

Appendix 1: Blueprint for the development of a scientific reasoning questionnaire (based on Krell 2018)

Theoretical background

Within scientific reasoning (SR), cognitive strategies include forming hypotheses, testing hypotheses, and analyzing data (Klahr and Dunbar 1988). Hence, we distinguish between three SR skills (Krell 2018; Hammann et al. 2008; Mannel, Walpuski, and Sumfleth 2015):

1. *Forming hypotheses*: Participants have to identify the hypothesis of a given research design.
2. *Testing hypotheses*: Participants have to identify the research design that is needed to test a given hypothesis.
3. *Analyzing data*: Participants have to identify the conclusion that can be validly drawn on the basis of results obtained from a given research design.

Item development

The item development accounted for three item features—SR skills, item complexity, and research context. Two other item features (i.e., text complexity and specialist terms) were also considered in our analysis (see the main manuscript).

With regard to the three SR skills, we accounted for this item feature by developing items that were concerned with either forming hypotheses, testing hypotheses, or analyzing data as described above.

We also accounted for the number of independent variables as another item feature called item complexity: items with lower (one independent variable) and higher complexity (two independent variables) were developed. For this SR questionnaire, we considered three independent variables (A, B, C), of which only one or two varied. The independent variables (IVs) and the dependent variable (DV) could either be manifest (+) or not manifest (–).

Please consider the following blueprints, specific to the three SR skills, in Tables A1–3 of this Appendix 1.

Furthermore, the SR skills are contextualized in different research topics (e.g., wildlife ecology). Citizen science project designers might want to look for research topics that stem from actual research in previous projects. We provide an example of how we adapted the blueprint for one research context (urban wildlife ecology) in the following section.

Example of item development

In this example, we describe the development of an item in our SR questionnaire regarding three item features: SR skills, item complexity, and research context. In our SRQ, Item 6 concerned the SR skill of analyzing data, the number of independent variables was two (i.e., high item complexity), and the research context was urban wildlife ecology. We present the developed item in Figure 1 in the main manuscript and in Appendix 2.

Identification of relevant research

For the contextualization of items, we suggest that researchers and practitioners identify authentic studies from the respective discipline that fit to their CS project context. One co-author recently published a study on wild boars' behavior in urban environments (Stillfried et al. 2017). They were able to show that urban boars (in contrast to rural boars) were more tolerant of human-caused disturbances by measuring the flight distance. This research was highly relevant for our CS project that concerned the urban ecology of wildlife. Wild boars in particular are a common concern in the city of Berlin and are covered in local media due to human-wildlife conflicts.

Review of the research

For the research review, we suggest forming a team with members who have expertise in the scientific discipline and in educational assessment. In our study, three of the co-authors, two with expertise in urban ecology and one with expertise in educational assessment, discussed this research in several rounds to identify the study's key variables in relation to wild boars' flight distance (Stillfried et al. 2017). Based on human encounters with wild boars, we

decided to focus on the phenomenon of how humans actively approaching wild boars while making noise affect the flight distance of wild boars from either urban or rural areas. In the study, we identified the humans' active approach toward the animal as a variable that was held constant. The two variables that varied in the research setup were the kind of wild boars (rural vs. urban boars) and the noises that humans made during their active approach (silence vs. noises). The dependent variable was the measured flight distance, which is when a boar runs away when a person actively approaches.

Adaptation of the blueprint

For the blueprint adaptation, we suggest accounting for the item features—SR skills, number of independent variables, and research context—in the development of every item in the SR questionnaire. We now describe the adaptation process for one item that concerned the SR skill of analyzing data (i.e., item feature: SR skills) with two variables (i.e., item feature: item complexity) in urban wildlife ecology (i.e., item feature: research context; see Table A3 on the left side for the blueprint). First, we introduced the research setup for the investigated phenomenon (i.e., wild boars' behavior) to the participant in the text stem. To introduce the research setup, we described for each independent variable (i.e., active approach, the habitat of wild boars, and noises) whether they were manifest (+) or not manifest (–) in the four setups. In a figure, we then summarized the four setups and provided information on each independent variable. Second, we presented the observations obtained by the researchers. In the item, we described the setup in which the researchers observed that the wild boars ran away later (i.e., the wild boars had a lower flight distance). Third, we asked the participant to decide which valid conclusion scientists could draw based on these observations. Fourth, we provided the participant with four answer options, of which only one option was correct. The answer options provided possible conclusions that the scientists might draw. The conclusions on the observed behavior were based on different combinations of the independent variables. A valid conclusion could be drawn for only one combination of the independent variables in the answer options. The participant should recognize this combination and tick the respective answer option.

References

- Hammann, M, Phan, TTH, Ehmer, M and Grimm, T. 2008. Assessing pupils' skills in experimentation. *Journal of Biological Education*, 42(2): 66–72.
- Klahr, D and Dunbar, K. 1988. Dual Space Search During Scientific Reasoning. *Cognitive Science*, 12(1): 1–48.
- Krell, M. 2018. Schwierigkeitserzeugende Aufgabenmerkmale bei Multiple-Choice-Aufgaben zur Experimentierkompetenz im Biologieunterricht: Eine Replikationsstudie [Difficulty-generating task characteristics in multiple-choice tasks for experimental competence in biology teaching: A replication study]. *Zeitschrift für Didaktik der Naturwissenschaften*, 24(1):1-15.
- Mannel, S, Walpuski, M and Sumfleth, E. 2015. Erkenntnisgewinnung: Schülerkompetenzen zu Beginn der Jahrgangsstufe 5 im naturwissenschaftlichen Anfangsunterricht [Acquirement of Knowledge: Students' Competencies at the Beginning of Secondary Schooling in the Natural Sciences (Grade 5)]. *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1): 99–110.
- Stillfried, M, Gras, P, Börner, K, Göritz, F, Painer, J and Röllig, K, *et al.* 2017. Secrets of Success in a Landscape of Fear: Urban Wild Boar Adjust Risk Perception and Tolerate Disturbance. *Frontiers in Ecology and Evolution*, 5: 683.

Table A1: Blueprint for developing items on *forming hypotheses* (item feature: SR skills) while varying either two independent variables (high item complexity) or one independent variable (low item complexity).

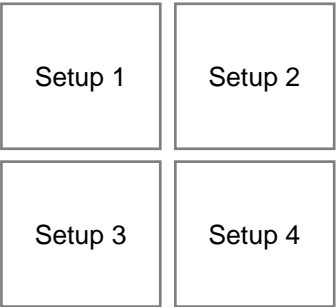

High item complexity	Low item complexity
<p>Item stem</p> <p>[Person X] conducts an investigation on the [phenomenon Y].</p> <p>[Person X] uses the following four setups (see figure):</p> <p>Setup 1: A+, B+, C+</p> <p>Setup 2: A+, B–, C+</p> <p>Setup 3: A+, B+, C–</p> <p>Setup 4: A+, B–, C–</p> 	<p>Item stem</p> <p>[Person X] conducts an investigation on the [phenomenon Y].</p> <p>[Person X] uses the following two setups (see figure):</p> <p>Setup 1: A+, B+, C+</p> <p>Setup 2: A+, B–, C+</p> 
<p>Item</p> <p>What hypothesis can [Person X] test with this investigation?</p> <p><input type="checkbox"/> From A+ and B+ follows DV+</p> <p><input type="checkbox"/> From A+ and C+ follows DV+</p> <p><input type="checkbox"/> From B+ and C+ follows DV+</p> <p><input type="checkbox"/> From A+, B+, and C+ follows DV+</p>	<p>Item</p> <p>What hypothesis can [Person X] test with this investigation?</p> <p><input type="checkbox"/> From A+ follows DV+</p> <p><input type="checkbox"/> From B+ follows DV+</p> <p><input type="checkbox"/> From C+ follows DV+</p> <p><input type="checkbox"/> From A+, B+, and C+ follows DV+</p>

Table A2: Blueprint for developing items on *testing hypotheses* (item feature: SR skills) while varying either two independent variables (high item complexity) or one independent variable (low item complexity).

High item complexity	Low item complexity
<p>Item stem</p> <p>[Person X] assumes that B and C influence AV+.</p> <p>[Person X] creates the setups A+, B+, C+; A+, B-, C-; and A+, B+, C-.</p> <p>[Person X] needs an additional setup.</p> <div> <div>Setup 1</div> <div>Setup 2</div> <div>Setup 3</div> </div> <p>Item</p> <p>Which additional setup is needed?</p> <p><input type="checkbox"/> Setup 4a: A-, B+, C+ Setup 4a</p> <p><input type="checkbox"/> Setup 4b: A+, B-, C+ Setup 4b</p> <p><input type="checkbox"/> Setup 4c: A-, B+, C- Setup 4c</p> <p><input type="checkbox"/> Setup 4d: A-, B-, C- Setup 4d</p>	<p>Item stem</p> <p>[Person X] assumes that B influences AV+.</p> <p>[Person X] creates the setup A+, B+, C+.</p> <p>[Person X] needs an additional setup.</p> <div> <div>Setup 1</div> </div> <p>Item</p> <p>Which additional setup is needed?</p> <p><input type="checkbox"/> Setup 2a: A-, B+, C+ Setup 2a</p> <p><input type="checkbox"/> Setup 2b: A+, B-, C+ Setup 2b</p> <p><input type="checkbox"/> Setup 2c: A+, B+, C- Setup 2c</p> <p><input type="checkbox"/> Setup 2d: A-, B-, C- Setup 2d</p>

Table A3: Blueprint for developing items on *analyzing data* (item feature: SR skills) while varying either two independent variables (high item complexity) or one independent variable (low item complexity).

High item complexity	Low item complexity
<p>Item stem</p> <p>[Person X] creates the following four setups:</p> <p>Setup 1: A+, B+, C+</p> <p>Setup 2: A+, B–, C+</p> <p>Setup 3: A+, B+, C–</p> <p>Setup 4: A+, B–, C–</p> <p>[Person X] recognizes DV+ only in Setup 1</p> <div data-bbox="293 981 632 1285"> </div> <p>(see figure).</p> <p>Item</p> <p>What conclusion can [Person X] draw?</p> <p><input type="checkbox"/> From A+ and B+ follows DV+</p> <p><input type="checkbox"/> From A+ and C+ follows DV+</p> <p><input type="checkbox"/> From B+ and C+ follows DV+</p> <p><input type="checkbox"/> From A+, B+, and C+ follows DV+</p>	<p>Item stem</p> <p>[Person X] creates the following two setups:</p> <p>Setup 1: A+, B+, C+</p> <p>Setup 2: A+, B–, C+</p> <p>[Person X] recognizes DV+ only in Setup 1 (see figure).</p> <div data-bbox="925 1008 1264 1155"> </div> <p>Item</p> <p>What conclusion can [Person X] draw?</p> <p><input type="checkbox"/> From A+ follows DV+</p> <p><input type="checkbox"/> From B+ follows DV+</p> <p><input type="checkbox"/> From C+ follows DV+</p> <p><input type="checkbox"/> From A+, B+, and C+ follows DV+</p>