Sailer, M., Kiesewetter, J., Fischer, M. R., & Fischer, F. (2022). Diagnostic Argumentation in Teacher Education: Making the Case for Justification, Disconfirmation, and Transparency. *Frontiers in Education, 7*, Article 977631.

Bauer, E., Sailer, M., Kiesewetter, J., Schulz, C., Pfeiffer, J., Gurevych, I., Fischer, M. R., & Fischer, F. (2019). Using ENA to Analyze Pre-service Teachers' Diagnostic Argumentations: A Conceptual Framework and Initial Applications. In B. Eagan, M. Misfeldt, & A. Siebert-Evenstone (Eds.), *Advances in Quantitative Ethnography* (pp. 14–25). Springer.

Bimba, A. T., Idris, N., Al-Hunaiyyan, A., Mahmud, R. B., & Shuib, L. (2017). Adaptive feedback in computer-based learning environments: a review. *Adaptive Behavior, 25*(5), 217–234.

Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-Based Learning in Higher Education: A Meta-Analysis. *Review of Educational Research, 90*(4), 499–541.

Chinn, C. A., & Rinehart, R. W. (2016). Commentary: Advances in research on sourcing – source credibility and reliable processes for producing knowledge claims. *Reading and Writing, 29*(8), 1701–1717.

Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching, 15*(2), 273–289.

Lawson, A. (2003). The nature and development of hypothetico-predictive argumentation with implications for science teaching. *International Journal of Science Education, 25*(11), 1387–1408.

Machts, N., Chernikova, O., Jansen, T., Weidenbusch, M., Fischer, F.& Möller, J. (2024). *Categorization of simulated diagnostic situations and the salience of diagnostic information – Conceptual framework. Zeitschrift für Pädagogische Psychologie, 38*(1–2), 3–13. https://doi.org/10.1024/1010-0652/a000364

Omarchevska, Y., Lachner, A., Richter, J., & Scheiter, K. (2021). It takes two to tango: How scientific reasoning and self-regulation processes impact argumentation quality. *Journal of the Learning Sciences, 32*(3), 1–41.

Pfeiffer, J., Meyer, C. M., Schulz, C., Kiesewetter, J., Zottmann, J., Sailer, M., Bauer, E., Fischer, F., Fischer, M. R., & Gurevych, I.

(2019). Famulus: Interactive annotation and feedback generation for teaching diagnostic reasoning. In S. Padó (Ed.), *Proceedings of the EMNLP-IJCNLP Conference: Nov 3–7, 2019, Hong Kong* (pp. 73–78).

Sailer, M., Bauer, E., Hofmann, R., Kiesewetter, J., Glas, J., Gurevych, I., & Fischer, F. (2023). Adaptive feedback from artificial neural networks facilitates pre-service teachers' diagnostic reasoning in simulation-based learning. *Learning and Instruction, 83*, Article 101620.

Schulz, C., Meyer, C. M., & Gurevych, I. (2019). Challenges in the Automatic Analysis of Students' Diagnostic Reasoning. In A. Korhonen, D. Traum, & L. Màrquez (Eds.), *Proceedings of the 57th AAAI Conference: July 28-Aug 2, 2019, Florence* (pp. 6974–6981).

Shaffer, D. W. (2017). *Quantitative ethnography*. Cathcart Press.

Südkamp, A., Praetorius, A.-K., & Spinath, B. (2018). Teachers' judgment accuracy concerning consistent and inconsistent student profiles. *Teaching and Teacher Education, 76*, 204–213.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review, 31*(2), 261–292.

Toulmin, S. E. (1958). *The uses of argument*. Cambridge university press.

Wambsganss, T., Kueng, T., Soellner, M., & Leimeister, J. M. (2021). ArgueTutor: An Adaptive Dialog-Based Learning System for Argumentation Skills. In *Proceedings of the CHI Conference* (pp. 1–13). ACM.

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research. *Frontiers in Psychology, 10*, Article 3087.

ORCID
Elisabeth Bauer
 https://orcid.org/0000-0003-4078-0999
Michael Sailer
 https://orcid.org/0000-0001-6831-5429
Jan Kiesewetter
 https://orcid.org/0000-0001-8165-402X
Frank Fischer
 https://orcid.org/0000-0003-0253-659X

Elisabeth Bauer
Education and Educational Psychology
LMU Munich
Leopoldstr. 13
80802 Munich
Germany
elisabeth.bauer@psy.lmu.de