– Originalbeitrag –

Errors Made by 5th-, 6th-, and 9th-Graders When Planning and Performing Experiments: Results of Video-Based Comparisons

Armin Baur

University of Education, Heidelberg
Biology and Biology Education

ABSTRACT

Some previously-published research deals with pupils' errors in experimentation (manual mistakes, methodological errors, misconceptions and pupil-specific approaches). However, many of these studies work neither with pupils who experiment themselves nor with individually working pupils. The article presents the observed pupils' errors of 9th ($n_1 = 12$), 5th and 6th graders ($n_2 = 18$), recorded during individual, open experimentation using videography (60 minutes per experiment) and subsequent analysis of observation protocols based on the video material and pupils' experimentation report. In the inquiry, each pupil carried out two different experiments. The results replicate many errors described in the literature and also reveal new ones. When comparing 5th and 6th graders with 9th graders, very few differences could be observed in the pupils' errors. The analysis of the strategies (pupils' course of action) shows that five strategies (four strategies, one of which consist of two sub-strategies) could be assigned to all pupils, which are categorised as follows: 'Initially, only one experimental trial with all variables', 'Control variables are varied in the experimental trials', 'Only the assumed variable(s) and the necessary variable(s) are included in the experimental trials', 'Working in engineer mode', 'Procedure without hypothesis'. Not a single pupil used the control of variables strategy correctly. The calculation of correlations between the occurrence of errors in the two different contents shows few significant correlations, except in hypothesis formulation, indicating that the occurrence of pupils' errors is content dependent.

Key words: experimentation; pupils' errors; scientific inquiry

ZUSAMMENFASSUNG

Einige Forschungsarbeiten befassten sich mit Schüler*innenfehlern beim Experimentieren (manuelle Fehler, methodische Fehler, Präkonzepte und schülerspezifisches Vorgehen). Viele dieser Studien arbeiteten jedoch weder mit Schülern*innen, die selbst experimentieren (hands-on), noch mit individuell arbeitenden Schülern*innen. Im Artikel werden beobachtete Schüler*innenfehler von Neunt- ($n_1 = 12$), Fünft- und Sechstklässler*innen ($n_2 = 18$) dargestellt. Diese wurden beim offenen Experimentieren in Einzelarbeit über Videographie (60 Minuten Film pro Experiment) und anschließender Analyse der aus den Videos erstellten Beobachtungsprotokolle und der Experimentierprotokolle der Schüler*innen identifiziert. In der Untersuchung führte jede Schülerin/jeder Schüler zwei verschiedene Experimente durch. Die Ergebnisse replizieren viele Schüler*innenfehler, die in der Literatur beschrieben werden, und zeigen neue Fehler auf. Beim Vergleich von Fünft- und Sechstklässler*innen mit Neuntklässler*innen waren nur sehr wenige Unterschiede in den auftretenden Schüler*innenfehlern zu beobachten. Die Analyse der Strategien (Vorgehensweisen der Schüler*innen) zeigt, dass alle Schüler*innen fünf Strategien (vier Strategien, eine davon mit zwei Teilstrategien) zugeordnet werden können: "Zuerst nur ein Ansatz mit allen Variablen", "In den Versuchsansätzen werden Kontrollvariablen variiert", "Aufbau besteht nur aus notwendiger und unabhängiger Variable", "Arbeiten im Ingenieurmodus", "Vorgehen ohne Hypothese". Kein/e einzige/r Schülerin/Schüler hat die Variablenkontrollstrategie richtig angewandt. Die Berechnung von Korrelationen zwischen dem Auftreten von Fehlern in den beiden verschiedenen Inhalten ergab nur wenige signifikante Korrelationen in der Hypothesenformulierung der Schüler*innen, was darauf hindeutet, dass das Auftreten von Schüler*innenfehlern vom Inhalt abhängig ist.

Schlüsselwörter: Experimentieren; Schüler*innenfehler; Scientific inquiry

1 Background and Objective

Knowledge of the methods of scientific inquiry and the respective competencies necessary to answer scientific questions using scientific methods are considered relevant parts of scientific literacy (e.g. OECD, 2017; Hodson, 2014; Bybee, 1997). For this reason, the educational standards of many countries include competencies (or practices) for scientific inquiry (e.g. DfE, 2014; NRC, 2012; KMK, 2005). However, it should be kept in mind for teaching that the methods of scientific inquiry are not learned effectively when pupils only take part in the performance of an inquiry method (Bell, Blair, Crawford & Lederman, 2003). Studies have shown that scientific inquiry is not learned best through forms of open learning (Kirschner, Sweller & Clark, 2006). This suggests that competencies (or practices) of scientific inquiry must be explicitly taught and learned at school. Therefore, natural science didactics must provide impulses, as well as educational and methodological teaching support in order to enable effective teaching, culminating in inquiry skills and enabling pupils to perform guided and in the long term, open inquiry.
Paying attention to pupils' errors could be helpful for the development of such impulses. *Pupils' errors* include manual mistakes, methodological errors, misconceptions, and pupil-specific approaches (see Baur, 2018). The latter refers to the fact that pupils occasionally take steps that, while not necessarily false from a scientific point of view, do not lead towards a solution in the scientific problem-solving process. Knowledge about pupils' common errors is useful for diagnostics (Baur, 2015), feedback (learning from mistakes: Weinert, 1999; Oser, Hascher & Spychinger 1999), and lesson planning (Schumacher, 2008). It may also be very useful for the development of scaffolding for independent, problem-oriented scientific inquiry. Errors are therefore not perceived as a deficit, but as a learning opportunity (Metcalfe, 2017; Schumacher, 2008). Experimentation – the focus of this article – is one facet of scientific inquiry (NRC, 1996).

2 Previous Results of Research

Research on pupils' errors in experimenting has been done in psychology and various domains of science education (biology, chemistry, physics, and earth & space science).
The results show a multitude of pupils' errors in the different phases of experimentation. The following list contains the most common errors made within each phase:
<u>Formulate a research question:</u> Instead of causal questions which can be tested with experiments, pupils often formulate content-related questions (Neber & Anton, 2008 [C][1]; Hofstein, Navon, Kipnis & Mamlok-Naaman, 2005 [C]; Cuccio-Schirripa & Steiner, 2000 [B, Ph]).
<u>State a hypothesis:</u> Many learners do not formulate hypotheses when experimenting (Zhai, Jocz & Tan, 2013 [B, C, Ph]; Millar & Lubben, 1996 [Ph]; Dunbar & Klahr, 1989 [P]). From the pupils' perspectives, the purpose of an experiment is to achieve an effect (Hammann, Phan, Ehmer & Bayrhuber, 2006 [B]; Schauble, Klopfer & Raghavan, 1991 [Ph]). Often, pupils assume only a positive covariation between the independent and dependent variables (Kanari & Millar, 2004 [Ph]). Pupils only establish hypotheses when they appear plausible (Hammann, 2006 [B]; Klahr; Fay & Dunbar, 1993 [E]).
<u>Design and experiment:</u> Many learners plan experiments that include only one approach (one experimental trial) and cannot, therefore, compare the influence of the independent variable (Hammann, Phan, Ehmer & Grimm, 2008 [B]; Hammann et al., 2006 [B]). While some pupils do not use a control trial (Germann, Aram & Burke, 1996 [Ph]), others do not vary the independent variable (Chen & Klahr, 1999 [Ph]). The control of variables strategy is also often disregarded (Siler & Klahr, 2012 [E, Ph]; Hammann et al. 2006 [B]; Kuhn & Dean, 2005 [E]; Schauble et al. 1991 [Ph]). In many cases, pupils simply try things out without being strictly scientific (Meier & Mayer 2012 [B]; Wahser & Sumfleth 2008 [C]; Hammann et al., 2008 [B]). As a matter of fact, pupils rarely repeat their measurements (Lubben & Millar, 1996 [B, C, Ph]) and some are not able to use (simple) measuring instruments and laboratory equipment correctly (Kechel, 2016 [Ph]).

[1] The source references in this section are assigned with a letter to their domains: [B]: Biology; [C]: Chemistry; [E]: Earth & Space Science, [Ph]: Physics; [P]: Psychology.

doi: 10.11576/zdb-3576

<u>Observe and analyse data:</u> If the data determined in an experiment do not match the expected data, pupils suspect an error in the performance of their experiment (Ludwig, Priemer & Lewalter, 2019 [Ph]; Wahser & Sumfleth, 2008 [C]; Chinn & Brewer, 1998 [P]; Carey, Evants, Honda, Jay & Unger, 1989 [B]). Many pupils ignore data that do not match their imagination (Chinn & Brewer, 1993 [P]; Schauble et al., 1991 [Ph]; Watson & Konicek, 1990 [Ph]; Kuhn, 1989 [P]; Gauld, 1986 [Ph]). Pupils as well as adults tend to maintain hypotheses and try to confirm them (Chinn & Brewer, 1993 [P]; Klayman & Ha, 1989 [P]; Wason, 1960 [P]). For repeated measurements, pupils often choose the first, last or some value between the highest and lowest measurement and do not calculate the arithmetic mean (Kanari & Millar, 2004 [Ph]; Masnick & Klahr, 2003 [Ph]; Lubben & Millar, 1996 [B, C, Ph]). Some pupils confuse the result (observation, measurement) of an experiment with its conclusion (Boaventura, Faria, Chagas & Galvão, 2013 [B]). There is a tendency for pupils to change too many variables, making it difficult for them to draw a conclusion (Glaser, Schauble, Raghavan & Zeitz, 1992 [P]).

A comparison of the methods used in the presented studies reveals that data was collected with paper-pencil tests, interviews, computer-based methods, hands-on tests with small groups, or hands-on tests with pupils working individually. Analyses of assessment formats show that the results of hands-on tests differ from those of other formats (paper-pencil tests, computer-based testing) in terms of convergence (Schreiber, 2012; Emden, 2011; Shavelson, Ruiz-Primo & Wiley, 1999; Ruiz-Primo & Shavelson, 1996). Various authors confirm the suitability of computer-based tests for experimental tasks to diagnose pupils' work processes (e.g. Schreiber, 2012). Nevertheless, it should not be overlooked that non-hands-on tests simplify or even omit steps of experimentation (Schecker & Parchmann, 2006). To illustrate, Ludwig, Priemer, and Lewalter (2018) show that pupils tend to react differently when identifying a false hypothesis in real experiments and in computer simulations, with pupils holding onto a false hypothesis in the real experiment settings. Considering the correct handling of substances, experimental objects, measuring instruments and laboratory materials as an important component of scientific work, the use of hands-on tests is necessary. In some of the investigations mentioned above, which included hands-on tests, data was collected from pupils working in small groups. For individual diagnosis, data must be collected from pupils who work individually. There are some limitations to some of the studies in which hands-on tests were used and pupils who worked individually were analysed:

- scientists intruded on pupils' experimental processes (investigation of Hammann et. al., 2008: pupils were given pictures of the result; investigation of Germann et al., 1996: pupils received model answers after each step to be able to continue working),
- guidelines were set (intervention of Kanari & Millar, 2004: selection of hypotheses),
- only partial competencies of the experimental process were focused on; e.g. using the variable control strategy (investigation of Chen & Klahr, 1999), the rejection of hypotheses (investigation of Ludwig et al., 2019).

The only hands-on test survey carried out on individual working pupils in which the experimental process was viewed in its entirety took place in a single experiment, which did not involve the use of laboratory equipment or laboratory materials, nor did it address scientific inquiry as such. It approached the process from a general problem-solving angle: in the experiment, Dunbar and Klahr (1989) asked test subjects to find out the function of a button on a computer-controlled vehicle.

The work of Shavelson, Ruiz-Primo, and Wiley (1999) and Ruiz-Primo and Shavelson (1996) shows that the results of performance assessments can vary with content. Further surveys with hands-on tests must therefore be carried out with individually working pupils and with other experimental contents.

Dunbar and Klahr's study focussed on pupils aged 8 to 12 and adults and did not compare the different age groups of pupils. Piaget describes, in his theory, that a change in hypothetical-deductive thinking occurs when changing from the *concrete operational stage* to the *formal operational stage* at the age of 11-15 years old (Ginsburg & Opper, 1998). In more recent developmental psychological approaches to formal-scientific thinking, however, it is not assumed that there are large interindividual differences, but rather continuous cognitive development (not in stages, as in Piaget's theory) and an earlier basic understanding of the logic of hypothesis testing and evidence evaluating are assumed (Koerber, Mayer, Osterhaus, Schwippert & Sodian, 2014). Consequently, more recent approaches also assume continuous development. As Dunbar and Klahr

doi: 10.11576/zdb-3576

did not analyse the differences of the different aged pupils, studies conducted with older pupils and with comparisons of the pupils of different ages may also complement the findings.

3 Research Questions and Hypotheses

The following research questions regarding pupils' errors (manual mistakes, methodological errors, misconceptions, and pupil-specific approaches) can be derived from the aspects described above.

- Research question 1: Which of the described pupils' errors in experimenting can be replicated with other content and survey methods?
- Research question 2: What additional, yet-to-be described pupil errors in experimenting can be identified with individual diagnostics?
- Research question 3: Does the occurrence of pupil error in experimenting depend on the content?
- Research question 4: Does the occurrence of pupil error in experimenting differ among pupils of different ages?

The research questions 1, 2, and 3 led to an investigation of 5th and 6th graders between 10 and 12 years old (Baur, 2018, 2016). In the analysis described in this article, existing data (5th and 6th graders) was complemented with data from 9th graders. For this purpose, 9th graders from age 14 to 16 years old were observed while planning and conducting experiments, as with the 5th and 6th graders. This allows research questions 1, 2, and 3 to be examined using a larger sample and research question 4 to be examined by comparing 5th and 6th graders to 9th graders. The investigation of research questions 1 and 2 was carried out in an explorative manner. Research questions 3 and 4 were examined based on hypotheses. The hypothesis for research question 3 is that content has an influence on the occurrence of pupils' errors. This is suggested by the research results of Shavelson, Ruiz-Primo, and Wiley (1999) and Ruiz-Primo and Shavelson (1996), who found out that the results of an assessment may depend on its content.
The hypothesis for research question 4 is that age influences the outcome and that the results of the two samples, '5th and 6th graders', and '9th graders', differ. The hypothesis is based on Piaget's theory of cognitive development and on recent developmental psychological approaches, in which it is assumed that there is an increase of the abilities in formal operational thinking in adolescence.

4 Method

4.1 Sample

The data was collected from $n_1 = 12$ 9th graders (aged 14-16 years, M = 15.04, SD = 0.61). The pupils attended a secondary school in Germany (Baden-Württemberg).
The sample (n_1) consisted of six boys and six girls. Teachers asked pupils who met the specified achievement levels (three categories of achievement levels: good school performance, average school performance, and poor school performance) to participate in the experiment. All pupils volunteered to take part in the investigation. In addition, the study was undertaken with the pupils' parents' consent. The school grades in mathematics, German and natural science subjects were relevant for the classification. For each category, the sample consisted of an equal number of boys and girls. Using school grades for the assessment was a compromise for investigating a heterogeneous group of learners and provided a reasonable workload for the schools. The use of competence tests would have resulted in prolonged testing periods. As explained above, the data are compared with the data of a previous study of 5th and 6th graders (Baur, 2016, 2018). The pupils were 5th and 6th graders (age: 10-12 years, M = 11.51, SD = 0.64) from two different secondary schools in Germany (Baden-Württemberg) recruited in the same way as in sample n_1. The sample size is $n_2 = 18$ (six pupils from the 5th grade – three girls and three boys – and twelve pupils from 6th grade – six girls and six boys).

4.2 Data recording

The pupils were filmed while experimenting and had to report their work in an experimentation report (a sheet with the headings: Hypothesis, Material, Sketch of the Experiment, Description, Observation, Result and with space for own texts for each heading). They were given two tasks for which they had to plan an experiment, perform the experiment, and interpret the results. Task 1 was a yeast experiment, in which pupils were told to: 'Find out what yeast needs to form carbon dioxide. Material for experimentation is on the table.' Task 2, a cone scale experiment, instructed pupils to: 'Find out what triggers cone scales to close. Material for experimentation is on the table.' For the planning and implementation of the experiment, pupils had a wide range of materials to choose from. Pupils were told that they don't have to use all the materials, but they can if they want to. The material for the yeast experiment consisted of

doi: 10.11576/zdb-3576

Erlenmeyer flasks, beakers, test tubes, stoppers, balloons, test tube rack, precision scale, spatula, stirring rod, paper towels, flour, salt, cold and warm water, and yeast. The impulse for the experimental process was a short film in which pizza dough expands in a closed can with the resulting carbon dioxide blowing the lid off. Before the film was shown, the ingredients for pizza dough were listed in a conversation. After the film, the pupils were told that carbon dioxide is produced during fermentation and task 1 was discussed. For the second experiment (cone scale experiment) the following materials were provided: dried conifer cones, beakers, cardboard, paper towels, thermometer, hairdryer, cooler, ice cubes, and water. Two images were used as prompt: the first picture showed a conifer cone with opened scale, and the second picture showed that the cone scales were closed, one could see that it had started raining and that had become darker and colder outside (on each picture a thermometer was visible). After discussing the differences between the pictures, task 2 was discussed.

If both tasks are compared, the cone scale task could be considered to be easier for pupils because:

- the amount of substances (quantity of water) or size of objects (size of cones) is not so decisive for a comparison between two or more experimental trials in the cone scale experiment
- the closing of the cones can be seen directly in contrast to the occurring gas (CO_2) in the yeast experiment

The pupils worked on both tasks in the same day and had 60 minutes for each task. Pupils performed the tasks in the same order: the yeast experiment followed by the cone scale experiment. They worked individually and each of them was supervised by a student assistant.

The student assistants were trained beforehand on carrying out the following tasks:

- check whether pupils understood the task properly and provide explanations if needed,
- explain how to complete the experimentation report when experimenting,
- remind to continue filling in the experimentation report,
- cue pupils' 'thinking out loud'.

The student assistants did not help in the investigation execution.

Due to organisational reasons, data recording of the 9th graders differs in two aspects from the data recording of the 5th and 6th graders:

- pupils carried out both experiments on the same day (It proved difficult to find 9th graders willing to participate in the survey as it lasted two days and took place after regular classes. Because of this, school principals agreed to spare lesson time, provided it could be done in one day),
- no hot water could be provided for the cone scale experiment, which was ultimately not essential for the experiment anyway.

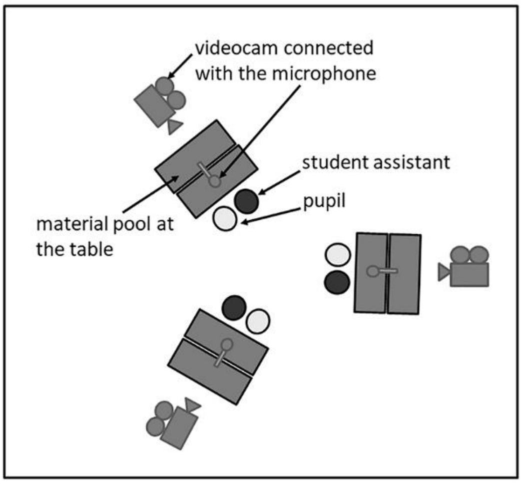

Figure 1. The setting of data recording

4.3 Data analysis

The video data (9th graders: 1440 minutes; 5th and 6th graders: 2100 minutes) were transferred to observation protocols (integration of audio and video tracks). The software ELAN was used for transcriptions. The experimentation reports were included in the observation protocols. Mayring's (2008) Qualitative Content Analysis was used to reduce discourse of each observation protocol without reducing content. This was achieved in two steps; first by *paraphrasing,* which entails leaving out all parts of the text that do not carry content and simplifying it; and second, by *generalising,* which means generalising observations to a wider level of abstraction (Mayring, 2008, p.62).

After minimizing the amount of text, the data were analysed in a deductive manner using a category system developed in prior research (see Baur, 2016, 2018), which was open to inductive extension. Since new errors could be discovered during analysis of further pupils, the category system must be expandable. Assuming that new errors are discovered, all previous observation protocols would have to be checked for those. The category system used is listed in Table 1 and 2 respectively, and in Table 6.

doi: 10.11576/zdb-3576

The analysis of the data material differentiated two levels:

- Level 1 *microanalysis*: consideration of each step in a pupil's experimental activity (category system: see Table 1 and Table 2).
- Level 2 *macroanalysis*: consideration of the pupil's overall approach to understanding the pupil's strategy to solve the given problem (category system: see Table 6).

The analysis' objectivity was checked with an interrater, who also analysed 25% of the sample (interrater values are listed in Table 3, 4 and 6). Observation protocols for the interrater were chosen at random, but equivalent numbers of observation protocols were used for both contents the yeast experiment and the cone scale experiment. The interratings were used to calculate Cohen's Kappa (κ) and the percentage match (PM). In some cases, κ could not be calculated because of a division by zero[2] or when one of the variables appeared as constant[3]. In these cases, only the PM was used in the analysis. κ and PM could not be calculated separately for the two tasks in 9th graders, due to the sample size (n_1). If the interrating-samples of the yeast experiment and the cone scale experiment of the 9th graders had not been merged, only very few κ values could have been calculated. The merging of the data resulted in fewer constants and divisions by zero (see footnotes 2 and 3). The code-recorde method (the data was re-analysed at intervals of 2 to 3 months) was used to reliably analyse (cf. Krefting, 1991) and the interrating method (the measure of objectivity was also used as a reliability measure) was used to check reliability.

Two of the video recordings of the 9th graders could not be analysed, both of which were on the subject of yeast fermentation. The audio track was missing in one of the films and the student assistant helped in the other.

For the investigation of research question 3, correlations were calculated (Pearson correlation and Yule's Y) for the microanalysis. In the case of dichotomously distributed data, the correlation analysis results in a value that takes into account the coincidence of subject-related measurement agreements and non-agreements in the event of an error in the two experiments. The correlations control the connectedness between variable pupils' error X of yeast experiment with the variable pupils' error X of cone scale experiment. For example, the correlation of error H1 of the yeast experiment with H1 of the cone scale experiment, the correlation of error H2 of the yeast experiment with H2 of the cone scale experiment, and so on. If the correlation coefficient for a pupil's error is close to 1.0, most of the test subjects in the yeast experiment have a comparable result with regard to an error as in the cone scale experiment. In general, the Phi coefficient is an accepted correlation with clear interpretations for 2x2 tables (Bonett & Price, 2007). The Pearson correlation is, for dichotomous data, equivalent to the Phi coefficient. The odds ratio is another correlation for 2x2 tables which have excellent properties for small samples (Bonett & Price, 2007). Yule's Y is a function of the odds ratio and leads to values between -1 and 1.

To examine research question 4, Fisher's exact test was used. The Fisher exact test is applied to small samples ($n < 40$) and is required when a value in the 2x2 table is less than 5 (Ludbrook, 2008).

All statistical computing was carried out using statistics software *R* (Version 3.6.1). In general, analysis methods which are recommended for small samples were used.

5 Results

5.1 Results of microanalysis

5.1.1 Identifiable pupils errors. Tables 3 and 4 present errors that occurred in at least one of the two experimental tasks and, therefore, ones that were exhibited by more than two pupils of one of both samples (n_1 or n_2). In Tables 1 and 2, all pupils' errors are defined, and examples of the samples (n_1 and n_2) are presented. Tables 1 and 2 show the entire category system, all non-occurred pupils' errors are marked in grey font colour. For a better overview, the identified errors were assigned to subprocesses of the experiment in accordance with Friedler, Nachmias, and Linn (1990). In addition, the type of error was distinguished in the presentation: pupil-specific approaches are presented in Table 4 and manual mistakes, methodological errors, and misconceptions in Table 3.

[2] If an error in the entire sample is assessed as having occurred or not occurred by both raters, κ cannot be calculated. Thus, a zero is generated under the fraction bar of the calculation.

[3] If an error is assessed as having occurred or not occurred by only one of the interraters in the entire sample, κ automatically becomes zero, even though the consensus of both raters might be eminent.

doi: 10.11576/zdb-3576

Table 1

Category system (section 1) with descriptions and examples

	Issue	Description	Example
State a hypotesis	Hypothesis is not based on a variable, but on an expected observation. (H1)	*P. ignores the different variables and limits him/herself to considerations concerning a possible observation (effect).*	– *P-003(♀, Y): "Hmm, what should I write down now [protocol]? So what do I, um, what I expect, right?" Student: "Mhm" P-003: "I expect that the lid pops off"* – *P-004(♂, C): writes assumption in the protocol: The cone scales close.*
	Hypothesis consists of variable combination. (H2)	*P. hypothesises that it does not consist of a single variable, but of a combination of variables. P. must/would have to create a multi factorial experimental design to examine all combinations.*	– *P-007(♂, C): writes assumption in protocol: cold, wet, dark.* – *Student: "What do you suspect?"; P-005(♂, Y): „Uh, flour, salt and yeast“*
	Hypothesis is changed during the experiment without being checked. (H3)	*While working, the p. spontaneously changes the hypothesis (several times) without having confirmed or refuted the previous hypothesis.*	– *P-001(♀, C) gradually writes down three assumptions without the experimental approaches having been completed (in some cases the approaches (set-ups) were only developed).*
	Hypothesis is only proposed during (or at the end of) the experiment. (H4)	*P. first conducts the experiment without a hypothesis, but then stating a hypothesis during or at the end of the experiment. The p. first tries it out.*	– *P-015(♀, Y) asks student after setting up an experimental approach: "Um, should I write down my assumption?" Student: "Mhm, what do you think the yeast needs?" P-015: "Err, maybe err, yeast dough, err, no, that's what you do, maybe salt."*
	No hypothesis is proposed. (H5)	*P. performs an experiment without explicit hypothesis.*	– *After P-011(♀, Y) has set up two experimental approaches, student asks: "Did you have an assumption at the beginning? P-011: "What needs to be put in there?", P-011 is reflecting, Student: "Or else you can continue " P-011: "Ok ".*
Design and conduct an experiment	Missing test trial. (D1)	*P. creates a test series without a test trial or with not all test trials. No variation of the test variables.*	– *P-001(♀, Y) puts forward the hypothesis 'salt' and sets up an experimental trial with salt, flour, warm water and yeast. After a short time P-001 says: "I suspected that it was the salt" Student: "Ok". S-001: "But I am not sure anymore".*
	Missing control trial. (D2)	*P. creates a test series without a control trial. In a control trial none of the variables is varied.*	– *P-013(♀, Y) hypothesises warm water during yeast fermentation and sets up different experimental trials, none of which contain all variables. P-013 then says: "I don't think that will happen".*
	Approach without necessary component. (D3)	*P. plans and prepares experimental trials and forgets at least one necessary component (e.g. yeast).*	– *P-010(♀, Y) sets up experimental batch, fetches warm water in a beaker, adds salt and flour and mixes. P-010: "I don't know what to do now" Student: "What have you done so far?" P-010: "I put warm water, salt and flour in it. Student: "And what were you trying to, uh, show or find out?" P-010: "I just wanted to try it".*
	Use of different laboratory equipment between the individual experimental trials. (D4)	*P. plans coherent experimental approaches with varying set-up (laboratory equipment).*	– *P-018(♂, Y) builds up a test batch to his hypothesis 'flour is necessary for yeast fermentation' with yeast, flour, warm water in a test tube (closure: balloon) and a test batch with yeast, flour, salt, cold water in an Erlenmeyer flask (closure: stopper).*
	The quantities of substances used in the different trials are not equivalent. (D5)	*P. does not take into account that the quantities of substances used in the trials of the test series must be equivalent if they are not to be specifically modified.*	– *P-017(♂, Y) pours all substances into the test vessels without measuring (weighing) them beforehand.*
	Intentional (but illogical) variation of sets. (D6)	*P. intentionally varies the quantities in the approaches - without establishing a reference to the variable.*	*This is not evaluated in the survey described.*

P.: Pupil; Student: Student assistant; (Y): Experimental content ‘yeast fermentation’; (C): Experimental content ‘closing movement of cone scales’

doi: 10.11576/zdb-3576

Table 2

Category system (section 2) with descriptions and examples

	Issue	Description	Example
Design and conduct an experiment	Focus on proper execution of some selected trials. (D7)	*P. only works precisely and neatly with one or several trials, the others are not assembled precisely.*	*Did not occur in described survey.*
	Trials with the same content (no variation). (D8)	*Several test trials of a test series with the same content.*	– *P-002(♀, Y) twice sets up an experimental approach with the same ingredients.*
	Pupil no longer knows what is included in one (or some) experimental trials. (D9)	*P. has not described approaches and no longer knows what is in which trial or mixes up the trials.*	– *P-010(♀, C) asks student: "Which was the fourth [Trial No.4]?"*
	When conducting the experiment, the pupil forgets to add recorded materials to the experimental set-ups. (D10)	*P. plans to use certain substances/laboratory equipment but forgets to use them during implementation.*	– *P-005(♂, Y) puts warm water, flour and salt into a test tube and closes it with a stopper. The protocol also lists yeast as an ingredient.*
	Experimental trials are altered. (D11)	*P. changes experimental trials (several times) without deliberating about effects - adds something else, opens the stopper, stirs etc.*	– *P-002(♀, C) places cones from the 2nd trial in the glass of the 3rd trial. For this purpose, P-002 removes the cone from trial 3 and sets it aside.* – *P-005(♂, Y) opens test tube no. 4, stirs and closes it again.*
Observe and analyse data	Observation only in one or a few trials. (O1)	*P. makes observation only on some, not all, trials. Concentration on one / few trials.*	*Did not occur in the survey described.*
	Pupil believes that something must happen/be observed immediately. (O2)	*If nothing happens / nothing becomes visible immediately, the p. assumes it doesn't work.*	– *P-011(♀, Y) sets up the experimental trial; she is a bit insecure, stirs in it and says: "It doesn't look like that", "I think I'll try it again, it won't work".*
	The result describes which trial provides the best result (no independent variable statement). (A1)	*The result describes which experimental trial works best but no variable is discussed.*	– *P-023(♂, C) writes in the protocol: Cone scales closed best in ice water, then follows cold and dark and least when only dark.*
	The final result cannot be derived from the trials/observations. (A2)	*The determined result cannot be derived from the implementation / observation.*	*This is not evaluated in the survey described.*
	Observation/assumption as formulated result of the experiment. (A3)	*P. indicates only one observation or assumption as a result.*	*This is not evaluated in the survey described.*
	No final result. (A4)	*P. gives no final result.*	– *In general: Neither a written result in the protocol, nor an oral result.*
	Pupil believes that he/she can only write down a result when his/her intended result occurs. (A5)	*P. believes that he/she can only write down a result if what was previously expected occurs.*	– *Student: "Based on what you saw, can you already write down a result?"* *P-003(♀, Y): "Hmm, well, no, not really, because [pause] no, I don't think so, because, if that doesn't rise [P. points to trial 1], then I've got it wrong".*

P.: Pupil; Student: Student assistant; (Y): Experimental content 'yeast fermentation'; (C): Experimental content 'closing movement of cone scales'

In Tables 3 and 4, the results of the 5th and 6th graders, as well as the 9th graders, are shown for each type of error. The corresponding p-value concerning the difference between the two groups is shown in the right column. Each cell of the results section contains the number of pupils presenting the error. The percentage (in italics) is displayed in the middle of each result cell. It is unusual to use the percentage for small samples (if the sample size is less than 100, not all percentages are possible). The use of percentage was a compromise be-

doi: 10.11576/zdb-3576

cause of the comparison of the two samples with different sample sizes. At the bottom of each cell, the number of boys and girls showing the error is indicated in parentheses. In the set of the 9th graders, two pupils could not be analysed in the yeast experiment (see 4.3 Data analysis). In the sample of 5th and 6th graders two pupils from the cone scale experiment could not be analysed. In one case the student assistant made a mistake, and in the other, the pupil did not take part in the experiment.

Table 3.
Recorded pupils' manual mistakes, methodological errors and misconceptions

		Ninth Graders				**Fifth and Sixth Graders**				**Differences: Ninth G. and Fifth+Sixth G.** Y: Yeast Exp. C: Cone Scale Exp.
		Yeast Experiment		Cone Scale Experiment		Yeast Experiment		Cone Scale Experiment		
	Error	Count ***Percent*** [♂;♀]	Kappa (PM)	Count ***Percent*** [♂;♀]	Kappa (PM)	Count ***Percent*** [♂;♀]	Kappa (PM)	Count ***Percent*** [♂;♀]	Kappa (PM)	
State a hypothesis	Hypothesis is not based on a variable, but on an expected observation. (H1)	4 ***40%*** [3;1]	nc_2 (83%)	1 ***8%*** [0;1]	nc_2 (83%)	1 ***6%*** [0;1]	nc_1 (100%)	2 ***13%*** [1;1]	nc_1 (100%)	***Y: p = .04*** C: ns
	Hypothesis is changed during the experiment without being checked. (H3)	6 ***60%*** [2;4]	0.57 (83%)	7 ***58%*** [3;4]	0.57 (83%)	7 ***39%*** [5;2]	1.0 (100%)	6 ***38%*** [3;3]	1.0 (100%)	Y: ns C: ns
Design and conduct an experiment	Missing test trial. (D1)	7 ***70%*** [2;5]	nc_2 (83%)	11 ***92%*** [6;5]	nc_2 (83%)	12 ***67%*** [8;4]	0.50 (75%)	10 ***63%*** [4;6]	nc_1 (100%)	Y: ns C: ns
	Missing control trial. (D2)	5 ***50%*** [1;4]	0.67 (83%)	7 ***58%*** [3;4]	0.67 (83%)	6 ***33%*** [2;4]	0.50 (75%)	6 ***38%*** [3;3]	0.50 (75%)	Y: ns C: ns
	Approach without necessary component. (D3)	**0**	nc_1 (100%)	**0**	nc_1 (100%)	6 ***33%*** [3;3]	1.0 (100%)	**0**	nc_1 (100%)	Y: ns C: ns
	Use of different laboratory equipment between the individual experimental trials. (D4)	5 ***50%*** [1;4]	nc_2 (83%)	**0**	nc_2 (83%)	15 ***83%*** [8;7]	0.50 (75%)	1 ***6%*** [1;0]	nc_1 (100%)	Y: ns C: ns
	The quantities of substances used in the different trials are not equivalent. (D5)	5 ***50%*** [1;4]	0.57 (83%)	**0**	0.57 (83%)	14 ***78%*** [6;8]	1.0 (100%)	**0**	nc_2 (75%)	Y: ns C: ns
	Experimental trials are altered. (D11)	3 ***30%*** [2;1]	nc_2 (83%)	5 ***42%*** [2;3]	nc_2 (83%)	8 ***44%*** [5;3]	1.0 (100%)	12 ***75%*** [7;5]	1.0 (100%)	Y: ns C: ns
Observe and analyse data	Pupil believes that something must happen / be observed immediately. (O2)	3 ***30%*** [0;3]	1.0 (100%)	**0**	1.0 (100%)	6 ***33%*** [2;4]	1.0 (100%)	**0**	nc_2 (75%)	Y: ns C: ns
	No final result. (A4)	2 ***20%*** [1;1]	nc_2 (83%)	1 ***8%*** [1;0]	nc_2 (83%)	8 ***44%*** [3;5]	1.0 (100%)	**0**	nc_1 (100%)	Y: ns C: ns

PM: Percentage match
G.: Graders
nc_1: Not calculable; all pupils are assigned a value by the raters in accordance with each other
nc_2: Not calculable; a variable is a constant (a rater assigns a value to all pupils)
ns: Not significant

doi: 10.11576/zdb-3576

Table 4.
Recorded pupils' specific approaches

	Errors	Ninth Graders: Yeast Experiment Count *Percent* [♂;♀]	Ninth Graders: Yeast Experiment Kappa (PM)	Ninth Graders: Cone Scale Experiment Count *Percent* [♂;♀]	Ninth Graders: Cone Scale Experiment Kappa (PM)	Fifth and Sixth Graders: Yeast Experiment Count *Percent* [♂;♀]	Fifth and Sixth Graders: Yeast Experiment Kappa (PM)	Fifth and Sixth Graders: Cone Scale Experiment Count *Percent* [♂;♀]	Fifth and Sixth Graders: Cone Scale Experiment Kappa (PM)	**Differences: Ninth G. and Fifth+Sixth G.** Y: Yeast Exp. C: Cone Scale Exp.
State a hypothesis	Hypothesis consists of variable combination. (H2)	6 ***60%*** [2;4]	1.0 (100%)	11 ***92%*** [6;5]	1.0 (100%)	7 ***39%*** [6;1]	1.0 (100%)	5 ***31%*** [5;0]	1.0 (100%)	Y: ns ***C: p = .002***
Design and conduct an experiment	Trials with the same content. (D8)	**0**	nc₁ (100%)	2 ***17%*** [2;0]	nc₁ (100%)	2 ***11%*** [1;1]	nc₁ (100%)	3 ***19%*** [2;1]	nc₁ (100%)	Y: ns C: ns

PM: Percentage match
G.: Graders
nc_1: Not calculable; all pupils are assigned a value by the raters in accordance with each other
nc_2: Not calculable; a variable is a constant (a rater assigns a value to all pupils)
ns: Not significant

Table 5.
Correlations

Variables	Yule's Y	Pearson
$H1_{yeast}$ <> $H1_{cone}$	***.56***	***.43 p < .05***
$H2_{yeast}$ <> $H2_{cone}$	***.55***	***.55 p < .001***
$H3_{yeast}$ <> $H3_{cone}$	***.53***	***.54 p < .001***
$D1_{yeast}$ <> $D1_{cone}$	-.03	-.03 p > .05
$D2_{yeast}$ <> $D2_{cone}$	.06	.06 p > .05
$D3_{yeast}$ <> $D3_{cone}$	nc	nc
$D4_{yeast}$ <> $D4_{cone}$	.39	.13 p > .05
$D5_{yeast}$ <> $D5_{cone}$	nc	nc
$D8_{yeast}$ <> $D8_{cone}$	.45	.28 p > .05
$D11_{yeast}$ <> $D11_{cone}$	.04	.04 p > .05
$O2_{yeast}$ <> $O2_{cone}$	nc	nc
$A4_{yeast}$ <> $A4_{cone}$	-.43	-.15 p > .05

nc: Not calculable

5.1.2 Correlations between the pupils' errors in different contents. As before in the identification of pupils' errors, only errors were included to determine correlations, that occurred in at least one of the two experimental tasks and, therein, exhibited by more than two pupils of one of both samples. In total, three significant correlations were found. The correlation values are listed in Table 5.

5.1.3 Difference between the two samples (age effects). The calculations with Fisher exact test (see the right column in Table 3 and 4) show only two significant differences:

- Error (H1) 'Hypothesis is not based on a variable, but on an expected observation' is observed more frequently in 9th graders (in the yeast experiment).
- Error (H2) 'Hypothesis consists of variable combination' is more often identified in 9th graders (in the cone scale experiment).

5.2 Results macroanalysis

5.2.1 Oberseved and identified pupils' strategies. In the 9th grade sample, a new (sub-)strategy

doi: 10.11576/zdb-3576

Table 6.
Pupils' strategies

Approach/Strategy	Definition/Description	Ninth Graders: *Yeast Experiment* Kappa: 0.75 PM: 83% Count ***Percent*** [♂;♀]	Ninth Graders: *Cone Scale Experiment* Kappa: 0.75 PM: 83% Count ***Percent*** [♂;♀]	Fifth and Sixth Graders: *Yeast Experiment* Kappa: 0.60 PM: 75% Count ***Percent*** [♂;♀]	Fifth and Sixth Graders: *Cone Scale Experiment* Kappa: nc_2 PM: 75% Count ***Percent*** [♂;♀]	**Differences in frequency distribution in all approaches/ strategies: Ninth G. and Fifth+Sixth G.** Y: Yeast Exp. C: Cone Scale Exp.
At first only one experimental trial with all variables. (S1)	Pupil works with hypothesis based on variables. Pupil first uses only one experimental trial that includes all variables (not varied). If this does not lead to the intended result, the next trial is set up.	2 ***20%*** [1;1]	1 ***8%*** [1;0]	6 ***33%*** [4;2]	1 ***6%*** [0;1]	Y: ns ***C: p = .004***
Control variables are varied in the experimental trials. (S2)	Pupil works with hypothesis based on variables. Pupil plans experimental trials in which one or more control variables are varied. At the beginning, pupil ***does not*** limit him- or herself to creating only one experimental trial with all variables and waiting for the result; ***which means S1 does not apply.***	5 ***50%*** [1;4]	10 ***83%*** [5;5]	8 ***44%*** [5;3]	13 ***81%*** [8;5]	
Only suspected variable(s) and necessary variable(s) included. (S2-A)	Experimental set-up (experimental set-ups: if there are several hypotheses), always consists of only the necessary and the independent variable or variables. All other variables that are relevant to the question are not included.	2 ***20%*** [0;2]	6 ***50%*** [3;3]	1 ***6%*** [1;0]	***0***	
S2 but not S2-A (S2-B)	S2 but not S2-A.	3 ***30%*** [1;2]	4 ***33%*** [2;2]	7 ***39%*** [4;3]	13 ***81%*** [8;5]	
Working in engineer mode. (S3)	The only defined goal (also defined in the hypothesis) of the pupil is to achieve a certain effect (stopper pops up etc.) and not to answer the question.	3 ***30%*** [2;1]	1 ***8%*** [0;1]	1 ***6%*** [0;1]	2 ***13%*** [1;1]	
Work without hypothesis. (S4)	Pupil works without a proposed hypothesis.	***0***	***0***	3 ***17%*** [0;3]	***0***	

PM: Percentage match
G.: Graders
nc_1: Not calculable; all pupils are assigned a value by the raters in accordance with each other
nc_2: Not calculable; a variable is a constant (a rater assigns a value to all pupils)
ns: Not significant

(S2-A: Only suspected variable(s) and necessary variable(s) included) was found and therefore all S2 cases of the 5th and 6th grade sample have been analysed once more. The results are shown in Table 6. The structure of Table 6 is similar to Tables 3 and 4.

5.2.2 Difference between the two samples. The distributions of observable strategies differ percentage-wise between samples and are significant for the cone scale experiment (see the right column in Table 6). For the analysis of the differences in the strategies between the samples, all strategies were utilised to create a 5x2 table (5 strategies x 2 samples). For the two 5x2 tables (one for the yeast experiment and one for the cone scale experiment), differences were assessed utilising the Fisher exact test.

doi: 10.11576/zdb-3576

6 Discussion

With reference to question 1, the analysis showed that errors already described in the literature (see Chapter 2: Previous results of research) were replicated by the observations of the 5th, 6th, and 9th graders and can be summarized as follows:

- For many pupils, the aim of experimenting is to achieve an effect (cf. H1, S3). In each of the experiments, 13 subjects changed their hypothesis (repeatedly) without testing it (H3). This error can also be interpreted as a tendency to generate an effect or as impatience during work.
- Many pupils experiment without a test or control trial. (cf. D1, D2). One striking result of the study is that pupils comparatively often forgot to include the test trial (trial in which the test variable is varied) rather than the control trial (trial in which no variable is varied). Neglecting the test trial makes it extremely difficult, if not impossible, to draw evidence-based conclusions about the independent variable.
- Pupils often try things out without being strictly scientific (cf. H4, D8, D11).
- If the data obtained do not correspond to the expected data, pupils suspect an error in their experimental performance (cf. A5).
- Pupils tend to change too many variables at once and cannot draw conclusions. In the yeast experiment, 13 pupils proposed hypotheses, and in the cone scale experiment, 16 pupils proposed hypotheses consisting of a combination of variables (H2). While this issue may not be an error, it does make it difficult for pupils to plan experiments independently because a multi-factor experimental design is required.
- In part, learners do not formulate any hypothesis in their experimental work (cf. H5). However, this could only be observed in one pupil. This small number can possibly be explained by the fact that an experimentation report was kept which provided space for a hypothesis.

Replicating these pupils' errors listed above using a different method confirms the previously generated evidence.

In addition, further pupils' errors could be identified which have not yet been described. For the yeast experiment, six pupils planned and built experimental set-ups which did not respect the vital variable yeast (D3). This phenomenon was only observed in 5th and 6th graders and needs further investigation. However, compared to 9th graders, there is no significant difference. In the experimental content 'yeast' 20 pupils used different laboratory equipment between the related experimental trials (D4) and in the cone scale experiment, only one pupil did so. When working on the experimental content 'yeast', 19 pupils did not consider that they had to use equivalent quantities of substances in their related experimental trials (D5). Germann et al. (1996) were able to observe in their study that a lot of pupils do not explicitly describe the use of equivalent laboratory equipment and quantities of substances in their experiment protocols, which was already an indication that it is not relevant for pupils. The two issues (D4 and D5) can be summarised and generalised as a 'trial and error' approach of pupils. Other literature has also described it in this general way (Meier & Mayer 2012; Wahser & Sumfleth 2008; Hammann et al., 2008). Strictly speaking, the use of different laboratory equipment in related experimental trials (e.g. vessels with different volumes; one trial with a balloon to detect CO_2 development and the other trial with a stopper) and the use of non-comparable quantities in a series of experiments could also be considered an error in keeping the control variables constant. As a consequence of the error D3, to improve experimental skills, it seems to be of crucial importance in class to teach the pupils about vital variables as well as independent and dependent variables. Another point of particular relevance for the lesson is the necessity of teaching concerning the non-variation of laboratory material and amount of material used (conclusion on pupils' errors D4 and D5).

In the macroanalysis, which identified strategies describing the overall process, all pupils, in both contents, could be assigned to four classes. One class (S2: 'varying control variables') could be further subdivided into two subclasses S2-A and S2-B (S2-A: Only suspected variable(s) and necessary variable(s) included; S2-B: S2 but not S2-A). This subdivision is an extension of previously published results on the strategies (Baur, 2018). In terms of frequency, the 5th and 6th graders were most likely to 'vary control variables' (S2) and 'work with only one trial containing all variables' (S1), whereas the 9th graders were more likely to 'work in engineering mode' (S3) and 'vary control variables' (S2). It is noticeable that no pupil used the 'control of variables' strategy adequately.

doi: 10.11576/zdb-3576

With regards to question 3, only a few correlations between pupils' errors were found. Some of these were found in the subprocess of 'hypothesis generation'. This suggests that the content determines the occurrence of an error in the other subprocesses. The results need further examination, as a learning effect as a result of the first experiment is also conceivable. However, it also illustrates once more that an experimental competence should not only be investigated with one experimental content (see also Shavelson et al., 1999; Ruiz-Primo & Shavelson, 1996), but with several contents in cross-design, e.g. a change of the presented order of the experimental contents within the sample. A non-significant correlation at D5 (the quantities of substances used in the different trials are not equivalent) cannot be interpreted. The error could not be unambiguously evaluated at D5 in the cone scale experiment (difficult to recognize in the video) and was therefore only evaluated for the yeast experiment.

The comparison of the age groups (question 4) reveals only a few significant differences. This result stands in contrast to findings on scientific thinking, in which an age-related development of abilities is described (Koerber et al. 2014; Piaget, 2001; Ginsburg & Opper, 1998). It may be possible to detect other differences if the sample is expanded to include even younger pupils. The differences concern the percentage distribution of the strategies as well as the errors H1 (hypothesis is not based on a variable, but an expected observation) and H2 (hypothesis consists of a combination of variables). H1 occurred more frequently among older pupils in the context of the yeast experiment and H2 occurred more frequently among older pupils in the context of the cone scale experiment. Previous knowledge of pupils should be considered a limitation regarding the investigation of question 4, as it was not feasible to control the pupils' knowledge gained prior to the experimentation. The pupils do not differ, despite the fact that 9th graders should have learned more about experimenting and should have gained more experimental practice than 5th and 6th graders (see Educational Standards of Baden-Württemberg: MKJS, 2004). The results of question 4 suggest that gaining experimental competence does not take place solely through adolescence and cognitive development alone but requires instruction in which pupils learn explicitly about the process and the elements of an experiment. Example questions that hint at such explicit knowledge include: What exactly is the definition of an experiment? What is the definition of a hypothesis and the reason for hypothesis testing? What is the 'control of variable strategy' and which parameters can or need to be controlled in an experiment?

The results of the entire study open up approaches for lesson planning and might lead to further research questions. Applications for lesson planning include, but are not limited to, the following examples. First, for scaffolding, knowledge about pupils' errors could be helpful for addressing them individually. Thus, this knowledge enables the supervisors to support the pupils in moving up one level in a sub-process of experimentation. Second, for pedagogical diagnostics purposes, the pupils' errors could be general items of routine analysis that a teacher takes into account for the individual assessment and planning. Questions about further research that have emerged are the following questions. Are there any relationships between the different pupils' errors? What kind of teaching techniques are useful to address pupils' errors? Are there any pupils' errors that are more likely to be discussed with younger or older pupils?

In the evaluation of all results, possible effects of communication between student assistants and pupils must be considered as a limitation. Despite control of communication, pupils can be encouraged to think when student assistants ask them questions about the pupil's actions and thoughts.

doi: 10.11576/zdb-3576

Acknowledgements

The author would like to thank: Kathrin Beißwenger for her help in organizing and coordinating the implementation in the schools; the school management of the Realschule Heubach; the student assistants Pia Lena Baier, Aileen-Monique Schmidt, Stephanie Barth, Juliane Purr; Arne Bewersdorff for work as an interrater; and Sophie Le Boulanger and Markus Emden for proofreading. Many thanks to the pupils who took part.

Funding

This study was financially supported by Deutsche Telekom Stiftung.

Compliance with Ethical Standards

Conflict of Interest: The author declares that there are no conflicts of interest.

Ethical and Consent Statements: All procedures performed with human participants were in accordance with the ethical standards of the German Society of Psychology (DGPs) and the rules for surveys at schools in Baden-Württemberg (rules of the ministry of education Baden-Württemberg).

doi: 10.11576/zdb-3576

References

Baur, A. (2015). Inwieweit eignen sich bisherige Diagnoseverfahren des Bereichs Experimentieren für die Schulpraxis? [To what extent are present diagnosis procedures of the area of experimentation suitable for use in school?]. *Biologie Lehren und Lernen – Zeitschrift für Didaktik der Biologie*, *19*(1), 25–36.

Baur, A. (2016). Problempunkte von Schülerinnen und Schülern bei der experimentellen Methode: Manuelle Fehler, methodische Fehler und Fehlkonzepte [Pupils‘ problems in the application of the experimental method: Manual and metodical mistakes, misconceptions, and pupils' idiosyncratic approaches]. In U. Gebhard & M. Hammann (Ed.), *Lehr- und Lernforschung in der Biologiedidaktik 7* (p. 191-205) [Research in biology educational science]. Innsbruck: Studien Verlag.

Baur, A. (2018). Fehler, Fehlkonzepte und spezifische Vorgehensweisen von Schülerinnen und Schülern beim Experimentieren: Ergebnisse einer videogestützten Beobachtung [Mistakes, misconceptions, and pupils' idiosyncratic approaches to experimentation: Findings from an observation]. *Zeitschrift für Didaktik der Naturwissenschaften*, *24*(1), 115–129.

Bell, R. L., Blair, L. M., Crawford, B. A. & Lederman, N. G. (2003). Just do it? Impact of a science apprenticeship program on high school students' understanding of the nature of science and scientific inquiry. *Journal of Research in Science Teaching*, *40*(5), 487–509.

Boaventura, D., Faria, C., Chagas, I. & Galvão, C. (2013). Promoting science outdoor activities for elementary school children: Contributions from a research laboratory. *International Journal of Science Education*, *35*(5), 796–814.

Bonett, D. G. & Price R. M. (2007). Statistical inference for generalized Yule Coefficients in 2 x 2 contingency tables. *Sociological Methods & Research*, *35*(3), 429–446.

Bybee, R. W. (1997). *Achieving scientific literacy: From purposes to practices. Portsmouth*, NH: Heinemann.

Carey, S., Evants, R., Honda, M., Jay, E. & Unger, C. (1989). An experiment is when you try it and see if it works: A study of grade 7 students' understanding of the construction of scientific knowledge. *International Journal of Science Education*, *11*(special issue), 514–529.

Chen, Z. & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the Control of Variables Strategy. *Child Development*, *70*(5), 1098–1120.

Chinn, C.A. & Brewer, W.F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. *Journal of Research in Science Teaching*, *35*(6), 623–654.

Cuccio-Schirripa, S. & Steiner, H. E. (2000). Enhancement and analysis of science question level for middle school students. *Journal of Research in Science Teaching*, *37*(2), 210–224.

DfE [Department for Education] (2014). *The national curriculum in England: Key stages 3 and 4 framework documents.*

ELAN (Version 5.2) [Computer software]. (2018). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

Emden, M. (2011). *Prozessorientierte Leistungsmessung des naturwissenschaftlich-experimentellen Arbeitens: Eine vergleichende Studie zu Diagnoseinstrumenten zu Beginn der Sekundarstufe I* [Process-oriented performance measurement of scientific-experimental work: A comparative study of diagnostic tools at the beginning of lower secondary education]. Berlin: Logos Verlag.

Friedler, Y., Nachmias, R. & Linn, M. C. (1990). Learning scientific reasoning skills in microcomputer-based laboratories. *Journal of Research in Science Teaching*, *27*(2), 173–192.

Gauld, C. (1986). Models, meters and memory. *Research in Science Education,* (16), 49–54.

Germann, P.J., Aram, R. & Burke, G. (1996). Identifying patterns and relationships among the responses of seventh-grade students to the science process skill of designing experiments. *Journal of Research in Science Teaching*, *33*(1), 79–99.

Ginsburg, H. P. & Opper, S. (1998). *Piagets Theorie der geistigen Entwicklung* (8., völlig überarbeitete und ergänzte Auflage) [Piaget's theory of kognitive development]. Stuttgart: Klett-Cotta Verlag.

doi: 10.11576/zdb-3576

Glaser, R., Schauble, L., Raghavan, K. & Zeitz, C. (1992). Scientific reasoning across different domains. In E. de Corte, M.C. Linn, H. Mandl & L. Verschaffel (Ed.), *Computer-based learning environments and problem solving* (p. 345-371). Berlin: Springer-Verlag.

Hammann, M., Phan, T.T.H., Ehmer, M. & Bayrhuber, H. (2006). Fehlerfrei Experimentieren [Experiment correctly]. *Der mathematische und naturwissenschaftliche Unterricht*, *59*(5), 292–299.

Hammann, M., Phan, T.T.H., Ehmer, M. & Grimm, T. (2008). Assessing pupils' skills in experimentation. *Journal of Biological Education*, *42*(2), 66–72.

Hodson, D. (2014). Learning Science, Learning about Science, Doing Science: Different goals demand different learning methods. *International Journal of Science Education*, *36*(15), 2534–2553.

Hofstein, A., Navon, O., Kipnis, M. & Mamlok-Naaman, R. (2005). Developing students' ability to ask more and better questions resulting from inquiry-type chemistry laboratories. *Journal of Research in Science Teaching*, *42*(7), 791–806.

Kanari, Z. & Millar, R. (2004). Reasoning from data: How students collect and interpret data in science investigations. *Journal of Research in Science Teaching*, *41*(7), 748–769.

Kechel, J.-H. (2016). *Schülerschwierigkeiten beim eigenständigen Experimentieren: Eine qualitative Studie am Beispiel einer Experimentieraufgabe zum Hooke'schen Gesetz* [Difficulties of pupils when experimenting independently: A qualitative study using the example of an experimental task on Hooke's law]. Berlin: Logos Verlag.

Kirschner, P. A., Sweller, J. & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86.

Klahr, D., Fay, A.L. & Dunbar, K. (1993). Heuristics for scientific experimentation. A developmental study. *Cognitive Psychology*, *25*, 111–146.

Klayman, J. & Ha, Y.-w. (1989). Hypothesis testing in rule discovery: Strategy, structure, and content. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 596–604.

KMK [Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland]. (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss* [Educational standards in biology secondary education I]. München: Luchterhand.

Koerber, S., Mayer, D., Osterhaus, C., Schwippert, K. & Sodian, B. (2014). The development of scientific thinking in elementary school: A comprehensive inventory. *Child Development*, *86*(1), 327–336.

Krefting, L. (1991). Rigor in qualitative research: The assessment of trustworthiness. *The American Journal of Occupational Therapy*, *45*(3), 214–222.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, *96*(4), 674–689.

Kuhn, D., & Dean, D. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science*, *16*(11), 866–870.

Lubben, F. & Millar, R. (1996). Children's ideas about the reliability of experimental data. *International Journal of Science Education*, *18*(8), 955–968.

Ludbrook, J. (2008). Analysis of 2 x 2 tables of frequencies: Matching test on experimental design. *International Journal of Epidemiology*, *37*, 1430–1435.

Ludwig, T., Priemer, B. & Lewalter, D. (2018). Decision-making in uncertainty-infused learning situations with experiments in physic classes. In M. A. Sorto, A. White, & L. Guyot (Ed.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg: The Netherlands: International Statistical Institute.

Ludwig, T., Priemer, B. & Lewalter, D. (2019). Assessing secondary school students‘ justifications for supporting or rejecting a scientific hypothesis in the physics lab. *Research in Science Education* [published online 01 June 2019].

Masnick, A.M. & Klahr, D. (2003). Error matters: An initial exploration of elementary school children's understanding of experimental error. *Journal of Cognition and Development*, *4*(1), 67–98.

Mayring, P. (2008). *Qualitative Inhaltsanalyse*: *Grundlagen und Techniken* [Qualitative content analysis: Basics and techniques]. Weinheim: Beltz Verlag.

doi: 10.11576/zdb-3576

Meier, M. & Mayer, M. (2012). Experimentierkompetenz praktisch erfassen: Entwicklung und Validierung eines anwendungsbezogenen Aufgabendesigns [Diagnosing experimental competence: Development and validation of an application-oriented task design]. In U. Harms, & F.X. Bogner (Ed.), *Lehr- und Lernforschung in der Biologiedidaktik* (p. 81-98) [Research in biology educational science]. Innsbruck: Studien Verlag.

Metcalfe, J. (2017). Learning from Errors. *Annual Review of Psychology*, *68*, 465–489.

MKJS [Ministerium für Kultus, Jugend und Sport Baden-Württemberg]. (2004). Bildungsplan 2004 Realschule [Educational plan 2004 middle school].

Neber, H. & Anton, M.A. (2008). Förderung präexperimenteller epistemischer Aktivitäten im Chemieunterricht [Fostering of pre-experimental epistemic activities in chemistry lessons]. *Zeitschrift für Pädagogische Psychologie*, *22*(2), 143–150.

NRC [National Research Council]. (1996). *National science education standards*. Washington, D.C.: National Academy Press.

NRC [National Research Council]. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, D.C: National Acadamy Press.

OECD [Organisation for Econonomic Co-Operation and Development]. (2017). *PISA 2015 assessment and analytical framework*. Paris: OECD.

Oser, F., Hascher, T. & Spychinger, M. (1999). Lernen aus Fehlern: Zur Psychologie des „negativen“ Wissens [Learning from errors: Psychology of negative knowledge]. In W. Althof (Ed.), *Fehlerwelten: Vom Fehlermachen und Lernen aus Fehlern: Beiträge und Nachträge zu einem interdisziplinären Symposium aus Anlass des 60. Geburtstags von Fritz Oser* [Error worlds: Making errors and learning from errors: Contributions and supplements to an interdisciplinary symposium on the occasion of the 60th birthday of Fritz Oser] (S. 11–41). Opladen: Leske und Budrich Verlag.

Ruiz-Primo, M.A. & Shavelson, R.J. (1996). Rhetoric and reality in science performance assessments: An update. *Journal of Research in Science Teaching, 33*(10), 1045–1063.

Schauble, L., Klopfer, L.E. & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching*, *28*(9), 859–882.

Schecker, H. & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz [Modelling of scientific competence]. *Zeitschrift für Didaktik der Naturwissenschaften*, *12*, 45–66.

Schreiber, N. (2012). *Diagnostik experimenteller Kompetenz*: *Validierung technologiegestützter Testverfahren im Rahmen eines Kompetenzstrukturmodells* [Diagnostics of experimental competence: Validation of technology-supported test procedures within the framework of a competence structure model]. Berlin: Logos Verlag.

Schumacher, R. (2008). Der produktive Umgang mit Fehlern: Fehler als Lerngelegenheit und Orientierungshilfe [The productive use of errors: Errors as learning opportunity and orientation aid]. In R. Caspary (Ed.), *Nur wer Fehler macht, kommt weiter: Wege zu einer neuen Lernkultur* (p. 49-72) [Only those who make errors will get ahead: Ways to new learning culture]. Freiburg i. B.: Herder.

Schwichow, M., Croker, S., Zimmerman, C., Höffler, T. & Härtig, H. (2016). Teaching the control-of-variables strategy: A meta-analysis. *Developmental Review. 39*(1), 37–63.

Shavelson, R.J., Ruiz-Primo, M.A. & Wiley, E.W. (1999). Note on sources of sampling variability in science performance assessments. *Journal of Educational Measurement*, *36*(1), 61–71.

Siler, S. A. & Klahr, D. (2012). Detecting, classifying, and remediating: Children’s explicit and implicit misconceptions about experimental design. In R. W. Proctor & E. J. Capaldi (Ed.), *Psychology of Science* (p. 137–180). Oxford University Press.

Wahser, I. & Sumfleth, E. (2008). Training experimenteller Arbeitsweisen zur Unterstützung kooperativer Kleingruppenarbeit im Fach Chemie [Training of experimental working to support cooperative small group work in chemistry]. *Zeitschrift für Didaktik der Naturwissenschaften, 14*, 219–241.

Wason, P.C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*(3), 129–140.

Watson, B. & Konicek, R. (1990). Teaching for conceptual change: Confronting children's experience. *Phi Delta Kappan*, 680–684.

doi: 10.11576/zdb-3576

Weinert, F.E. (1999). Aus Fehlern lernen und Fehler vermeiden lernen [Learning from errors and learn to prevent errors]. In Althof, W. (Ed.), *Fehlerwelten: Vom Fehlermachen und Lernen aus Fehlern: Beiträge und Nachträge zu einem interdisziplinären Symposium aus Anlass des 60. Geburtstags von Fritz Oser* [Error worlds: Making errors and learning from errors: Contributions and supplements to an interdisciplinary symposium on the occasion of the 60th birthday of Fritz Oser] (S. 101–110). Opladen: Leske u. Budrich.

Zhai, J., Jocz, J. A. & Tan, A.-L. (2013). 'Am I Like a Scientist?': Primary children's images of doing science in school. *International Journal of Science Education, 36*(4), 553–576.

doi: 10.11576/zdb-3576

Kontakt

Armin Baur
Pädagogische Hochschule Heidelberg
University of Education Heidelberg
Biology and Biology Education
Im Neuenheimer Feld 561
69120 Heidelberg
E-Mail: baur@ph-heidelberg.de
https://orcid.org/0000-0001-5950-7613

Zitationshinweis:
Baur, A. (2021). Errors made by 5th-, 6th-, and 9th-graders when planning and performing experiments: Results of video-based comparisons. *Zeitschrift für Didaktik der Biologie (ZDB) – Biologie Lehren und Lernen*, *25*, 45-63. doi: 10.11576/zdb-3576

Veröffentlicht: 18.07.2021