Judgment bias in diagnosing misconceptions with decimal fractions

Andreas Rieu, Timo Leuders, Katharina Loibl

Judgment bias in diagnosing misconceptions with decimal fractions

Andreas Rieu[1], Timo Leuders and Katharina Loibl

[1]University of Education, Freiburg, Germany; andreas.rieu@ph-freiburg.de

A prerequisite for remediating student misconceptions is their accurate diagnosis by the teacher. However, studies on judgment accuracy show that teachers differ substantially regarding their judgments even though little is known about the reasons so far. The present study investigates whether teachers' diagnosis of misconceptions with decimal fractions is subject to judgment biases. For this purpose, we propose a cognitive model for diagnosing misconceptions based on the process of hypothesis testing. The study results show that the formulation of alternative hypothesis and the processing of relevant information are predictors of high judgment accuracy when diagnosing misconceptions. Furthermore, normative and confirmatory biased judgment processes could be distinguished. Implications for teacher education are discussed.

Keywords: Assessment, misconceptions, confirmation bias, cognitive processes.

Judgment processes in diagnosing misconceptions

The term misconception suggests that students commit a systematic error due to a naïve theory. From a student's point of view, it is rather a strategy based on a hitherto reasonable idea. In mathematics education, various student misconceptions are well researched (cf. Confrey & Kazak, 2006), such as the assumption that "multiplication makes bigger, division makes smaller" which is correct with natural numbers and is thus often overgeneralised to fractions. From a teachers' perspective, misconceptions are constructs that emerged as a result of the learners' experiences in different contexts but that no longer function correctly when transferred to another area of knowledge (Fujii, 2014). In order to create adaptive learning opportunities, learner misconceptions must first be diagnosed. Then the teacher can trigger a cognitive conflict and resolve it by introducing the actual mathematical concept (Corno, 2008). Thus, teachers' judgment accuracy when diagnosing misconceptions – that is to determine it precisely - seems crucial for student learning. However, various studies have shown that teachers' judgment accuracy on student performance varies widely (mean effect size between teachers' judgments of students' academic achievement and students' actual academic achievement of r = 0.63 in the meta-analysis by Südkamp, Kaiser, & Möller, 2012).

In their framework, Loibl et al. (2020) conceptualise diagnostic judgments in pedagogical contexts as a teacher's inference about learners (e.g., their abilities) or materials (e.g., task difficulty) based on the information that is explicitly or implicitly present in a diagnostic situation. This definition locates diagnostic judgments within the larger field of social judgment and cognitive information processing and allows investigating the genesis of (correct and incorrect) diagnostic judgments. In this line of research, recent studies focus their research interest on the judgment processes and examine which information teachers actually gather and process to form their judgment (e.g., Rieu et al., 2022).

The judgment process can be influenced by personal expectations (often leading to erroneous diagnoses), which have already been documented in the area of ethnicity, socioeconomic status, and

gender (McKown & Weinstein, 2003; Rubie-Davies, Hattie, & Hamilton, 2006; Südkamp et al., 2012). Biases like the confirmation bias – the tendency to selectively choose and process information supporting the initial hypothesis – influence the judgment process when the diagnostic judgments are based on hypothesis testing (Herppich et al., 2018; Oswald & Grosjean, 2004; Westhoff & Kluck, 2014).

The current study

One diagnostic situation in which the teacher's cognitive processes can be modelled as hypothesis testing (Trope & Liberman, 1996), is the detection of misconceptions in decimal fraction (e.g., *longer-is-larger, shorter-is-larger,* Stacey, 2005). In this process, an erroneously solved task often cannot be clearly assigned to one single misconception, but only the structured processing of several tasks and its solutions by the student allow the precise diagnosis of the misconception.

We assume that cognitive biases occur in this knowledge-based process and that these biases systematically favour the confirmation of a hypothesis and make its rejection unlikely (overestimation of the a priori probability of the hypothesis, selective gathering of hypothesis-confirming information, hypothesis-consistent interpretation of ambiguous information according to Schulz-Hardt & Köhnken, 2000, figure 1).

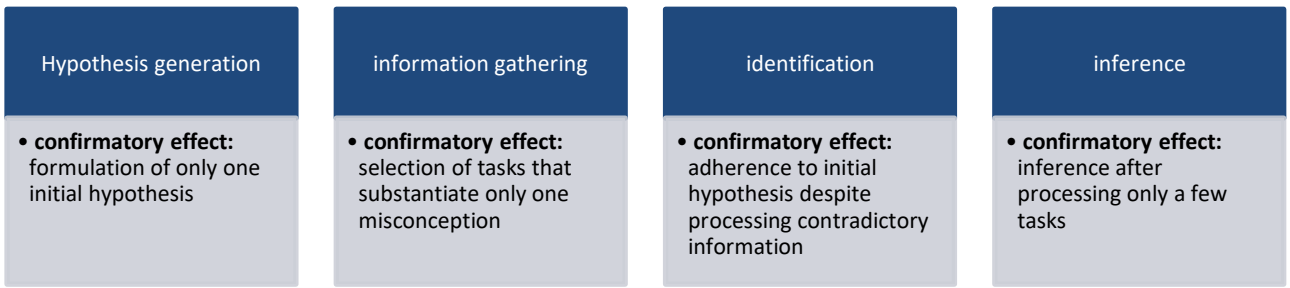

Figure 1: Hypothesis testing process and confirmatory effects in the domain of diagnosis misconceptions

The present study defines diagnostic judgments as information processing and analyses the causes of diagnostic errors. The collection of external indicators such as the formulation of the initial hypothesis and the number and type of information processed allow conclusions to be drawn about the genesis of diagnostic judgments (Loibl et al., 2020).

Specifically, it is assumed that in the ambiguous diagnostic situation of detecting misconceptions in decimal fraction comparison, confirmatory biases occur and prevent an accurate diagnosis. To this end, the following research question is investigated:

> *Is the diagnostic process of misconceptions in decimal fractions subject to confirmation biases?*

Based on the modelled judgment process and the theoretical assumptions, it is assumed that

① the generation of the initial hypothesis reveals if the (prospective) teachers recognise the ambiguous situation when diagnosing misconceptions in the area of decimal fractions based on one erroneously solved task and that

② the amount and type of processed information indicate possible confirmatory biases when gathering and identifying information. The processed information influences the accuracy of judgments when diagnosing misconceptions in the area of decimal fractions.

Method

To test these assumptions, prospective mathematics teachers (N = 79, average age = 21,7 years, 85% were female) at the beginning of their studies were confronted with one erroneously solved task by a virtual student in the domain of decimal fraction comparison (cf. figure 2). The first task and its incorrect solution represent an ambiguous diagnostic situation, as several misconceptions can be responsible for the student error. The diagnostic goal for the participants was to clearly determine the existing misconception of the presented student who consistently solves tasks according to a precise misconception. For this purpose, the necessary specific PCK concerning the misconceptions was visible to the participants and in total, 7 standardised cases had to be diagnosed.

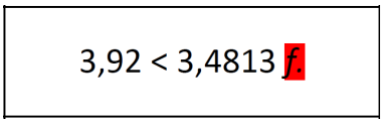

Figure 2: erroneously solved task in the domain of decimal fraction comparison

In an online survey, the prospective teachers first formulated an initial hypothesis about the misconception based on the erroneous student solution to one task. Afterwards they could choose further tasks which were solved by the student after their selection. That is, after clicking on it, teachers saw the solution to this task generated according to the misconception. Finally, the teachers submitted their final diagnosis.

The tasks offered for selection were shown in groups of four. The tasks differed with to the relevance of their information for the diagnosis. Tasks with relevant information for diagnosis allowed to distinguish between two possible misconceptions in the ambiguous diagnostic situation (i.e., these tasks are typically solved correctly with one misconception, but not with the other). Tasks with irrelevant information for diagnosis are either solved correctly by all learners despite the presence of a misconception or do not provide any additional information to the incorrectly solved tasks presented at the beginning.

A digital questionnaire was used to collect the initial hypothesis (single hypothesis or alternative hypotheses) and the number and type (diagnostically relevant or irrelevant) of information the participating persons used to get to their final diagnosis.

Results

Due to floor effects, two case diagnoses were excluded from the calculation. The judgment accuracy for the diagnosis of 5 cases over all participants was 63.3% (SD = 0.48) which designs the ratio of correct diagnoses of the presented students' misconception.

In a first step, the predictors for high judgment accuracy for the diagnosis of misconceptions in the area of decimal fractions are to be determined. Based on the theoretical modelling of the diagnostic process as hypothesis testing, the formulation of the initial hypothesis, the amount of information processed and the proportion of relevant diagnostic tasks are examined. Table 1 gives an overview of the average values of the three predictors.

Table 1: Representation of the average values of the assumed predictors for judgment accuracy according to accurate or incorrect diagnoses

	accurate diagnoses (n = 250)	incorrect diagnoses (n = 145)
average number of alternative initial hypotheses (SD)	0.26 (0.44)	0.15 (0.36)
average number of processed further tasks (SD)	2.73 (2.28)	2.94 (2.73)
average proportion of processed further relevant diagnostic tasks (SD)	0.68 (0.31)	0.54 (0.36)

The influence of the type of initial hypothesis (single hypothesis or alternative hypotheses) on the accuracy of the judgment was compared for accurate and incorrect diagnoses. The one-factor ANOVA indicates that significantly more accurate judgments are given after the formulation of an alternative hypotheses ($F(394) = 6.333$, $p = .012$, $d = 0.263$).

In addition, it was hypothesised that the amount of information processed, i.e. the number of tasks selected to see further solutions of the student, would also have an impact on judgment accuracy. To calculate this influence, the average number of processed tasks was compared for accurate and incorrect diagnoses. Group comparison using an ANOVA indicates no significant difference between accurate and incorrect judgments ($F(394) = 0.647$, $p = .422$, $d = 0.084$).

As a final predictor of judgment accuracy, we examined whether the type of information selected had an impact on judgment accuracy. For this purpose, the proportion of processed tasks that provide relevant diagnostic information was examined. The one-factor ANOVA indicates that accurate judgments, compared to incorrect judgments, are obtained by processing a significantly higher proportion of diagnostic tasks ($F(394) = 18.025$, $p \leq .001$, $d = 0.444$).

In a second step, the study aims to differentiate between different categories of information processing. Based on the finding that one of the predictors of high diagnostic accuracy is the formulation of an alternative hypothesis, different categories of judgment processes can be distinguished. First, after formulating an alternative hypothesis, a correct or incorrect diagnosis can be made. Secondly, a correct initial hypothesis can lead to correct or incorrect diagnoses. Thirdly, a correct or incorrect diagnosis can also be made after an incorrect initial hypothesis. Altogether, 6 different categories of judgment can be distinguished.

Out of the 6 possible judgment categories, two seems of special interest: On the one hand, the process that maps the ambiguity of the situation and subsequently leads to a correct result. Founded on the model assumptions, the ambiguous diagnostic situation should start with the formulation of alternative hypotheses and the given information (the tasks to select) must be processed as hints to correct diagnoses (normative judgment process). On the other hand, and taking into account the research interest, the process that starts from a single confirmatory misguided hypothesis and despite contrary information leads to an incorrect diagnosis (confirmatory-biased judgment process). Table 2 provides the descriptive overview on these two judgment categories concerning the number of processed information and the proportion of relevant diagnostic information.

Table 2: Descriptive overview of the information processing of the two judgment categories correct diagnoses based on alternative initial hypothesis and incorrect confirmatory diagnoses

	N	average number of processed information (SD)	average proportion of relevant diagnostic information (SD)
Normative judgment process: correct diagnoses based on alternative initial hypothesis	65	2.51 (2.02)	0.72 (0.28)
Confirmatory-biased judgment process: incorrect confirmatory diagnoses	84	2.48 (2.05)	0.47 (0.36)

The comparison using a one-factor ANOVA shows that the two judgment processes do not differ in the amount of information processed ($F(148) = 0.009$, $p = .926$, $d = 0.016$). Significant differences are shown, however, in the proportion of relevant diagnostic information processed and in the certainty of the final diagnosis: correct judgments based on an alternative hypothesis are made using a greater proportion of relevant information ($F(148) = 1.734$, $p \leq .001$, $d = 0.219$).

Discussion

Accurate diagnosis of student misconceptions by the teacher is crucial for building resilient conceptions in learners (Bradshaw & Templin, 2014). The research interest of the present study focuses on the emergence of such judgments via the investigation of the underlying cognitive processes (Loibl et al., 2020; Rieu et al., 2020). For such a complex judgment situation, it is assumed that biases take place and have a negative impact on judgment accuracy (Oswald & Grosjean, 2004).

The present study defines diagnostic judgments of misconceptions in decimal fractions as information processing in form of hypothesis testing and examines the impact of the created initial hypothesis, the number and type of processed information on the final diagnosis. It is expected that the information processing of the participating persons indicates the presence of confirmatory biased diagnoses.

In a first step, the formulation of alternative hypothesis and the processing of relevant information could be identified as predictors for the accuracy of judgments in the assessment of misconceptions in the area of decimal fractions. Due to the fact that all participating students had the necessary specific PCK about the misconceptions, these results show that the ambiguity of the diagnostic situation must be perceived and subsequently appropriate strategies must be used to cope with it. These findings complement previous studies on knowledge-guided information processing in diagnostic situations (Ostermann et al., 2018; Rieu et al., 2022) and person-dependent diagnostic sensitivity as a disposition (Kron, Sommerhoff, Achtner, & Ufer, 2021).

In a second step, categorising and contrasting the normative process with the confirmatory process highlights the differences between the two approaches concerning the type of information processed to obtain accurate judgments. These results indicate that an information-integrating strategy leads more often to a correct diagnosis (Böhmer, Hörstermann, Gräsel, Krolak-Schwerdt, & Glock, 2015; Fiske & Neuberg, 1990).

Despite several limitations, the results of the present study allow first insights into the judgment processes of prospective teachers when diagnosing misconceptions in the area of decimal fractions. The categorisation carried out on the basis of the type of initial hypothesis and further information processing permits an initial distinction between normative-accurate and confirmatory-biased judgment processes. The normative procedure, which processes relevant information based on alternative hypotheses in order to obtain an accurate diagnosis, should be incorporated into teacher training as a judgment strategy in complex situations to achieve a higher diagnostic accuracy and thus increase the adaptivity of teaching.

References

Böhmer, I., Hörstermann, T., Gräsel, C., Krolak-Schwerdt, S., & Glock, S. (2015). Eine Analyse der Informationssuche bei der Erstellung der Übergangsempfehlung: Welcher Urteilsregel folgen Lehrkräfte? *Journal for educational research online, 7*(2), 59–81. https://doi.org/10.25656/01:11490

Bradshaw, L., & Templin, J. (2014). Combining item response theory and diagnostic classification models: A psychometric model for scaling ability and diagnosing misconceptions. *Psychometrika*, *79*(3), 403–425. https://doi.org/10.1007/s11336-013-9350-4

Confrey, J., & Kazak, S. (2006). A thirty-year reflection on constructivism in mathematics education in PME. In A. Gutiérrez & P. Boero (Eds.), *Handbook of research on the psychology of mathematics education: Past, present and future* (pp. 305–345). Rotterdam: Sense Publishers.

Corno, L. (2008). On Teaching Adaptively. *Educational Psychologist*, *43*(3), 161–173. https://doi.org/10.1080/00461520802178466

Fiske, S. T., & Neuberg, S. L. (1990). *A Continuum of Impression Formation, from Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation* (Advances in Experimental Social Psychology): Elsevier. https://doi.org/10.1016/S0065-2601(08)60317-2

Fujii, T. (2014). Misconceptions and Alternative Conceptions in Mathematics Education. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education* (pp. 453–455). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-4978-8_114

Herppich, S., Praetorius, A.-K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., . . . Südkamp, A. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education*, *76*(4), 181–193. https://doi.org/10.1016/j.tate.2017.12.001

Kron, S., Sommerhoff, D., Achtner, M., & Ufer, S. (2021). Selecting Mathematical Tasks for Assessing Student's Understanding: Pre-Service Teachers' Sensitivity to and Adaptive Use of Diagnostic Task Potential in Simulated Diagnostic One-To-One Interviews. *Frontiers in Education*, *6*, 738. https://doi.org/10.3389/feduc.2021.604568

Loibl, K., Leuders, T., & Dörfler, T. (2020). A Framework for Explaining Teachers' Diagnostic Judgements by Cognitive Modeling (DiaCoM). *Teaching and Teacher Education*, *91*(3), 103059. https://doi.org/10.1016/j.tate.2020.103059

McKown, C., & Weinstein, R. S. (2003). The development and consequences of stereotype consciousness in middle childhood. *Child Development*, *74*(2), 498–515. https://doi.org/10.1111/1467-8624.7402012

Ostermann, A., Leuders, T., & Nückles, M. (2018). Improving the judgment of task difficulties: Prospective teachers' diagnostic competence in the area of functions and graphs. *Journal of Mathematics Teacher Education*, *21*(6), 579–605. https://doi.org/10.1007/s10857-017-9369-z

Oswald, M. E., & Grosjean, S. (2004). Confirmation Bias. In R. Pohl (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory* (1st ed., pp. 79–96). Hove, New York: Psychology Press. https://doi.org/10.4324/978020372061

Rieu, A., Loibl, K., Leuders, T., & Herppich, S. (2020). Diagnostisches Urteilen als informationsverarbeitender Prozess – Wie nutzen Lehrkräfte ihr Wissen bei der Identifizierung und Gewichtung von Anforderungen in Aufgaben? *Unterrichtswissenschaft*, *48*(4), 503–529. https://doi.org/10.1007/s42010-020-00071-x

Rieu, A., Leuders, T., & Loibl, K. (2022). Teachers' diagnostic judgments on tasks as information processing – The role of pedagogical content knowledge for task diagnosis. *Teaching and Teacher Education*, *111*, 103621. https://doi.org/10.1016/j.tate.2021.103621

Rubie-Davies, C., Hattie, J., & Hamilton, R. (2006). Expecting the best for students: Teacher expectations and academic outcomes. *The British Journal of Educational Psychology*, *76*(Pt 3), 429–444. https://doi.org/10.1348/000709905x53589

Schulz-Hardt, S., & Köhnken, G. (2000). Wie ein Verdacht sich selbst bestätigen kann: Konfirmatorisches Hypothesentesten als Ursache von Falschbeschuldigungen wegen sexuellen Kindesmißbrauchs. *Praxis der Rechtspsychologie*, *10* (Sonderheft 1), 60–88.

Stacey, K. (2005). Travelling the road to expertise: A longitudinal study of learning. In H. L. Chick & J. L. Vincent (Eds.), *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education* (pp. 19–36). Melbourne: PME.

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*(3), 743–762. https://doi.org/10.1037/a0027627

Trope, Y., & Liberman, A. (1996). Social hypothesis testing: cognitive and motivational mechanisms. In E. T. Higgins & A.W. Kruglanski (Eds.), *Social Psychology: Handbook of basic principles* (pp. 239–270). New York: Guilford Press. https://doi.org/10.1111/j.1467-9221.2008.00665.x

Westhoff, K., & Kluck, M.-L. (2014). *Psychologische Gutachten schreiben und beurteilen*. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-35354-3