

Chapter 2

Science-P I: Modeling Conceptual Understanding in Primary School

Judith Pollmeier, Steffen Tröbst, Ilonca Hardy, Kornelia Möller,
Thilo Kleickmann, Astrid Jurecka, and Knut Schwippert

Abstract In the Science-P project (*Science Competency in Primary School*), we aimed at modeling scientific literacy in two dimensions—scientific reasoning and conceptual understanding—to describe science learning in primary school. The present chapter focuses on conceptual understanding exemplified by two content areas: floating and sinking (FS) and evaporation and condensation (EC). Drawing on results from conceptual change research in developmental psychology and science education, we devised a model with three hierarchically ordered levels of understanding—naïve, intermediate and scientifically advanced—as the foundation of item and test construction. The two content areas engendered a two-dimensional structure in our test instrument. A validation study underscored that responses to our paper-pencil items were systematically related to responses obtained in interviews. Our test instrument was used to capture the development of primary school students' conceptual understanding from second to fourth grade, in both a cross-sectional and a longitudinal study. For cross-sectional data, students' proficiency in scientific reasoning was found to predict their conceptual understanding. In future analyses, we will test this finding with our longitudinal data.

Keywords Conceptual understanding • Science competency • Primary school • Development

J. Pollmeier (✉) • S. Tröbst • T. Kleickmann
Kiel University, Kiel, Germany
e-mail: pollmeier@paedagogik.uni-kiel.de; troebst@paedagogik.uni-kiel.de;
kleickmann@paedagogik.uni-kiel.de

I. Hardy • A. Jurecka
Goethe University Frankfurt, Frankfurt/Main, Germany
e-mail: Hardy@em.uni-frankfurt.de; Jurecka@em.uni-frankfurt.de

K. Möller
University of Münster, Münster, Germany
e-mail: molleko@uni-muenster.de

K. Schwippert
University of Hamburg, Hamburg, Germany
e-mail: knut.schwippert@uni-hamburg.de

2.1 The Assessment of Science Competency in Primary School

In recent years, science learning has been described within the construct of scientific literacy, which has been conceptualized with various facets, distinguishing a component of conceptual understanding from a component of procedural understanding (Bybee 1997). While there are theoretical and empirically validated models of science competency for secondary schools, corresponding efforts are rare within the growing body of research on competency development for primary schools (e.g., Walpuski et al. 2011). Hence, in our project we aimed to model the development of science competency in primary school in the two dimensions of scientific reasoning (e.g., Koerber et al. 2017, in this volume) and conceptual understanding. The latter is the focus of this chapter.

To derive a theoretically plausible and empirically testable competency model of conceptual understanding in primary school science, it appears suitable to resort to the findings of conceptual change research in developmental psychology and science education. This research has revealed that students bring a wide range of individual content-specific ideas and conceptions to the science class; these have potential to hinder or foster formal science learning. Aside from the nature of students' naïve conceptions, conceptual change research has also explored the pathways along which these evolve (e.g., Schneider and Hardy 2013). In this context, we pursued three main goals: (a) modeling primary school students' conceptual understandings in the content areas of FS and EC with paper-pencil tests to empirically validate a competency model using large groups of students, (b) investigating the development of conceptual understanding over the course of primary school and (c) examining the relation between students' conceptual understanding and scientific reasoning.

2.2 Modeling Conceptual Understanding in Primary School Science

2.2.1 Model Specification and Item Construction

In the first place, we hypothesized a competency model with three hierarchical levels of increasing understanding: At the *naïve* level students hold scientifically inadequate conceptions which, through processes of restructuring or enrichment may possibly result in *intermediate* conceptions. These contain partly correct conceptualizations and are applicable in a broader range of situations than are naïve conceptions. At the *scientifically advanced* level, eventually, students hold conceptions in accordance with scientifically accepted views (Hardy et al. 2010). Within this framework, we designed a construct map as the foundation for item development (Wilson 2005). For each content area, this contained detailed descriptions of

possible student conceptions at each level of understanding. These conceptions were extracted from conceptual change research (e.g., Hsin and Wu 2011; Tytler 2000).

The translation of the conceptions identified in conceptual change research into a paper-pencil instrument suitable for testing groups of primary school students, posed a considerable challenge for item development. Specifically, the test instrument had to incorporate and represent different levels of conceptual understanding without inducing artificial response tendencies and preferences. Using the construct map, we designed items with mainly closed response formats; response alternatives represented varying levels of conceptual understanding. Response formats were either forced-choice (select the better of two alternatives), multiple-choice (select the best of three to six alternatives) or multiple-select (judge three to six alternatives consecutively as true or false). In addition, a few items with open and graphical response formats were constructed (Kleickmann et al. 2010).


For all items, the stems consisted of descriptions of physical phenomena relevant to the two content areas. Of these phenomena, those which could be presented in a classroom, were demonstrated during administration of the test (see Fig. 2.1). After presentation of a specific phenomenon, students had to select or, in the rare case of open response formats, produce an explanation for that phenomenon. For multiple-select items, students could select several explanations simultaneously (see Fig. 2.1). To minimize the impact of reading ability on students' performance, descriptions of phenomena and response alternatives were read out aloud. Students in participating classes proceeded simultaneously through the test within 90 min. The majority of items represented explanations on the naïve level, due to the wealth of naïve conceptions identified by previous research. In general, primary school students were considered to demonstrate proficient conceptual understanding by the dismissal of naïve explanations and the endorsement of intermediate or scientifically advanced explanations.

2.2.2 *Conceptual Understanding: Dimensions and Levels*

To examine the dimensionality of our test instrument, we fitted one-parametric logistic item response models with varying dimensionality to the data of a cross-sectional study with 1820 s, third and fourth graders, using ACER Conquest 2.0 (Wu et al. 2005). A likelihood ratio test of relative model fit demonstrated that a model featuring the two content areas as separate dimensions, fitted the data better than did a uni-dimensional model ($\Delta\chi^2(2) = 246.83$, $p < .001$, $\Delta\text{AIC} = 242.88$, $\Delta\text{BIC} = 231.71$; Pollmeier 2015; Pollmeier et al. in prep.). This finding supported the notion that competency in certain content areas might develop separately from that in other content domains. Thus, further analyses for the cross-sectional data were performed separately for each content area. The two-dimensionality established for the cross-sectional data set was consistent with the results of preliminary studies (Pollmeier et al. 2011).

The Cold Glass

Instruction:
 You fill a glass with cold water and ice cubes. At first the glass is dry on the outside. But after a couple of minutes you can see little droplets on the outside.
 Let's try this by ourselves.
 (*Demonstration of phenomenon with corresponding material.*)



Why do the droplets appear on the outside of the glass?

Select either “true” or “false” after each explanation!

| | true | false |
|--|---------------------|-------|
| The water droplets came from inside the glass through small pores. | <i>naïve</i> | |
| The water droplets condensed in the air, as the air cooled. | <i>scientific</i> | |
| Water from the air became visible because of the cold. | <i>intermediate</i> | |
| The water from the glass is now on the outside. | <i>naïve</i> | |

Fig. 2.1 Sample item: condensation

To clarify the influence of the hypothesized levels of understanding and sub-facets of content areas defined in the construct map on students' performance, we devised explanatory item response models, using the R-package lme4 (De Boeck and Wilson 2004; De Boeck et al. 2011). These models explored the impact of specific person and item characteristics on students' responses (Pollmeier et al. 2013). In particular, the analyses revealed differential proficiency in subgroups of students with regard to levels of understanding in the two content areas. We found an overall gender effect for the content area of FS, with boys outperforming girls. Furthermore, girls exhibited specific weaknesses for items on density and displacement, compared

to items on buoyancy. They also, relative to boys, had a preference for explanations on the intermediate level, whereas they neglected explanations based on scientifically advanced conceptions. In contrast, for the content area of EC, overall performance did not differ between girls and boys. Yet girls also displayed a relative preference for items featuring intermediate conceptions in the content area of EC, although they did not neglect scientifically advanced conceptions (Pollmeier et al. 2013).

To explain additional variance in item difficulties, we explored the relevance of characteristics derived from the classification of item stems. Again, we employed explanatory item response models, both for the cross-sectional data and for one further experimental study, in which certain contextual features were varied systematically across item stems (Pollmeier 2015). For the content area of FS we identified *congruence* as an important explanatory characteristic for items associated with the concepts of density and displacement: With these items, students had to compare two objects and decide which was made of the denser material or displaced more water. Items featuring congruent objects—that is, the denser object was also the heavier object, or the heavier object displaced more water—were easier to solve than incongruent items—that is, where the denser object was the lighter object or the lighter object displaced more water.

For the content area of EC we obtained no single explanatory characteristic of central importance. However, we found that the specific content used in items accounted for a large portion of the variance in item difficulties. The most difficult content for the facet of evaporation was a naïve conception: the anthropomorphic interpretation of the physical phenomena to be explained. Items with scientifically advanced content—that is, with the correct explanation for phenomena of evaporation in age-appropriate language—were not as difficult to solve as these items, or items proposing a mere change of matter as the explanation for evaporation phenomena. Items conveying a false description of the change of matter, a description of change of location, and non-conservation of matter as explanations for evaporation phenomena, were comparatively easy to solve. For the facet of condensation, items featuring a faulty description of a cause, a change of location, a change of matter and a scientifically advanced explanation for condensation phenomena, constituted an order of decreasing difficulty of content (Pollmeier 2015).

2.2.3 Validity

To assess the convergent and discriminant validity of our instrument, we conducted a validation study with four third grade classes (FS: $N = 41$, EC: $N = 32$). For each content area we presented two classes with 13 item stems, both as paper-pencil items with closed response format and as interview items with open response format (Pollmeier et al. 2011). Students were randomly assigned to an order of presentation of the two forms of item. Additionally, reading ability (Lenhard and Schneider 2006) and cognitive ability (Weiß 2006) were measured. We found substantial

correlations between the two modes of assessment for each content area, but also systematic differences in the responses: Students produced a wider range of answers on the naïve and intermediate levels, and fewer answers on the scientifically advanced level for interview items than for paper-pencil items. As expected, the production of explanations was more demanding than merely endorsing correct response alternatives. Apart from that, knowledge in the content area of FS, as assessed with the interviews, appeared to be more fragmented and context dependent than corresponding knowledge in the content area of EC; a discrepancy not evident in the paper-pencil items.

Moreover, for the content area of EC, performance on paper-pencil items was independent of reading ability and cognitive ability. This finding supports the claim that our instrument measured a form of science competency that was more specific than those general abilities. The substantial relation found between the test of cognitive ability and performance on the paper-pencil items for the content area of FS probably was induced by the similarity between items covering the facet of density and items assessing cognitive ability. The impact of socio-economic status on proficiency in the content of FS was evident both for interview and for paper-pencil items.

In sum, there was a systematic difference between responses to interview and paper-pencil items that can be readily explained by the discrepancy between free retrieval and recognition and that thus was not caused by a difference in the constructs assessed by the items. In other words, the positive associations between responses to interview and paper-pencil items indicate that our test instrument for assessment of conceptual understanding captured a form of science competency that is plausibly parallel to the conceptual understanding found in classic conceptual change research.

2.3 The Development of Conceptual Understanding in Primary School Science

Analyses of the cross-sectional data set (see Sect. 2.2.2 above) by means of explanatory item response models also yielded insights into the differences in average conceptual understanding between grade levels: Third and fourth graders outperformed students from second grade in terms of conceptual understanding, and we further unveiled the specific strengths of third and fourth graders. Within the content area of FS, third and fourth graders performed particularly well on items covering the facets of density and displacement and on items featuring scientifically advanced conceptions. In the content area of EC, students from third and fourth grade displayed a specific strength in items concerned with the facet of evaporation.

A longitudinal study with a total of 1578 students in 75 classes from primary schools in two federal states of Germany (Baden-Wuerttemberg, North Rhine-Westphalia) concluded our project. Students completed our tests on science

Table 2.1 Descriptive statistics for the longitudinal study

| Content area | Grade 3, end of school year | | | Grade 4, end of school year | | |
|-------------------------------|-----------------------------|-----|-----|-----------------------------|-----|-----|
| | M(SD) | Min | Max | M(SD) | Min | Max |
| Descriptive item statistics | | | | | | |
| Floating and sinking | .50(.43) | .16 | .96 | .56(.45) | .30 | .90 |
| Evaporation and condensation | .52(.47) | .30 | .86 | .59(.47) | .33 | .92 |
| Descriptive person statistics | | | | | | |
| Floating and sinking | 11.40(4.48) | 2 | 23 | 14.39(4.87) | 0 | 23 |
| Evaporation and condensation | 24.63(5.53) | 10 | 42 | 27.92(6.30) | 2 | 46 |

competency at the end of third and fourth grade. For the preliminary analyses we used 48 individual items from 23 anchoring item stems, of the total 27 stems that were used in this study.

For both content areas, on average, the solution rates of anchor items increased in fourth grade and were accompanied by relatively large standard deviations (see Table 2.1, item statistics); a first hint that our instrument covered a sensible amount of variety. Also, the number of correctly solved items reveals that students on average solved more items at posttest than at pretest (see Table 2.1, person statistics). In relation to the number of anchor items assigned to each content area, this implies a relatively smaller gain in conceptual understanding for the content area of EC.

In sum, our preliminary explorations suggest that we succeeded in assessing naturally occurring growth in conceptual understanding in this longitudinal study. In future analyses based on all items, we will examine whether the small growth in the content area of EC is attributable to the general difficulty of this content or rather to deficiencies in the amount and quality of formal instruction. Furthermore, we will investigate students' performance with regard to the various characteristics of items and item stems (e.g., the assigned level of conceptual understanding). Finally, future analyses will focus on investigating the conjoint development of conceptual understanding and scientific reasoning.

2.4 Conceptual Understanding and Scientific Reasoning

The issue of the relation between conceptual understanding and scientific reasoning was also tackled with the cross-sectional study data (for detailed analyses of primary school students' competency in scientific reasoning see Koerber et al. 2017, in this volume). After calibrating our tests by the use of simple Rasch models, we retrieved weighted likelihood estimates of person ability for subsequent analyses. Multilevel analyses revealed substantial associations between scientific reasoning and conceptual understanding in both content areas that were not readily explained by relevant covariates like fluid intelligence, reading ability, interest in science, socioeconomic status, and immigrant status. Furthermore, in the content area FS, the predictive effect of scientific reasoning on conceptual knowledge slightly

increased with grade, even after controlling for fluid ability. These findings hint at the possibility that proficient scientific reasoning facilitates the acquisition of conceptual understanding. Specifically, having a command of the processes of scientific reasoning could enhance the evaluation of evidence with respect to existing conceptions, which could take the role of hypotheses or even theories. This could also account for the possibly cumulative effect of proficient scientific reasoning on conceptual understanding, suggested by its interaction with the content area FS in the cross-sectional data. Future analyses of the longitudinal data will yield deeper insights into this issue.

Acknowledgements The preparation of this chapter was supported by grants to Kornelia Möller (MO 942/4-1, MO 942/4-2 and MO 942/4-3), from the German Research Foundation (DFG) in the Priority Programme “Competence Models for Assessing Individual Learning Outcomes and Evaluating Educational Processes” (SPP 1293).

References

- Bybee, R. W. (1997). Toward an understanding of scientific literacy. In W. Gräber & C. Bolte (Eds.), *Scientific literacy* (pp. 37–68). Kiel: IPN.
- De Boeck, P., & Wilson, M. (Eds.). (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York: Springer.
- De Boeck, P., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., Tuerlinckx, F., & Partchev, I. (2011). The estimation of item response models with the lmer function from the lme4 package in R. *Journal of Statistical Software*, 39(12), 1–28.
- Hardy, I., Kleickmann, T., Koerber, S., Mayer, D., Möller, K., Pollmeier, J., ... Sodian, B. (2010). *Die Modellierung naturwissenschaftlicher Kompetenz im Grundschulalter: Projekt Science-P* [Modeling science competence in primary school: The project Science-P.]. *Zeitschrift für Pädagogik, Beiheft* 56, 115–125.
- Hsin, C., & Wu, H.-K. (2011). Using scaffolding strategies to promote young children’s scientific understandings of floating and sinking. *Journal of Science Education and Technology*, 20, 656–666. doi:10.1007/s10956-011-9310-7.
- Kleickmann, T., Hardy, I., Möller, K., Pollmeier, J., Tröbst, S., & Beinbrech, C. (2010). Die Modellierung naturwissenschaftlicher Kompetenz im Grundschulalter: Theoretische Konzeption und Testkonstruktion [Modeling science competence in primary school: Theoretical conception and test construction]. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 265–283.
- Koerber, S., Sodian, B., Osterhaus, C., Mayer, D., Kropf, N., & Schwippert, K. (2017). Science-P II: Modeling scientific reasoning in primary school. In D. Leutner, J. Fleischer, J. Grünkorn, E. Klieme, *Competence assessment in education: Research, models and instruments* (pp. 19–29). Berlin: Springer.
- Lenhard, W., & Schneider, W. (Eds.). (2006). *ELFE 1–6. Ein Leseverständnistest für Erst- bis Sechstklässler* [A reading literacy test for Grades 1–6]. Göttingen: Hogrefe.
- Pollmeier, J. (2015). *Kontextmerkmale und die Bearbeitung von Aufgaben in einem Test naturwissenschaftlicher Kompetenz in der Grundschule* [Contextual features of items in a test of science competence in primary school] (Unpublished doctoral dissertation). WWU Münster, Münster, Germany.

- Pollmeier, J., Hardy, I., Koerber, S., & Möller, K. (2011). Lassen sich naturwissenschaftliche Lernstände im Grundschulalter mit schriftlichen Aufgaben valide erfassen [Valid assessment of scientific knowledge in primary school with written tests]? *Zeitschrift für Pädagogik*, *57*, 834–853.
- Pollmeier, J., Tröbst, S., & Möller, K. (2013, August). *Conceptual competence in science in primary school*. Paper presented at the 15th Biennial EARLI Conference, München, Germany.
- Pollmeier, J., Tröbst, S., Hardy, I., Osterhaus, C., Koerber, S., Mayer, D., ... Sodian, B. (in preparation). *The effect of scientific reasoning on conceptual scientific understanding in the course of primary school (Grade 2–4)*.
- Schneider, M., & Hardy, I. (2013). Profiles of inconsistent knowledge in children's pathways of conceptual change. *Developmental Psychology*, *49*, 1639–1649. doi:[10.1037/a0030976](https://doi.org/10.1037/a0030976).
- Tytler, R. (2000). A comparison of year 1 and year 6 students' conceptions of evaporation and condensation: Dimensions of conceptual progression. *International Journal of Science Education*, *22*, 447–467. doi:[10.1080/095006900289723](https://doi.org/10.1080/095006900289723).
- Walpuski, M., Ropohl, M., & Sumfleth, E. (2011). Students' knowledge about chemical reactions development and analysis of standard-based test items. *Chemistry Education Research and Practice*, *12*, 174–183. doi:[10.1039/C1RP90022F](https://doi.org/10.1039/C1RP90022F).
- Weiß, R. H. (2006). *Grundintelligenztest Skala 2–Revision (CFT 20-R)* [Test for general mental capacity]. Göttingen: Hogrefe.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah: Erlbaum.
- Wu, M., Adams, R., & Wilson, M. (2005). *ACER ConQuest*. Camberwell: ACER Press.