International Society of the Learning Sciences

Pre-Service Teachers' Argumentations in the Context of Assessment

Elisabeth Bauer, Ludwig Maximilian University of Munich, elisabeth.bauer@psy.lmu.de
Michael Sailer, Ludwig Maximilian University of Munich, michael.sailer@psy.lmu.de
Jan Kiesewetter, University Hospital, LMU Munich, jan.kiesewetter@med.uni-muenchen.de
Martin R. Fischer, University Hospital, LMU Munich, martin.fischer@med.uni-muenchen.de
Frank Fischer, Ludwig Maximilian University of Munich, frank.fischer@psy.lmu.de

Abstract: In this article, we introduce a conception of diagnostic argumentation skills and define three relevant facets, namely justification, disconfirmation and transparency. We present initial investigations in the three facets' (co-)occurrences in diagnostic argumentation and their relation with the accuracy of diagnostic decisions in the context of teacher education. For this purpose, we analyze data of 118 pre-service teachers, who were learning with simulated cases concerned with the topic of students' clinical problems. We interpret the results as justification, disconfirmation and transparency being three distinct facets of diagnostic argumentation skills. We also assume that justification may relate to cross-domain transferable argumentation skills, while disconfirmation and transparency are rather context-specific aspects of diagnostic argumentation, relating to standards in diagnosing. Moreover, we discuss the type of reasoning and prior processing of the case as potential explanation for a correlation found between justification in diagnostic argumentation and the accuracy of diagnostic decisions.

Pre-service teachers' diagnostic argumentation and diagnostic accuracy

Skills of diagnostic reasoning and argumentation are relevant in many disciplines (Heitzmann et al., 2019). In this article, we outline a conception of diagnostic argumentation skills and define three relevant facets, namely justification, disconfirmation and transparency. Furthermore, we present initial empirical investigations concerning their (co-)occurrences in diagnostic argumentation and their relation with the accuracy of diagnostic decisions in the context of teacher education.

The learning of diagnostic skills is a matter of professionalization within various disciplines (Heitzmann et al., 2019). Beyond medicine and medical education, also other disciplines like teacher education emphasize the role of diagnosing as "a process of goal-oriented collection and integration of case-specific information [aiming] to reduce uncertainty in order to make […] decisions", but use other terminology like *assessment* (Heitzmann et al., 2019, p. 4). Teachers need to diagnose students' performance, progress, and among other learning prerequisites, initially identify learning difficulties such as dyslexia or even behavioral disorders like ADHD (Praetorius et al., 2013; Reinke et al., 2011). Indeed, higher education programs in teaching and other disciplines increasingly acknowledge the relevance of diagnostic skills (Heitzmann et al., 2019). Simultaneously, research places a stronger emphasis on the facilitation of diagnostic skills, e.g., by exploring the effects of simulation-based learning on facilitating diagnostic skills (Chernikova et al., 2019). Advancing the understanding of the nature and relation of diagnostic skills is an essential foundation for learning and researching diagnostic skills.

Diagnostic skills reflect the application of professional diagnostic knowledge, e.g., in terms of effective diagnostic action and accurate diagnostic decisions (Heitzmann et al., 2019). Effective diagnostic action is researched e.g., in terms of epistemic activities like generating hypotheses, generating and evaluating evidence, and drawing conclusions (Fischer et al., 2014). However, research and education generally refer to the accuracy of diagnostic decisions as main indicator of diagnostic skills (e.g., Praetorius et al., 2013). The literature suggests that reasoning processes, which are underlying diagnostic action and diagnostic decision-making, can be either of fast and intuitive or of slow and reflective nature (Mercier & Heintz, 2013). While reflective inferences are conscious and explicable, intuitive inferences are less accessible for explication (Mercier & Heintz, 2013). Nonetheless, communicating diagnostic considerations, activities and decisions is crucial in various situations like collaborative diagnosing or verbalized post-hoc argumentation, e.g., when writing a report. In such situations, diagnostic reasoning is ideally explicated in the form of coherently interrelated arguments (Walton, 1990). We broadly subsume such instances of interrelated diagnostic arguments as *diagnostic argumentation*. Moreover, we assume that diagnostic argumentation does pursue different epistemic aims: Diagnostic argumentation may for example aim for sense-making during diagnostic action as well as for articulating or persuasively presenting diagnostic decisions (Berland & Reiser, 2008; Chinn, Rinehart, & Buckland, 2014). So far, research investigating verbal outcomes in diagnosing mostly focused on justificatory accuracy, e.g., in verbalized sense-making during the diagnostic process (e.g., Braun et al., 2018) or in articulating diagnostic decisions (e.g., Praetorius et al., 2013).

In our research, we build on the broader literature on scientific argumentation, which pointed out that beyond accuracy there are other, especially structural features contributing to an argumentation's persuasiveness (e.g., Hitchcock, 2005). It is not clear however, if structural features of diagnostic argumentation necessarily correlate with the accuracy of diagnostic decisions or if they reflect distinct diagnostic skills. Qualitative findings indicated that formal aspects of presenting arguments about diagnostic decisions may not necessarily correlate with accurate diagnostic decisions; instead, structural features of diagnostic argumentation may rather depend on additional skills (Braun et al., 2018), like cross-domain argumentation skills (Hetmanek et al., 2018). In addition, diagnostic argumentation may be a matter of internalized standards that are specific to the context of diagnosing (Bauer et al., 2020; Chinn, Rinehart, & Buckland, 2014). We conclude that further research is needed to investigate the nature of diagnostic argumentation and its distinction from other diagnostic skills, particularly accurate diagnostic decision-making. For this purpose, we outline a conception of diagnostic argumentation skills in the following and define three relevant facets, namely justification, disconfirmation and transparency.

The majority of studies considering argumentation structure refer to Toulmin´s argumentation scheme (2003). This model especially highlights the role of *justification*, since any claim requires some grounding to build a complete argument. Transferred to the context of diagnostic argumentation, diagnoses are comparable to claims that need to be justified with evidence. Moreover, the Toulmin scheme (2003) emphasizes the role of rebutting counterarguments and competing claims in supporting an overall conclusion. This rationale functionally resembles the scientific approach of *disconfirmation*: In the case of two competing hypotheses, a finding of evidence disconfirming one hypothesis supports in turn the other hypothesis. Disconfirmation is particularly relevant in cases that involve uncertainty, e.g., due to an incomplete data basis, thus requiring probabilistic reasoning (Kind & Osborne, 2017). Simultaneously, rejecting relevant differential diagnoses supports the plausibility of a diagnostic conclusion, especially in diagnosing cases that involve ambiguous evidence. Furthermore, elaborating on scientific argumentation and standards in science, other major themes are the reliability of processes and the evaluation of evidence in reference to methodology and data sources (Chinn, Rinehart, & Buckland, 2014; Fischer et al., 2014). Accordingly, *transparency* with respect to the processes is considered essential in underlining the quality of the presented evidence and conclusions (Vazire, 2017). In diagnostic argumentation, this is achieved by describing the processes undertaken to generate evidence.

We assume that justification, disconfirmation and transparency are three relevant facets of diagnostic argumentation. It is not clear however, if they are correlated, reflecting an overall diagnostic argumentation skill, or if they are distinct diagnostic argumentation skills. Therefore, to approach this issue, we explore how justification, disconfirmation and transparency (co-)occur in diagnostic argumentation (RQ1). Moreover, referring to the question of distinguishing diagnostic argumentation from diagnostic decision-making, we investigate, how justification, disconfirmation and transparency in diagnostic argumentation relate to diagnostic accuracy (RQ2).

Methods

We recruited 118 pre-service teachers to participate in a study that evaluated a simulation-based learning environment concerned with the topic of students' clinical problems that we developed for a teacher education program. The participants processed eight simulated cases of primary and secondary students displaying behavior that might indicate a disorder in the spectrum of either ADHD or dyslexia. Doing so, they could access several informational sources such as samples of the student's written exercises and school certificates, reports of observations from inside and outside of the classroom, conversations with the respective student, the parents, and other teachers. Participants had two tasks per case: (A) they had to make a diagnosis by indicating whether the simulated student might have a clinical problem and if so, which one it might be; (B) we asked the participants to write an argumentation about their conclusion and their approach in reasoning about the case. The data collection was computer-based and took place in a laboratory setting. We introduced participants to the aims and procedure of the study and familiarized them with the simulation-based learning environment. After giving informed consent to participate in the study, they answered a pretest that took around 35 minutes. Afterwards they entered the learning phase, consisting of the eight simulated cases. Time on task for all cases was M = 51.8 minutes (SD = 16.5). After four cases, participants took a break of ten minutes before continuing with solving cases five to eight. Subsequently, they answered a posttest, which took again around 35 minutes.

The data sources used for the analyses presented in this paper are the diagnoses (A) and the diagnostic argumentations (B) from six of the eight cases. We coded diagnostic accuracy of all the written diagnoses (A) as accurate (1 point), partially accurate (0.5 point) or inaccurate (0 points). Two raters coded 12.5% of the diagnoses, resulting in an inter-rater reliability of Cohen's κ = .80. The internal consistency across the six cases was rather low (McDonald's ω = .37). For further analyses, we calculated a sum score of the points achieved in terms of diagnostic accuracy with a possible range from 0 to 6 points.

Afterwards, we coded diagnostic argumentations (B) in two independent rounds of coding. First, we segmented and coded all argumentations in terms of the four epistemic activities *generating hypotheses, generating evidence, evaluating evidence* and *drawing conclusions*. Four raters coded 15% of the data in parallel before dividing the rest of the data. Subsequent to the initial coding, we decided to include *generating hypotheses* into *drawing conclusions*, resulting in a merged category of *drawing conclusions* (Krippendorff's a_U = .62) and the two additional coding categories *generating evidence* (Krippendorff's a_U = .56) and *evaluating evidence* (Krippendorff's a_U = .75). In a second round, we coded the diagnostic argumentations of six cases regarding the content dimension of *differential diagnoses*. 15% was double coded resulting in an overall inter-rater reliability of Cohen's κ = .92. We operationalized *justification* in diagnostic argumentation as one or more co-occurrences of evaluating evidence and drawing conclusions within a temporal context of two sentences. Moreover, we considered *disconfirmation* in diagnostic argumentation to apply if the argumentation included at least two differential diagnoses. Thirdly, we defined *transparency* in diagnostic argumentation as at least one explication of generating evidence. Internal consistencies across six cases was satisfactory, with McDonald's ω = .60 for justification, McDonald's ω = .60 for disconfirmation, and McDonald's ω = .71 for transparency. For all three variables, we calculated again sum scores with a possible range from 0 to 6 points.

To explore both research questions, we used Pearson's correlation analysis, including the variables justification, disconfirmation, transparency, and diagnostic accuracy. The alpha level was defined as α = .05.

Results

To explore, how justification, disconfirmation and transparency (co-)occur in diagnostic argumentation (RQ1), we report descriptive results and a Pearson correlation analysis. Across the six cases, participants put most emphasis on justification, which we found on average in around two thirds of their diagnostic argumentations (M = 3.83; SD = 1.58). Moreover, we found that participants hardly applied disconfirmation, which only occurred in around a quarter of the diagnostic argumentations (M = 1.52; SD = 1.41). We found transparency in around half of the diagnostic argumentations (M = 2.67; SD = 1.81), which is considerably less frequent than justification but more frequent than disconfirmation. Pearson correlation analysis indicated that justification and disconfirmation may be related, since the correlation was significant with a large effect (r = .568, p < .01). In contrast, transparency was not significantly correlated with justification (r = .055, p = .55) or disconfirmation (r = .025, p = .79).

To investigate, how justification, disconfirmation and transparency in diagnostic argumentation relate to diagnostic accuracy (RQ2), we report again a Pearson correlation analysis. Participants' average diagnostic accuracy was M = 4.41 (maximum of 6 achievable points; SD = .94). We found diagnostic accuracy and justification to correlate significantly, with a small effect (r = .284, p < .01). However, there were no significant correlations between diagnostic accuracy and disconfirmation (r = .105, p = .26) as well as transparency (r = .059, p = .53), indicating that diagnostic accuracy discriminated between justification and disconfirmation.

Discussion

Exploring the (co-)occurrences of justification, disconfirmation and transparency in diagnostic argumentation (RQ1), we found that pre-service teachers rather applied justification in their diagnostic argumentations than disconfirmation and transparency. Since there was no specific instruction that prompted students to emphasize justification in particular, we interpret that justification is less specific to the context of diagnosing, but may relate to a more generalizable skill level like cross-domain transferable argumentation skills (Hetmanek et al., 2018). Compared to justification, especially disconfirmation but also transparency were less prevalent in diagnostic argumentations. We suggest considering them as more specific argumentation skills (Hetmanek et al., 2018) that relate to standards in diagnosing or even broader standards of science (Chinn, Rinehart, & Buckland, 2014). Previous research has issued that such diagnostic standards may be insufficiently taught in teacher education (Bauer et al., 2020). Moreover, despite the finding of justification correlating highly with disconfirmation, we also found that diagnostic accuracy correlated only with justification but not with disconfirmation, thus discriminating between the two facets. Overall, we interpret that justification, disconfirmation and transparency seem to be rather independent diagnostic argumentation skills. In contrast to disconfirmation and transparency, justification significantly correlated with diagnostic accuracy (RQ2). Since the effect was only small, we still consider justification in diagnostic argumentation and the accuracy in diagnostic decision-making as two different skills. We assume that the correlation may be explained by reflective inferences made during the prior processing of the case (Mercier & Heintz, 2013): Reflective inferences are a conscious and explicable type of reasoning, which may facilitate both justification in diagnostic argumentation and accuracy in diagnostic decision-making.

Some methodological remarks seem relevant for our interpretation: Inter-rater reliabilities of generating evidence and drawing conclusions are rather low. However, the initial coding reflects the simultaneous segmentation and coding of individual epistemic activities; hence, we consider the agreement as sufficient for the

combined task. Moreover, for our operationalization, we did not use the initially coded individual epistemic activities, but abstracted their overall presence or absence, on the level of diagnostic argumentations. Thus, we suppose that the actual agreement on the analyzed level of diagnostic argumentations is most likely superior to the agreement presented. A second limitation consists in the low internal consistency of diagnostic accuracy. This issue is in line with other studies suggesting that the accuracy of diagnostic-decision-making is comparably content-specific and thus, internal consistency of diagnostic accuracy only increases across a large number of cases (Monteiro et al., 2020).

Based on our initial considerations and results presented in this paper, we suggest that justification, disconfirmation and transparency represent three distinct facets of diagnostic argumentation skills. We derive the assumption that justification relates to cross-domain transferable argumentation skills, while disconfirmation and transparency are rather context-specific aspects of diagnostic argumentation. We also assume that reflective inferences made during the processing of the case may facilitate both justification in diagnostic argumentation and accuracy in diagnostic decision-making. Justification in diagnostic argumentation and accuracy of diagnostic decision-making seem to be distinct outcomes, reflecting different diagnostic skills. Therefore, as a matter of further professionalization, research and learning programs interested in diagnosing may consider diagnostic argumentation skills as relevant outcomes and distinguish them from established indicators for diagnostic skills, like accurate diagnostic decisions.

References

Bauer, E., Fischer, F., Kiesewetter, J., Shaffer, D. W., Fischer, M. R., Zottmann, J. M., & Sailer, M. (2020). Diagnostic Activities and Diagnostic Practices in Medical Education and Teacher Education: An Interdisciplinary Comparison. *Frontiers in Psychology*, *11*(562665).

Berland, L. K., & Reiser, B. J. (2008). Making sense of argumentation and explanation. *Science education*, *93*(1), 26-55.

Braun, L. T., Lenzer, B., Kiesewetter, J., Fischer, M. R., & Schmidmaier, R. (2018). How case representations of medical students change during case processing–results of a qualitative study. *GMS journal for medical education*, *35*(3).

Chernikova, O., Heitzmann, N., Fink, M. C., Timothy, V., Seidel, T., & Fischer, F. (2020). Facilitating diagnostic competences in higher education - A meta-analysis in medical and teacher education.

Chinn, C. A., Rinehart, R. W., & Buckland, L. A. (2014). Epistemic cognition and evaluating information: Applying the AIR model of epistemic cognition. *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences*, 425-453.

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., ... & Strijbos, J. W. (2014). Scientific Reasoning and Argumentation: Advancing an Interdisciplinary Research Agenda in Education. *Frontline Learning Research*, *2*(3), 28-45.

Heitzmann, N., Seidel, T., Opitz, A., Hetmanek, A., Wecker, C., Fischer, M., ... & Stürmer, K. (2019). Facilitating Diagnostic Competences in Simulations: A Conceptual Framework and a Research Agenda for Medical and Teacher Education. *Frontline Learning Research*, *7*(4), 1-24.

Hetmanek, A., Engelmann, K., Opitz, A., & Fischer, F. (2018). Beyond intelligence and domain knowledge: Scientific reasoning and argumentation as a set of cross-domain skills. *Scientific reasoning and argumentation: The roles of domain-specific and domain-general knowledge.* New York: Routledge.

Hitchcock, D. (2005). Good reasoning on the Toulmin model. *Argumentation*, *19*(3), 373-391.

Kind, P. E. R., & Osborne, J. (2017). Styles of scientific reasoning: a cultural rationale for science education?. *Science Education*, *101*(1), 8-31.

Mercier, H., & Heintz, C. (2014). Scientists' argumentative reasoning. *Topoi*, *33*(2), 513-524.

Monteiro, S. D., Sherbino, J., Schmidt, H., Mamede, S., Ilgen, J., & Norman, G. (2020). It's the destination: diagnostic accuracy and reasoning. *Advances in Health Sciences Education*, *25*(1), 19-29.

Praetorius, A. K., Berner, V. D., Zeinz, H., Scheunpflug, A., & Dresel, M. (2013). Judgment confidence and judgment accuracy of teachers in judging self-concepts of students. *The Journal of Educational Research*, *106*(1), 64-76.

Reinke, W. M., Stormont, M., Herman, K. C., Puri, R., & Goel, N. (2011). Supporting children's mental health in schools: Teacher perceptions of needs, roles, and barriers. *School Psychology Quarterly*, *26*(1), 1.

Toulmin, S. E. (2003). *The uses of argument*. Cambridge university press.

Vazire, S. (2017). Quality uncertainty erodes trust in science. *Collabra: Psychology*, *3*(1).

Walton, D. N. (1990). What is reasoning? What is an argument?. *The Journal of Philosophy*, *87*(8), 399-419.