Using ENA to Analyze Pre-service Teachers' Diagnostic Argumentations: A Conceptual Framework and Initial Applications

Elisabeth Bauer[1(✉)], Michael Sailer[1], Jan Kiesewetter[2], Claudia Schulz[3], Jonas Pfeiffer[3], Iryna Gurevych[3], Martin R. Fischer[2], and Frank Fischer[1]

[1] Ludwig-Maximilians-University of Munich, Leopoldstr. 13, 80802 Munich, Germany
Elisabeth.Bauer@psy.lmu.de

[2] University Hospital, Ludwig-Maximilians-University of Munich, Pettenkoferstraße 8a, 80336 Munich, Germany

[3] Technical University of Darmstadt, Hochschulstraße 10, 64289 Darmstadt, Germany

Abstract. Diagnostic argumentation can be decomposed referring to the dimensions of content (see Toulmin 2003) and explicated strategy use indicated by epistemic activities (see Fischer et al. 2014). We propose a conceptual framework to analyze these two dimensions within diagnostic argumentation and explore its use within initial applications using the method of Epistemic Network Analysis (Shaffer 2017). The results indicate that both approaches of solely analyzing the dimension of content and solely analyzing the dimension of epistemic activities offer less insights into diagnostic argumentations than an analysis that includes both dimensions.

Keywords: Diagnosing · Argumentation · Pre-service teachers

1 A Diagnostic Argumentation Framework

Diagnosing is considered as a specialized type of scientific reasoning: When being confronted with a problem, diagnosticians generate hypotheses, gather and evaluate evidence for and against these and draw diagnostic conclusions (see Fischer et al. 2014). In this regard, it can be decomposed in the application of a diagnostic strategy and relevant concepts (Coderre et al. 2003). One example from medicine would be to specifically generate and evaluate evidence required for the conclusion or exclusion of several hypothesized differential diagnoses. Likewise, this strategy can be applied to various other domains and problem sets as for example in the context of teaching: Teachers need to diagnose single students' learning progress, understanding and learning preconditions. They also need to be able to communicate their diagnoses professionally (Lawson, Daniel 2011), e.g. to colleagues or in interdisciplinary professional communication. By relating arguments for and (if applicable) against differential diagnoses and rebutting potential counterarguments, they formulate a diagnostic argumentation.

In this paper, we therefore propose that it is a relevant learning objective for university students within several domains to formulate good diagnostic argumentations as a matter of professionalizing their communication. Corresponding interventions like role-play or peer-feedback are very resource-intensive and hardly applicable on a large scale (see Gartmeier et al. 2015). To approach this issue, it might be feasible to automate the analysis of diagnostic argumentations, for example using methods of natural language processing (Schulz et al. 2018). Such automation might be a useful basis for large-scale interventions. However, an automated analysis firstly requires a basic definition of and detailed insights into diagnostic argumentative structures, particularly regarding the quality characteristics of diagnostic argumentation. The literature and previous research in diagnosing rather focus on diagnostic decision-making processes and suggest different sequences of activities like the hypothetico-deductive type of reasoning (see Coderre et al. 2003). Such in-process reasoning consists of inferences, which are based on diagnostic schemata (Charlin et al. 2007). These process-related models assume a chronological temporality that is not necessarily applicable to post-hoc argumentation. Moreover, argumentation serves the purpose of explicitly presenting previous inferences in a comprehensible and conclusive manner (see Berland, Reiser 2009). We argue that comprehensibility and conclusiveness express in the explicated application of diagnostic strategy and concepts. Hence, analyzing the quality of diagnostic argumentations requires a framework that captures both, the application of strategy and concepts.

Since to our knowledge there is no such framework, we suggest the literature of argumentation schemes as a potential starting point. More specifically, we refer to the Toulmin argumentation model (Toulmin 2003) to analyze the application of concepts within a case-specific content dimension of diagnostic argumentation. The basic assumption of the Toulmin model is that conclusions (C) need to be grounded (G) and logically warranted (W) to represent an argument. In a more complex form, the scheme suggests to include more categories as for example evidence (E) that supports the ground (G), or a backing (B) of the warrant (W). The scheme becomes applicable to discursive argumentation by adding the element of the rebuttal (R) that attacks an initial conclusion (C_i) and might obtain its replacement or limitation by adding a qualifier (Q) to the final conclusion (C_f). The model is well transferable to diagnostic concepts as conclusions and rebuttals represent (differential) diagnoses that require being grounded on accumulated evidences; diagnostic conclusions also require being warranted by the knowledge of relations between symptoms and disorders. This knowledge is ideally backed by diagnostic guidelines or epidemiological data. Referring to the Toulmin model, the comprehensibility and conclusiveness of single arguments can be judged by several quality criteria as the acceptability, relevancy and sufficiency of the grounds in supporting the conclusion or the applicability of the warrant to the case under discussion (Toulmin 2003).

However, the Toulmin model does only account for the application of concepts within the case-specific content dimension of diagnostic argumentation. Analyzing the explication of a diagnostic strategy requires the integration of a second, strategy-related dimension within diagnostic argumentation. The strategy dimension can be conceptualized referring to epistemic activities (Fischer et al. 2014) as hypothesis generation (HG), evidence generation (EG), evidence evaluation (EE) and drawing of conclusions (DC). These activities were found to be relevant across a wide range of reasoning and argumentation (Hetmanek et al. 2018), e.g. the analysis of pre-service teachers' problem solving (Csanadi et al. 2018). The study found differences in the activity use of individual vs. collaborative learners, which became only observable by analyzing co-occurrences of epistemic activities. These were depicted as activity patterns by applying the method of Epistemic Network Analysis (Shaffer 2017). Building on this research, we suggest analyzing co-occurrences of diagnostic concepts and epistemic activities to extract similar patterns from diagnostic argumentations.

An exemplary mapping of epistemic activities on the structure of the Toulmin model (2003) is shown in Fig. 1. Introducing this framework we suggest that the structure of diagnostic argumentation might be better captured by analyzing the application of case-specific content and epistemic activities in combination rather than separately. This might point to more implicit aspects within diagnostic argumentations in terms of the application of strategy and concepts. For example a diagnostic conclusion that is linked to the activity of drawing conclusions (DC) might rather be considered as a final and therefore rather certain conclusion (C_f) which is sufficiently grounded; a diagnostic conclusion that co-occurs with the activity of generating evidence (EG) might indicate that the diagnostician considers the evidence to finally support or exclude a diagnosis as being insufficiently grounded. Therefore, we want to explore which insights about diagnostic argumentation can be gained from (1) analyzing co-occurrences solely within the dimension of content as well as from (2) analyzing co-occurrences solely within the dimension of epistemic activities and finally (3) from analyzing co-occurrences across both dimensions. We expect that the combination of both dimensions provides not only the insights of both separate analyses but might reveal some additional information about more implicit structures within diagnostic argumentation.

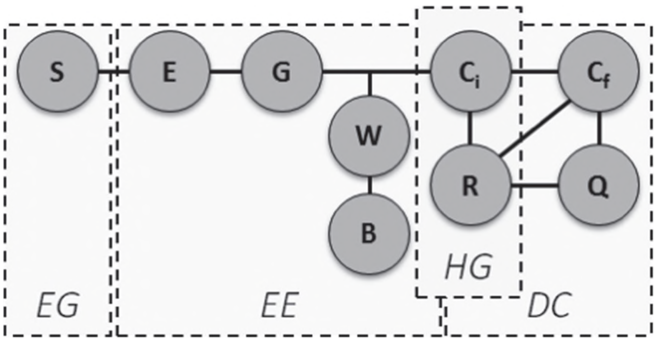

Fig. 1. Diagnostic Argumentation Framework. Abbreviations: S = Source of evidence. Adapted from Toulmin (2003): E = Evidence, G = Grounds, W = Warrant, B = Backing, C_i = initial conclusion, R = Rebuttal, Q = Qualifier, C_f = final conclusion. Adapted from Fischer et al. (2014): EG = generating evidence, EE = evaluating evidence, HG = generating hypotheses, DC = drawing conclusions.

2 Method

To explore the application of the framework we used data from a previous study. N = 118 pre-service teachers participated in a simulation-based training designed to foster their diagnostic reasoning and argumentation regarding school students' ADHD and dyslexia. The pre-service teachers were confronted with eight simulated cases about single school students displaying either behavioral problems or performance impairment specifically related to reading and writing. While working on the cases, they could generate pieces of evidence, for example by looking at reports of the school student's work and social behavior, reviewing exercises or tests, and by examining reports of conversations e.g. with the parents. The participants' final task was to indicate a diagnosis and formulate a diagnostic argumentation referring to their diagnostic strategy as well as evidences and diagnoses that they considered. We coded the resulting argumentations regarding two dimensions in two independent coding processes. In a first round, we coded the four diagnostic activities hypothesis generation (HG), evidence generation (EG), evidence evaluation (EE) and drawing conclusions (DC) (Fischer et al. 2014). For this purpose, fifteen percent of the argumentations were coded by four coders, resulting in an interrater reliability of Krippendorff's α_U = .65. In a second round, we coded the concept dimension, using a coding scheme that included 26 different diagnoses and 36 different categories of evidence. The coding manual offered specific coding recommendations for each of the eight cases on both sub-dimensions diagnoses and evidences. We coded fifteen percent of the argumentations per case with two coders each. Across the cases and sub-dimensions we reached interrater reliabilities ranging from Krippendorff's α_U = .73 to α_U = 1.00. Because of the case-specificity of the concept coding, we focus on the data of one training case for the current purpose. The third training case of the fictitious school student Anna displaying symptoms of ADD was considered as representative for the data because of its medium difficulty. The case-specific interrater reliability relating to the content dimension was Krippendorff's α_U = 1.00 for the diagnoses and Krippendorff's α_U = .81 for the evidences. For the current analysis, we accumulated the single categories by the correctness with respect to the principal case solution. This resulted in four sub-dimensions labeled with the codes "correct diagnoses" (diaC), "false diagnoses" (diaF), "correct evidences" (eviC) and "false evidences" (eviF).

To explore co-occurrences of codes within pre-service teachers' coded diagnostic argumentations of the training case Anna, we applied the method of Epistemic Network Analysis (Shaffer 2017). The ENA algorithm operationalizes sequences of codes by a moving window, which we defined as three lines, to construct a network model that shows the connections between codes. It further aggregates the resulting networks using a binary summation that reflects the presence or absence of co-occurrences of each pair of codes. The resulting networks are visualized using network graphs, where nodes correspond to codes and edges reflect the relative frequency of co-occurrence between each pair of codes. This results in two representations for each diagnostic argumentation: a plotted point in the low-dimensional projected space, plus a weighted network graph for every single argumentation. To make initial attempts in exploring the potential value of the combined dimensions concepts and epistemic activities suggested by the framework, we performed the ENA three times: (1) once only including the content dimension of the four codes correct and false evidences and diagnoses; (2) once only including the four epistemic activities hypothesis generation, evidence generation, evidence evaluation and drawing conclusions; (3) and once including all of these eight codes. To better examine the resulting networks, we created three plots for every of the tree analyses: The overall network across all participants' diagnostic argumentations for the respective case; one exemplary diagnostic argumentation formulated by the participant with the user ID 76N8058 who's argumentation quality was considered as high although the final diagnostic conclusion was false (see Table 1); and a comparison plot subtracting the overall network and the network of user 76N8058 indicating their discrepancies.

Table 1. Diagnostic argumentation formulated by the participant with the user ID 76N8058. Translated from German to English. Abbreviations: eviC = correct evidences, eviF = false evidences, diaC = correct diagnoses, diaF = false diagnoses, HG = hypothesis generation, EG = evidence generation, EE = evidence evalua-tion, DC = drawing conclusions.

	HG	EG	EE	DC	diaC	diaF	eviC	eviF
First, I observe Anna during class and initially I considered it to be only a temper, but over a longer period it became clear to me that she had a problem	0	1	1	0	0	0	0	0
For this reason, I had a look on her performance records, which corresponded to her general attention and her German grades	0	1	1	0	0	0	0	0
Then I talked to other subject teachers about Anna's problems	0	0	1	0	0	0	0	0
That made me reject my initial suspicion that she might have dyslexia	0	0	0	1	1	0	0	0
Eventually, she might have ADD, so I observe her during school recess, particularly her interaction with other kids, and talk to her mother	1	1	0	0	1	0	0	0
Her behavior during recess doesn't match the pattern of ADD, however, I took notice	0	0	1	0	0	1	0	0

(*continued*)

Table 1. (*continued*)

	HG	EG	EE	DC	diaC	diaF	eviC	eviF
of the problematic family environment								
Perhaps, the financial worries and absence of her mother are causal for her performance drop and her flight into her daydreaming and drawing	1	0	0	0	0	0	0	1
As a next step, I would consult a psychologist who might talk to Anna as well and check if my suspicion was correct	0	1	0	0	0	0	0	0
Besides, I make him aware of her seemingly depressive characteristics that he might keep an eye on	0	0	1	0	1	0	0	0

3 Results

Figure 2 shows the resulting plots of the first analysis solely including the content dimension. In Fig. 2a we see that the two strongest lines connect the codes correct and false evidences as well as correct evidences and correct diagnoses. Accordingly, in the context of three lines, correct evidences mostly co-occurred with false evidences and correct diagnoses. These co-occurrences represent firstly the selection and integration of different pieces of evidence in a grounding and secondly the grounding of correct diagnostic conclusions by acceptable pieces of evidence. The second strongest line is depicted between correct evidences and false diagnoses, which represents again the grounding of conclusions. In this case, the line can either indicate that false diagnostic conclusions were partially supported by correct evidences; however, it is more likely that the line represents the correct exclusion of false diagnoses grounded on correct evidences. The argumentation network of user 76N8058 shown in Fig. 2b presents a very different pattern of a triangle relating false diagnoses, false evidences and correct diagnoses. The network omits the code of correct evidences. Therefore, the user only related false evidences to correct and incorrect diagnoses. Integrating only the false evidences in a ground, the user most likely made a logically acceptable and yet false conclusion by accepting false diagnoses and excluding correct diagnoses.

The results of the second analysis solely considering the dimension of epistemic activities are shown in Fig. 3. In the overall network across all users (3a) we see that the thickest line relates the activities evidence evaluation and drawing conclusions. This relation shows that the evaluations performed were mostly considered as sufficient to draw rather certain conclusions. The second strongest relation is depicted between

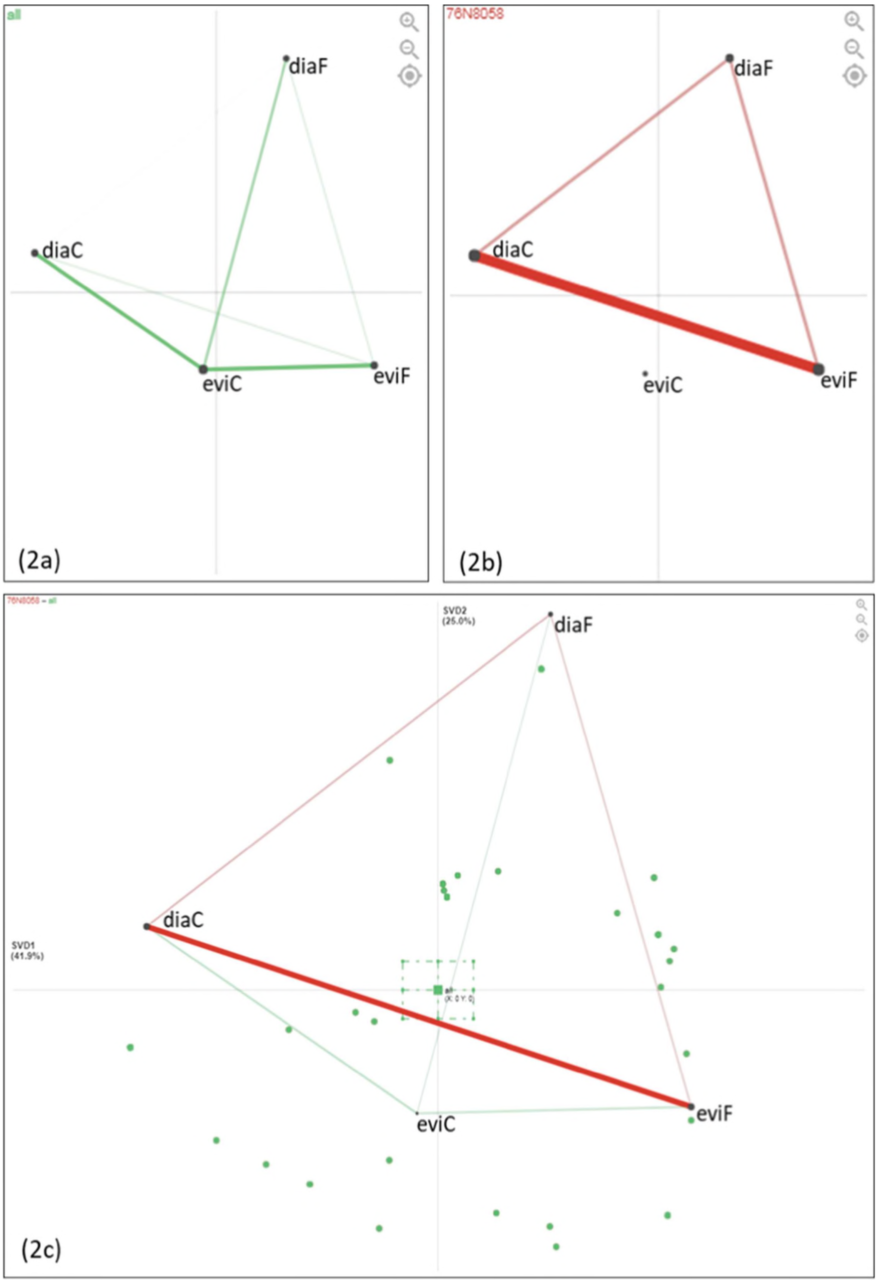

Fig. 2. Results of the first analysis including the content dimension only. Epistemic networks across all users (2a), one exemplary user (2b) and the difference between the networks (2c). Abbreviations: eviC = correct evidences, eviF = false evidences, diaC = correct diagnoses, diaF = false diagnoses.

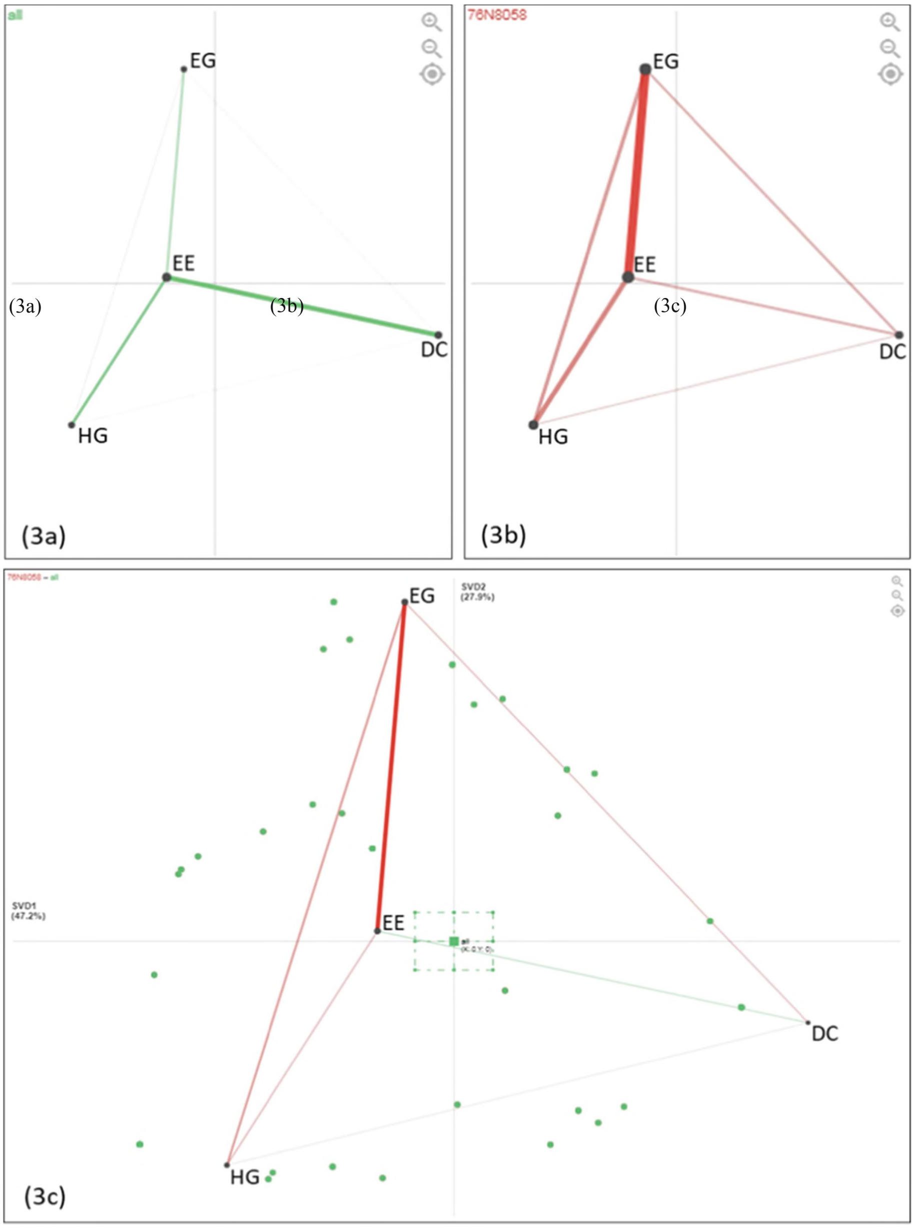

Fig. 3. Results of the second analysis including the epistemic activity dimension only. Epistemic networks across all users (3a), one exemplary user (3b) and the difference between the networks (3c). Abbreviations: HG = hypothesis generation, EG = evidence generation, EE = evidence evaluation, DC = drawing conclusions.

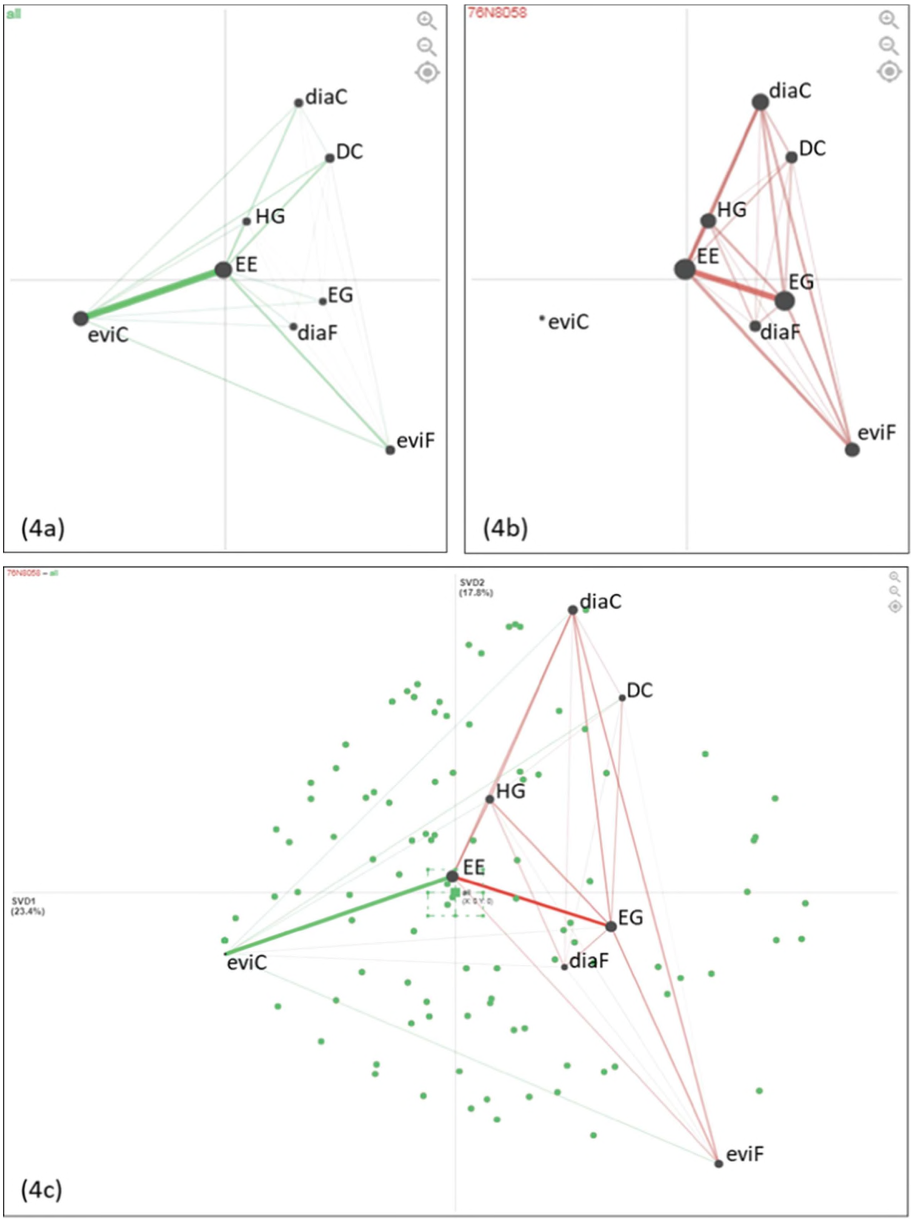

Fig. 4. Results of the third analysis including both dimensions. Epistemic networks across all users (4a), one exemplary user (4b) and the difference between the networks (4c). Abbreviations: eviC = correct evidences, eviF = false evidences, diaC = correct diagnoses, diaF = false diagnoses, HG = hypothesis generation, EG = evidence generation, EE = evidence evaluation, DC = drawing conclusions.

evidence evaluation and hypothesis generation. This line represents a higher degree of uncertainty in interpreting evaluated evidences. Looking at the argumentation network of user 76N8058 (3b) we see again some differences from the overall network. Above all, user 76N8058 rather refers to the activity of evidence generation compared to other users (3c). Hence, evidence generation and the other three epistemic activities are stronger related in the argumentation network of 76N8058. In addition, the link between hypothesis generation and evidence evaluation is slightly more apparent which indicates less certainty in concluding with respect to the evaluated evidences.

Figure 4 shows the resulting networks of the third analysis, which combines the two dimensions of content and epistemic activities. The resulting overall network (4a) shows the strongest line between evidence evaluation and correct evidences. This main line forms three triangles with false evidences, with drawing conclusions and with correct diagnoses. This shows that the majority of argumentations focus on the evaluation of the correct evidences but also evaluate false evidences partially with respect to correct evidences. These evaluations are used as grounding for the correct diagnoses, with more and less degrees of certainty. Comparing the overall network (4a) with the network of user 76N8058 (4b), one of the most obvious differences is the absence of the link between evidence evaluation and correct evidences; this is due to the general absence of the evaluation of the code for correct evidence which we also saw as a result from the first analysis including only the content dimension. All of the other codes co-occur within the diagnostic argumentation 76N8058. Apart from the absence of correct evidences, all lines from the overall network (4a) are also visible in the argumentation of user 76N8058 (4b). Nevertheless, the combination of both dimensions in this third analysis provides two additional insights in the user's diagnostic argumentation, particularly when considering the comparison plot of subtracted networks (4c). Firstly, we see that in argumentation 76N8058 (4b) both correct and false diagnoses rather co-occur with hypothesis generation as compared to other users (see 4c); at the same time the strength of the lines between diagnoses and the activity of drawing conclusions are comparable between network 76N8058 and the overall network (see 4c). Argumentation 76N8058 therefore expresses a comparably higher proportion of uncertainty, particularly with respect to the correct diagnoses but also regarding the false diagnoses. Secondly, we see a stronger link between all of the other codes (apart from correct evidences) with evidence generation. This represents a more transparent explanation of how the user generated the false evidences; moreover, the lines linking evidence generation with correct and false diagnoses show that user 76N8058 points to the necessity of generating further evidence as grounding for any final diagnostic conclusion.

4 Discussion

From the results of the first analysis solely including the content dimension we can draw conclusions about pre-service teachers' case-specific application of diagnostic concepts within diagnostic argumentation. Focusing on the content dimension allows comparing a diagnostician's performance to the correct answer or an expert solution. By applying the Toulmin model (2003) to the content dimension within diagnostic argumentation, we can progress beyond the correctness of a solution as a single indicator for diagnostic argumentative quality. In our first analysis we saw that the pre-service teacher with the user ID 76N8058 drew the false diagnostic conclusion probably because he or she missed central pieces of correct evidence. However, the false diagnostic conclusion was grounded on evidences. We argue to further focus on analyzing diagnostic argumentation in terms of quality criteria as for example the acceptability, relevancy and sufficiency of the grounds in supporting a conclusion (Toulmin 2003). These criteria may facilitate the analysis of single arguments' comprehensibility and conclusiveness, independently from the criteria of correctness.

The results of the second analysis solely considering the dimension of epistemic activities (see Fischer et al. 2014) in diagnostic argumentation presents insights into the explicated application of a diagnostic strategy. We see for example that pre-service teachers' single argumentations can differ largely in their report of evidence generation. However, reporting activities of evidence generation may indicate higher transparency regarding the quality of evidences for example in terms of appropriate application of methods or selection of sources of information (see Fischer et al. 2014). Moreover, analyzing epistemic activities in diagnostic argumentation provides information about different degrees of certainty in diagnosing. The exemplary analysis indicated that pre-service teachers differ in terms of the proportion to which they combine evaluative with hypothetical or concluding parts within their diagnostic argumentations. This result might even go beyond pointing to the sufficiency of the available evidences to draw a conclusion; the indication of uncertainty in solving a case might also partially reflect a subjective self-evaluation. We suggest to further research this question by manipulating variables like prior knowledge and complexity of the diagnostic task (see Charlin et al. 2007).

The results of the third analysis combining the two dimensions of content and epistemic activities capture the same insights as the previously discussed analyses referring to the two dimensions separately. Moreover, it combines the advantages of analyzing the two dimensions for example by indicating the certainty in referring to a correct or false diagnostic conclusion with respect to a sufficient grounding on correct or false evidences. Co-occurrences of certain epistemic activities with certain diagnostic concepts may also be able to indicate rather implicit assumptions within diagnostic argumentations. One example is the argumentative relation between diagnoses and the activity of evidence generation. This refers to the prior or further necessity of generating evidence as grounding for a final diagnostic conclusion. Consequently, the evaluated evidence was considered as insufficient for drawing a diagnostic conclusion. This reflects an implicit evaluation of certain evidences that might not necessarily be explicitly stated.

It has to be noted that the validity of the codes relating to the content dimension is very limited due to the simplifying reduction resulting in the four codes of case-specifically correct and false evidences and diagnoses. We precede our analyses by stronger differentiating codes within the content dimension. These could be additionally grouped in terms of other characteristics as for example their significance for diagnosing the respective case (see Charlin et al. 2007). It might also be interesting to add a code for the "source of evidence" (see Nicolaidou et al. 2011) as a corresponding content-related code for the activity of evidence generation.

Concluding from the previous considerations, we consider the presented initial applications of ENA to analyze diagnostic argumentations referring to the proposed framework (Fig. 1) as a promising approach to better understand diagnostic argumentations of pre-service teachers and eventually other diagnosticians.

References

Berland, L.K., Reiser, B.J.: Making sense of argumentation and explanation. Sci. Educ. **93**(1), 26–55 (2009)

Charlin, B., Boshuizen, H., Custers, E.J., Feltovich, P.J.: Scripts and clinical reasoning. Med. Educ. **41**(12), 1178–1184 (2007)

Coderre, S., Mandin, H., Harasym, P.H., Fick, G.H.: Diagnostic reasoning strategies and diagnostic success. Med. Educ. **37**(8), 695–703 (2003)

Csanadi, A., Eagan, B., Kollar, I., Shaffer, D.W., Fischer, F.: When coding-and-counting is not enough: using epistemic network analysis (ENA) to analyze verbal data in CSCL research. Int. J. Comput.-Support. Collaborative Learn. **13**(4), 419–438 (2018)

Fischer, F., et al.: Scientific reasoning and argumentation: advancing an interdisciplinary research agenda in education. Frontline Learn. Res. **2**(3), 28–45 (2014)

Gartmeier, M., et al.: Fostering professional communication skills of future physicians and teachers: effects of e-learning with video cases and role-play. Instr. Sci. **43**(4), 443–462 (2015)

Hetmanek, A., Engelmann, K., Opitz, A., Fischer, F.: Beyond intelligence and domain knowledge. Scientific reasoning and argumentation as a set of cross-domain skills. In: Fischer, F., Chinn, C.A., Engelmann, K., Osborne, J. (eds.) Scientific Reasoning and Argumentation: The Roles of Domain-Specific and Domain-General Knowledge, pp. 203–226. Routledge, New York (2018)

Lawson, A.E., Daniel, E.S.: Inferences of clinical diagnostic reasoning and diagnostic error. J. Biomed. Inform. **44**(3), 402–412 (2011)

Nicolaidou, I., Kyza, E.A., Terzian, F., Hadjichambis, A., Kafouris, D.: A framework for scaffolding students' assessment of the credibility of evidence. J. Res. Sci. Teach. **48**(7), 711–744 (2011)

Schulz, C., et al.: Challenges in the Automatic Analysis of Students' Diagnostic Reasoning. arXiv preprint arXiv:1811.10550 (2018)

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

Toulmin, S.E.: The Uses of Argument. Cambridge University Press, Cambridge (2003)