DIMENSIONS OF PRE-SERVICE TEACHERS' DIAGNOSTIC JUDGEMENTS OF STUDENT SOLUTIONS

Christin Laschke[1], Lars Jenßen[2], Bettina Rösken-Winter[2]

[1]Leibniz-Institute for Science and Mathematics Education

[2]Humboldt-Universität zu Berlin

Judgments are part of teachers' daily practice and crucial for students' educational careers. Previous evidence indicated that judgments are informed by various criteria. But how pre-service teachers (PSTs) judge student solutions and how these judgments are structured are still open questions. In two studies we shed light on the construct. First, we investigated PSTs' judgements of an exemplary student solution regarding the applied categories (n_1=110). Based on the results, we then constructed items and investigated the structure of the construct by applying EFA and CFA (n_{2a}=168, n_{2b}= 209). The results revealed the following judgment dimensions: understanding, solution quality, presentation of procedure, and motivation. In addition to evidence on the structure of the construct, we gained an instrument to measure PSTs' judgments.

INTRODUCTION

When planning lessons and making daily decisions regarding instruction, teachers rely on their diagnostic judgment of students' knowledge and potential. Diagnostic judgment informs not only the assessment of students' performances, but also their grades and transition recommendations and is therefore crucial for students' academic development (Zhu et al., 2018) and their educational careers (Fischbach et al., 2013). Thus, teachers' diagnostic judgment plays an important role and must be given special attention during teacher education (Ready & Wright, 2011). Judging student solutions against the background of learning goals, such as gaining conceptual and procedural knowledge, is crucial in all school subjects. Especially in mathematics, teachers often struggle with judging the variety of student solutions as tasks allow for multiple solution pathways (Durking et al., 2017). During teacher education, emphasis is thus put on pre-service teachers' (PSTs) judgments with respect to identifying the potential in students' solutions. Up to now, some evidence on how PSTs notice students' mathematical thinking as a pre-requisite of their judgments (Crespo, 2000; Talanquer et al., 2015; Baldinger, 2020) exists. Also, Loibl et al. (2020) contributed a framework focussing on the cognitive processes underlying diagnostic judgments. So far, no studies examined what teacher diagnostic judgments of student solutions are actually based on and how they are structured. Particularly, we are interested in exploring whether a content-related perspective is taken or, rather, a generic viewpoint.

In our first study, we utilized an open response instrument to assess the variety of criteria PSTs used to judge an exemplary student solution and reconstructed judgment criteria by content analysis. In our second study, we developed items based on the

2022. In C. Fernández, S. Llinares, A. Gutiérrez, & N. Planas (Eds.). *Proceedings of the 45th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 115-122). PME.

aforementioned results that were assumed to represent the criteria. We then equipped the student solution with these rating scales and assessed two different groups of PSTs to examine the dimensional structure of PSTs' diagnostic judgments.

THEORETICAL FRAMEWORK

As indicated by social cognitive (dual process) models (Grawonski & Creighton, 2013, Loibl et al., 2020), judgments can arise from automatic and spontaneous or from controlled and reflected strategies of processing information. In many countries, educational standards postulate competencies that students should acquire and thus can serve as a normative framework against which teachers judge student solutions. For the learning of mathematics, gaining conceptual and procedural knowledge is important (Goldin, 2018). Students need to acquire procedural knowledge, thus knowledge about how procedures, algorithms, or methods are to be applied, as well as conceptual knowledge, in the sense of a content-related understanding of essential concepts and procedures and their interrelationships (Rittle-Johnson & Schneider, 2015). Thus, mathematics teachers are requested to assess students' products with regard to the extent to which procedures were applied appropriately and correctly to the tasks and whether conceptual knowledge has been acquired.

Previous research on PSTs' judgments of students' products revealed that PSTs use three strategies when judging students' products: mathematical reasoning, pedagogical (content) reasoning, and reasoning through self-comparison (Baldinger, 2020). Furthermore, judgments are often restricted to describing students' work instead of sense making of students' ideas (Talanquer et al., 2015), merely evaluating instead of interpreting, and not building inferences on students thinking (Crespo, 2000). Also, studies showed that students' errors resulting from a lack of conceptual understanding were interpreted by PSTs as lacking procedural understanding (Son, 2013). As a consequence, PSTs tended to directly respond to students' utterances or to correct their mistakes instead of asking questions to reveal their mathematical thinking (Cai et al., 2021). However, findings from intervention studies imply that learning opportunities can strengthen PSTs' judgments towards a more detailed investigation of students' thinking (Monson et al., 2018).

In sum, previous studies revealed that PSTs seem to focus on content-related aspects, but base their judgement on rather surface characteristics as describing students' solution instead of drawing on deep structure characteristics such as student understanding. That is, the evidence provides insights into the variety of judgment criteria and suggests a multidimensional structure of the construct. Against that background our study was guided by the following aims and research question.

AIMS AND RESEARCH QUESTION

To investigate the dimensions of PSTs' diagnostic judgments of student solutions we combined a qualitative and a quantitative approach. We first approached possible dimensions inductively (study 1). Based on these findings, we then constructed scales and checked the dimensionality of the construct (study 2). Particularly, we pursued the following research questions: RQ1: What judgment criteria can be detected from PSTs' diagnostic judgments of an exemplary student solution? RQ2: What dimensions structure PSTs' diagnostic judgments of an exemplary student solution?

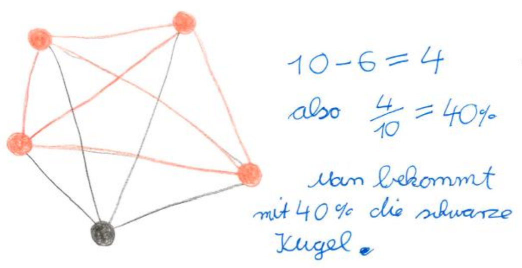
METHODOLOGY

To reveal the variety of judgment criteria, in study 1 we used an exemplary student solution of a probability task (see Figure 1) that allows for a diagnostic judgment with different focuses and by using different categories.

Task

There are 5 balls in a box. Four balls are red, one ball is black. Sarah pulls two balls out of the box with her eyes closed. What is the likelihood that Sarah will pull out the black ball?

Student solution



Translation of the solution:
10-6=4
thus 4/10=40%

You get the black ball with 40%.

Figure 1. Task and student solution.

The task was submitted to a sample of n_1 = 110 PSTs of a primary teacher education program, who attended a lecture on probability and stochastics.

The PSTs were asked to judge the result as well as the solution process and to justify their judgments. The open response data were analyzed by means of a step-wise inductive approach. First, the data-set was split into three subsets. One researcher analyzed one subset to identify a first set of categories that the PSTs used to rationalize their judgments. The research group then intensively discussed the categories. Second, code labels, definitions and examples were applied and revised through three rounds of coding and recoding, to identify the coding scheme that fits the data best (Kuckartz & Rädiker, 2019). Four categories of criteria were finally derived to code the whole dataset by two researchers. One sentence or more sentences with consistent meaning served as idea units and were coded to one category, if possible, or several categories, wherever necessary.

As a next step, items were constructed by extracting the most typical statements of each category. In study 2, we then combined the constructed items with a six-point response scale (from completely agree to completely disagree) to rate the student solution and submitted it to n_{2a} = 168 PSTs of two universities, enrolled in a master's program and who previously attended a lecture of probabilities and stochastics. To analyze the dimensional structure of the items, we first conducted an exploratory factor analysis (EFA) and, second, cross-checked the structure by applying confirmatory factor analyses (CFA). As factor loadings implied to add items, we constructed additional items based on the findings gained in study 1, and repeated the data analysis. The revised itemset was submitted to another sample of n_{2b} = 209 PSTs at the end of their bachelor teacher education program to validate the gained dimensionality. Again, CFA was carried out to analyze the structure of the construct. Quality of model fit was investigated by interpreting common fit indices (Hu & Bentler, 1999). Thereby, McDonalds ω was estimated as indicator for reliability (Hayes & Coutts, 2020).

RESULTS

The analysis of the open response data (RQ of study 1) revealed four categories PSTs used when judging the student solution. They pursued a focus on understanding, procedure, presentation, or motivation. PSTs with a focus on understanding usually emphasized that the student was able or not able to grasp the problem correctly (e.g. "student's solution shows that he or she understood the problem well"). PSTs who showed a focus on procedure pointed to details of how the student proceeded in either a correct or incorrect way or in a complete or incomplete way (e.g., "calculates correctly, converts to fractions, and gives correct percentages"). A focus on presentation was coded for judgments based on how the solution process was presented, arguing that the student created a picture of the problem, or wrote down a solution path and an answer or did not (e.g., "solution is not clearly arranged"). When PSTs recognized merely the student's effort to solve the problem, we coded it as focus on motivation (e.g., considering that he or she has a solution, and strained him- or herself). The four categories were thus coded regardless whether the PSTs pursued a deficit- or strength-based perspective.

In study 2, the EFA of the items constructed based on the most typical statements of each category indicated a three-dimensional model, that was proved by CFA against a four-dimensional structure (*presentation* and *procedure* modeled as two different factors in the second model, AIC = 5411,88, BIC = 5546,21, X^2 = 111.88, df = 47, p = .00; CFI = .89; RMSEA = .09 [.07 ; .11]; SRMR = .08). The results revealed a three-dimensional model as more appropriate than a four-dimensional model (AIC = 3812,44, BIC = 3909,28, X^2 = 37.25, df = 23, p = .03; CFI = .96; RMSEA = .06 [.02 ; .09]; SRMR = .06). Hence, *presentation* and *procedure* were building one factor which we labeled *presentation of procedure*, showing a high reliability (McDonalds ω = .80), in addition to the factors *understanding* (McDonalds ω = .78) and *motivation* (McDonalds ω = .79).

Each factor was presented by items with substantial loadings higher than .57. However, a closer look at item quality and loadings led to a revision of the dimension *understanding*. Two of the items of the dimension rather addressed the quality of the student's solution, e.g., "the solution is a smart one", with high loading on the factor.

Consequently, we added items based on the data gained by study 1 to test whether an additional dimension needs to be modelled.

Again, we conducted a CFA with an additional data set that revealed a four-dimensional model (see table 1) to fit the data best (see table 1). *Understanding* and *quality of solution* were building two different factors in addition to the factors *presentation of procedure* and *motivation* ($X^2 = 86.10$, $df = 48$, $p < .01$; CFI = .97; RMSEA = .06 [.04 ; .08]; SRMR = .05). The model fit was considered good. That is, the four dimensions each showed high reliability (McDonalds ω between .85 and .97). Also, they were represented by three items with a loading higher than .55.

The factors correlate moderately with each other (from r = .21, p < .01 to r = 59, p < .01), except for *understanding and quality of solution* with a high correlation (r = .82, p < .01). Nevertheless, the CFA confirmed a four-dimensional model as more appropriate than a three-dimensional model (ΔCFI = .14).

Dimensions	Items
Understanding $\omega = .97$	The student's solution shows that he or she understood the problem well. The student's solution indicates that he or she delved the problem. The student grasped the problem.
Quality of solution $\omega = .94$	The student carefully considered the solution. The student solution is smart. The student skilfully solved the problem.
Presentation of procedure $\omega = .85$	The student should have structured the solution better. The student should have chosen a different notation. The student solution does not show how he or she proceeded.
Motivation $\omega = .90$	The student tried hard to understand the task. The student strained to solve the task. The student gave thought to find a solution.

Table 1: Dimensions of PSTs' diagnostic judgment.

DISCUSSION AND CONCLUSION

In two studies, we shed light on the "black box" of PSTs' diagnostic judgments when confronted with an exemplary student solution – which will later be an important part of their daily practice. First, we could show that they pursued a focus on understanding, procedure, presentation or motivation. The results are in line with previous research, indicating the relevance of content-related aspects (Baldinger, 2020). However, some PSTs restricted their judgements to rather generic aspects when elaborating on how the solution was presented (Talanquer et al., 2015) or merely acknowledging motivational aspects such as the effort the student made. Thus, the diagnostic judgment criteria PSTs applied are of different quality with respect to fostering students' learning.

In the second study, we used the results of the qualitative study to further explore the dimensions of the construct. The results confirmed a multi-dimensional factor structure. Beyond the results of previous studies that identified content-related dimensions, our studies revealed that students' motivation as a generic dimension needs also to be considered as representing PSTs' judgments. Our results further revealed that the focus on procedure and on presenting the solution formed one dimension, in line with previous evidence on PSTs' judgments, showing that PSTs who focus on procedure take a rather descriptive than an interpretative view, not building inferences on students' thinking (Crespo, 2000). Furthermore, we discovered that the factor quality of solution needs to be considered in addition to the factor understanding. The factors presentation of procedure and motivation indicate a more surface view on student solutions as it was implicated by prior research (Talanquer et al., 2015). In contrast, the two content-related factors of understanding and quality of solution indicate a rather deep structure view, meeting the requirement to build inferences on students thinking (Rittle-Johnson & Schneider, 2015).

Our study on the one hand contributes to the field of teacher education by understanding the diagnostic judgment criteria PSTs use and how the construct is structured. On the other hand, we gained a standardized instrument to measure diagnostic judgment criteria PSTs apply when they judge an exemplary student solution. So far, we could conduct an additional study to test whether the identified diagnostic judgment dimensions fit the judgment of a student solving an arithmetic solution, proving the independency of the dimensions from the concrete task used. As a next step, starting from the study of Monson et al. (2020) who could show that PST learning opportunities can contribute to a stronger focus on students' thinking, we will apply the instrument to examine whether and what learning opportunities can affect a shift from focusing on surface to deep structure and content-related characteristics.

References

Baldinger, E. E. (2020). Reasoning about student written work through self-comparison: How pre-service secondary teachers use their own solutions to analyze student work. *Mathematical Thinking and Learning, 22*(1), 56-78. https://doi.org/10.1080/10986065.2019.1624930

Cai, J., LaRochelle, R., Hwang, S., & Kaiser, G. (2021). Expert and preservice secondary teachers' competencies for noticing student thinking about modelling. *Educational Studies in Mathematics.* https://doi.org/10.1007/s10649-021-10071-y

Crespo S. (2000). Seeing more than right and wrong answers: Prospective teachers' interpretations of students' mathematical work. *Journal of Mathematics Teacher Education,* 155–181.

Durkin, K., Star, J.R. & Rittle-Johnson, B. (2017). Using comparison of multiple strategies in the mathematics classroom: lessons learned and next steps. *ZDM Mathematics Education 49, 585–597.* https://doi.org/10.1007/s11858-017-0853-9

Fischbach, A., Baudson, T. G., Preckel, F., Martin, R., & Brunner, M. (2013). Do teacher judgments of student intelligence predict life outcomes? *Learning and Individual Differences, 27,* 109–119. doi:10.1016/j.lindif.2013.07.004.

Goldin, G. A. (2018). Discrete mathematics and the affective dimension of mathematical learning and engagement. In E. W. Hart & J. Sandefur (Eds.), *Teaching and learning discrete mathematics worldwide: curriculum and research* (pp. 53–65). Springer International Publishing.

Gawronski, B., & Creighton, L. A. (2013). Dual-process theories. In D. E. Carlston (Ed.), The *Oxford handbook of social cognition* (pp. 282–312). Oxford University Press.

Hayes, A. F. & Coutts, J. J. (2020). Use omega rather than Cronbach´s Alpha for estimating reliability. But… C*ommunication Methods and Measures, 14*(1), 1-24. https://doi.org/10.1080/19312458.2020.1718629

Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Kuckartz, U. & Rädiker, S. (2019). Analyzing Qualitative Data with MAXQDA. Springer International Publishing.

Loibl, K., Leuders, T., & Dörfler, T. (2020). A framework for explaining teachers' Diagnostic judgements by cognitive modeling (DiacoM). *Teaching and Teacher Education*, *91*, 1-10. https://doi.org/10.1016/j.tate.2020.103059

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276e282. https://doi.org/10.11613/BM.2012.031

Monson, D., Krupa, E., Lesseig, K., & Casey, S. (2020). Developing secondary prospective teachers' ability to respond to student work. *Journal of Mathematics Teacher Education, 23*(2), 209–232. https://doi.org/10.1007/s10857-018-9420-8

Ready, D. D., & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities. The role of child background and classroom context. *American Educational Research Journal*, *48,* 335–360. doi:10.3102/0002831210374874

Rittle-Johnson, B., & Schneider, M. (2015). Developing conceptual and procedural knowledge of mathematics. In R. C. Kadosh & A. Dowker (Eds.), Oxford library of psychology. *The Oxford handbook of numerical cognition* (pp. 1118–1134). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642342.013.014

Talanquer, V., Bolger, M., & Tomanek, D. (2015). Exploring prospective teachers' assessment practices: Noticing and interpreting student understanding in the assessment of written work. *Journal of Research in Science Teaching, 52*(5), 585–609. https://doi.org/10.1002/tea.21209

Zhu, M., Urhahne, D., & Rubie-Davies, C. M. (2018). The longitudinal effects of teacher judgement and different teacher treatment on students' academic outcomes. *Educational Psychology, 3410*, 1–21. https://doi.org/10.1080/01443410.2017.1412399