Developing a Diagnostic Instrument for Scientific Giftedness in the Context of Design-Based Research (DBR)

Colin Peperkorn

Claas Wegner

Abstract

Identifying and fostering gifted students is crucial in educational science and psychology. Giftedness diagnostics must be based on profound domain-specific concepts and acknowledge the variety of talents to enable a successful individual education. Growing challenges like digitalization, decarbonization, demographic changes, and pandemics increase the need for creative and productive professionals in STEM fields. Educational practitioners are expected to identify and foster these talents, confronted with the discourse between academic science and psychology about giftedness. Due to the lack of domain-specific diagnostic instruments, the individual expression of giftedness is often neglected during diagnostic procedures in educational practice and general intelligence tests are used instead. To address this problem, this paper presents the development process of a domain-specific diagnostic instrument for scientific giftedness. To bridge the gap between theory and practice and incorporate knowledge from various fields, the development process is integrated into Design-Based Research (DBR). Therefore, we theoretically examine how the principles of DBR and test development can be connected. We present our research approach and check how the planned development process can lead to a test instrument suitable for practice and further knowledge about giftedness diagnostics. The project demonstrates that DBR is suitable for developing didactic interventions but can also lead to innovations in psychometrics and improve giftedness diagnostics in practice through additional quality criteria.

Keywords

Giftedness, STEM, test development, DBR

Introduction

Identifying and nurturing gifted students is crucial to educational and psychological research. A thorough understanding of domain-specific gifted behaviors is necessary to acknowledge the variety of talents and provide adequate education to these students (Peperkorn & Wegner, 2023a; Renzulli & Reis, 2021). There is a growing demand for creative and productive professionals in STEM fields (Demary et al., 2021), and the fostering of scientific talents is of increasing interest due to various global challenges such as digitalization, decarbonization, demographic shifts, and pandemics. In Germany, about 496,50 positions in STEM professions are vacant (Anger et al., 2023). Pedagogical diagnostics represent an essential prerequisite for successful gifted education. It serves as a basis for sound decisions on fostering gifted students (Fischer & Fischer-Ontrup, 2022). Teachers face diagnosing their students' giftedness in school and considering subject-specific characteristics.

Giftedness diagnostics in school

There is an ongoing discourse between psychological and educational research on giftedness (Rost, 2013). The International Panel

Citation: Peperkorn, Colin & Wegner, Claas (2024). Developing a Diagnostic Instrument for Scientific Giftedness in the Context of Design-Based Research (DBR). *Global Education Review*, 11 (2), 25-45.

for Experts of Gifted Education defines giftedness as follows:

> Giftedness is generally understood to be the overall capacity for achievement. More specifically, giftedness refers to the individual's level of development of performance-related potential, i.e., those preconditions which, with the appropriate disposition and long-term, systematic stimulation, support, and encouragement, enable the individual to act in a meaningful and responsible manner and to carry out demanding activities in areas considered valuable in the respective culture. (iPEGE, 2009, p. 17)

Accordingly, giftedness develops through cognitive preconditions, appropriate environmental factors, and targeted fostering. This pedagogical approach emphasizes creating a supportive learning environment for gifted individuals (e.g., Heller et al., 2005; Renzulli, 1978). Intelligence is described as "a part of giftedness" (iPEGE, 2009, p. 18) and describes a high intellectual potential that enables a person to perform well under given conditions (Heller, 1976; Rost, 2015). However, psychologists have criticized pedagogy for not using the term intelligence in the context of giftedness research (e.g., Rost, 2015). Teachers and other educational practitioners are often caught between psychological and pedagogical approaches in diagnosing giftedness. While teacher assessments are commonly used, they are often inaccurate (Spinath, 2005). They can be influenced by a student's academic performance (Machts et al., 2016), making it challenging to diagnose underachievers (Rost & Hanses, 1997). Additionally, no consensus exists on whether checklists can significantly improve teachers' assessments (Jarosewich et al., 2002; but see Renzulli, 2009). Standardized procedures are utilized to diagnose giftedness, typically through general intelligence tests, due to the close connection between intelligence and giftedness (Peperkorn & Wegner, 2020). This approach is practically and methodologically sound, as intelligence tests are well-validated and easy to administer. Moreover, general intelligence is among the best-researched psychological constructs with a high psychometric quality (Rost, 2013). One limitation of intelligence tests in education is the requirement for trained psychological personnel to administer them, which poses difficulties in implementing them into routine school activities. It is not proven to what extent the abstract abilities measured in an IQ test allow statements about success in STEM fields (Sternberg, 2018). Research findings indicate a positive correlation between scientific reasoning skills across various content areas. However, the reasoning skills evaluated in intelligence tests differ from those related to scientific reasoning (Sternberg et al., 2019). According to Bergold (2014), teachers face challenges distinguishing between domains of giftedness when identifying gifted students. Consequently, subject specificity is often neglected in the context of giftedness diagnostics at schools. Additionally, their diagnostic performance is lower regarding highly intelligent students. A possible solution to this issue is incorporating subject-specific achievement tests that evaluate skills like scientific reasoning. However, these tests are rarely designed for school use (Opitz et al., 2017; Peperkorn & Wegner, 2023a). One way to improve giftedness diagnostics in STEM in educational practice is to develop suitable tests that measure cognitive abilities and subject-specific competencies. This paper aims to introduce test instruments designed for diagnosing scientific giftedness that can be utilized in school settings using Design-Based Research (DBR).

What is DBR?

The subsequent section expounds on the theoretical basis of the DBR approach, the test development process, and how they can be effectively combined. Problems from educational practice are often insufficiently understood in research or integrated into the research process (Reinmann, 2022). The DBR approach is a bridge between theory and practice and is becoming increasingly relevant for educational research (Anderson & Shattuck, 2012). The term emerged from the preliminary work of Brown (1992), who understood design experiments as a method that could fill research gaps in educational research. This meant that learning phenomena were no longer to be researched solely through laboratory experiments, large-scale studies, or ethnographies. Instead, central research questions of pedagogy should be explored in real situations, and the research design should be incorporated into the scientific process (Reinmann, 2005). In this way, insights could be gained directly from practice, and reality could be mapped in the best possible way to establish sustainable innovations in education. The approach is not limited to any specific method but forms a framework combining various methodologies to gain the best possible insights into educational practice (Design-Based Research Collective, 2003). Therefore, DBR has characteristics that go beyond usual measurement and quality criteria.

Reinmann (2005) and Euler (2014) summarized the diverse descriptions of the approach and presented its characteristics. The initial attribute pertains to the concentration of the research methodology on the development procedure. Implementing teaching-learning concepts is realized by working directly with people from practice. Bereiter (2002) argues that "design researchers, by contrast, are trying to make something happen, and this frequently means crossing the boundary between observer and actor" (p. 326). Another characteristic of DBR is its practice-oriented objective. Developing theories and interventions directly applicable to educational practice is crucial in solving related problems. Integrating established scientific theories with research findings from specific educational settings is an important factor in achieving sustained improvements in education (Shavelson et al., 2003). This motivation forms another feature of DBR (Reinmann, 2005). Researchers committed to the approach want to improve educational practice and directly influence it through their work. They aim to bridge the gap between theory and practice by continually imagining innovations and testing their value directly in practice (Anderson & Shattuck, 2012). Finally, implementing the above principles takes place in an iterative process consisting of several phases, which can be seen as another attribute and leading credo of DBR (Bereiter, 2002). There are numerous models of DBR, differing in the exact number of phases and intermediate steps (Euler, 2014). A well-known model was established by McKenney & Reeves (2018), in which the iterative process is divided into the phases "Needs and context analysis, Design development, and formative evaluation, and semi-summative evaluation" (p. 15). The initial stage of a DBR study involves defining the problem that needs to be resolved and determining the necessity of conducting the study. The methodological freedom allows a wide variety of approaches in this first step. Various methodologies can be employed to gather information about a specific field of practice. These include conducting on-site visits, interviewing knowledgeable practitioners or experts, scrutinizing public literature or curricula, or analyzing large-scale data. Once the problem has been clearly defined, the current state of the research needs to be identified so that previous findings can be incorporated into the study's design. Most authors recommend a systematic review of literature or meta-analysis to describe the state of research as precisely as possible. In this way, a sound theoretical basis for the development of innovations can be achieved, which, according to Euler (2014), describes another specific attribute of DBR. After conducting context analyses and preparing underlying theories, the next step is to develop initial prototypes. These prototypes are then used and evaluated in recurring processes (Reinmann, 2005). Using prototypes facilitates a perpetual enhancement of the designs, enabling the investigation of design principles and a comprehensive theory (Shavelson et al., 2003). Different types of theory building are described. First, *domain theories* can be established that bring together either knowledge about the practice context (*context theories*), e.g., explanations of the emergence of certain practice problems, or describe intended outcomes (*outcomes theories*) that are desired to be achieved by a prototype (Edelson, 2002). Furthermore, *design frameworks* that combine guidelines for creating interventions or didactic scenarios and provide generalized design templates can be developed (Edelson, 2002; Reinmann, 2005). Finally, *design methodologies* can be established, summarizing guidelines for the design process. The phase of "design, development, and formative evaluation" (McKenney & Reeves, 2018, p. 15) comprises repeated micro cycles that include a growing number of individuals, such as subjects, stakeholders, and experts, who are involved in the study. Following the development and re-design phase, a (semi) summative evaluation is conducted. This evaluation allows for a new initial problem analysis (Euler, 2014; Reinmann, 2005). The final summative analysis "make[s] it possible to anticipate outcomes in future designs" (Cobb et al., 2003, p. 13). In addition, new research questions can be derived from the process that may only have been discovered through practice-based research. The elaborations on DBR mainly refer to the developments of innovations for educational practice in the form of interventions. Since DBR can follow basic scientific principles (Fischer et al., 2003; Reinmann, 2022; Shavelson et al., 2003) and its framework allows for great methodological diversity, a transfer of the approach to other research fields, such as test development, is quite conceivable. For this to be successful, it must be examined whether the basic principles of test development can be incorporated into the phases of DBR.

DBR and Test Development

In the following, we will examine to what extent the DBR as a research direction is suitable for psychometric test development. For this purpose, basic test development guidelines (e.g., Lane et al., 2015; Irwing & Hughes, 2018) are applied to DBR and classified according to its phases (see Table 1).

Table 1: Stages of test development (Irwing & Hughes, 2018).

Stages and substages	
1	Construct definition, specification of test need, test structure.
2	Overall planning.
3	Item development
	a. Construct definition.
	b. Item generation: theory versus sampling.
	c. Item review.
	d. Piloting of items.
4	Scale Construction – factor analysis and Item Response Theory (IRT)
5	Reliability.
6	Validation.
7	Test scoring and norming.
8	Test specification.
9	Implementation and testing.
10	Technical Manual

The elaboration of a *construct definition*, the *specification of the test need*, and the development of a *test structure* can be placed in the first two phases of DBR, where the initial problem is specified, and corresponding literature is evaluated (e.g., Euler, 2014). Here, the first connections between the two research directions emerge. Irwing & Hughes (2018) describe: "The motivation for test development often stems from a practical concern" (p. 5). In the framework of DBR, test development can benefit, as the test structure can be adapted precisely to the observed conditions of the practice field. Different test methods (e.g., multiple choice, performance-based, open format) can be examined for their applicability in specific contexts, and an optimal fit of the instrument can be realized. Close collaboration with practitioners allows the test instrument's precise requirement to be determined for a specific setting. This ensures accuracy and effectiveness in its implementation. "In short, for a test to address a market need, it should be both technically sound (in terms of theoretical grounding and psychometric properties) and practically useful" (Irwing & Hughes, 2018, p. 7). The construct definition and the determination of the test components are carried out, for example, via meta-analyses. The aim is to obtain assessments of specific parameters that are adjusted for measurement errors or measurement artifacts and allow statements about psychological relationships, such as correlations to later career choices (Wetterich & Plänitz, 2021). An essential requirement for establishing such a foundation is a researched database. When developing tests on so far undefined constructs, the underlying theory should be systematically recorded and analyzed through a literature review. In test development, the *overall planning* describes answering various questions about the scope, format, structure, and evaluation of the test instrument, as well as initial considerations about the procedure for piloting or validation (Irwing & Hughes, 2018). This phase and the first steps of the *Item Development* phase can still be assigned to the Design phase in the context of DBR, in which several correction cycles can already be conducted without testing the prototype in the field (Design-Based Research Collective, 2003; Reinmann, 2020). Toward the end of the item development phase, the first field tests or test runs are conducted to determine item characteristics. In this phase, additional methods such as think-aloud tests can be added, and the application of possible scoring rubrics can be tested to improve the test instrument's construction further (Lane et al., 2015). These steps can be assigned to the beginning of the first micro cycle of the testing phase in DBR (Reinmann, 2020). After the first data on the developed test instrument have been collected, they should be checked for their *scale structure* and *psychometric properties*. In most cases, a combination of confirmatory *factor analysis* (CFA; Brown, 2015) and *item response theory* (IRT; e.g., Bock & Gibbons, 2021) is recommended (Irwing & Hughes, 2018). The *reliability* check can also be assigned to the development phase in DBR, as minor changes can always be made to the design of the test instrument to improve it (Reinmann, 2020). McKenney & Reeves (2018) describe that during the *design, development, and formative evaluation* phase in DBR, more and more test subjects must be involved. Like in test development procedures, the *validation* process loops larger and larger until enough data are available for the *test norming* (Lane et al., 2015). Concluding the underlying theory is crucial when analyzing and generalizing results (Reinmann, 2020; Shavelson et al., 2003). These conclusions are based on the data obtained, leading to a new phase of problem definition. In this way, another DBR cycle can be started. The *test* specification, *implementation and testing*, and the preparation of a *test manual* are all part of a transfer of previously researched design principles to the further use of the Instrument and, therefore, describe steps of the summative analysis phase (Cobb et al., 2003).

The explanations confirm that test development processes can be theoretically applied well to DBR, meeting and even extending scientific criteria. Design criteria can be incorporated into the developments in addition to the usual quality criteria. In this way, both an improvement of educational practice and insights into designing suitable test instruments for practitioners can be generated. Furthermore, valuable insights for test development and the overall improvement of psychometric measurement are yielded (Reinmann, 2022). As already described, there are hardly any psychometric tests in giftedness diagnostics designed or validated for use in schools (Peperkorn & Wegner, 2023a).

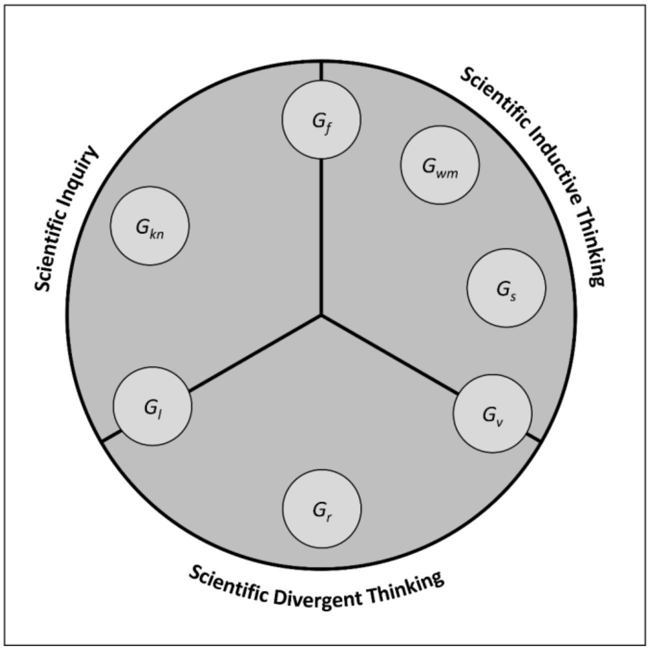
The development of a test instrument for the diagnosis of scientific giftedness

In the following, a project for the development of a test instrument for the diagnosis of scientific giftedness is presented. The concept of giftedness pertains to an individual's complete range of abilities. In particular, " giftedness refers to the individual's level of development of performance-related potential, [...] enable the individual to act in a meaningful and responsible manner and to carry out demanding activities in areas considered valuable in the respective culture." (iPEGE, 2009, p. 17). Scientific giftedness describes a person's potential to perform these demanding activities in the field of the natural sciences. The instrument is designed in a DBR process to be usable in educational contexts. Both criteria of DBR and test development are pursued and combined.

Needs and context analysis

A needs analysis must be conducted in test development processes and DBR (e.g., Edelson, 2002; Euler, 2014; Irwing & Hughes, 2018). Giftedness diagnostics in schools are rarely conducted in a subject-specific manner (Peperkorn & Wegner, 2020). Instead, general assessments by teachers, intelligence and achievement tests are used. Standardized tests on subject-specific skills, like *Scientific Reasoning,* are rarely used (Opitz et al., 2017), as they are not adapted for use in school. Because scientific giftedness is not solely reliant on inherent abilities, several diagnostically significant factors may go unnoticed (Peperkorn & Wegner, 2021; Peperkorn & Wegner, 2023a; Peperkorn & Wegner, 2023b; Wegner, 2014). Teachers' assessments often reach their limits in giftedness diagnostics and there is an ongoing discourse about their value (Machts et al., 2016; Rost & Hanses, 1997). Even supports like adapting theories of giftedness to different subjects (iPEGE, 2014) or providing checklists do not improve educational decisions (Jarosewich et al., 2002; but see Renzulli et al., 2009). Hence, there is a need for standardized subject-specific test instruments designed for use in school. Furthermore, there is a lack of empirically based theory on scientific giftedness. Existing concepts describe *a priori* models that have not been empirically tested (Peperkorn & Wegner, 2023a). Finally, there is a need for action to close the discourse between psychology and education in giftedness diagnostics (e.g., Rost & Sparfeldt, 2017). Integrating psychometric test development into the DBR framework can promote the transfer between psychological and pedagogical principles. In the context of the presented project, a comprehensive review of literature on existing test instruments for diagnosing giftedness in STEM subjects was conducted to define the initial problem and provide a sound theoretical basis for the planned test instrument (Peperkorn & Wegner, 2023a). Several problems with the existing instruments became apparent during the review. First, many of the test instruments are not suitable for use in practice because they are too time-consuming to administer and evaluate. Most studies utilized convenience sampling methods, resulting in pre-selected samples that complicate assertions regarding their applicability in practical settings. Furthermore, it was found that the development studies were mainly conducted with upper-school students and that hardly any instruments were developed for the school transition phase. However, this phase must be considered critically, as a loss of interest in STEM subjects can be measured (Gebhard et al., 2017). This decline should not be disregarded for students who may show giftedness in these areas. Moreover, the existing test instruments are not based on a uniform definition or model of scientific giftedness, which makes it very difficult to compare their results. Although studies describe scientific giftedness as multifactorial, few test instruments aim for a holistic measurement. Most instruments only assess partial areas of giftedness in a particular subject and assign the results to a multifactorial framework. To address the problem of the definitional disagreement and to provide a sound theoretical basis for the planned test instrument (Irwing & Hughes, 2018), a CHC-based analysis was conducted as part of the review (Mickley & Renner, 2019). In this process, the diagnostically relevant abilities assessed by the described test instruments were placed in the framework of the CHC theory (e.g., Schneider & McGrew, 2018). The diagnostically relevant abilities were classified into the broad ability domains Fluid Reasoning (G_f), Working Memory Capacity (G_{wm}), Processing speed (G_s), Visual Processing (G_v), Retrieval Fluency (G_r), Learning efficiency (G_l) and Domain-Specific Knowledge (G_{kn}; Flanagan & Dixon, 2013; Schneider & McGrew, 2018). Based on the CHC-based analysis, a hypothetical a priori measurement model was established and used as a theoretical framework for test development (see Figure 1). Three different subtests were derived from the results of the CHC-based analysis as an initial design solution (Edelson, 2002) to enable the measurement of relevant abilities.

Figure 1: Hypothetical measurement model for scientific giftedness



Note. A hypothetical measurement model was developed from the broad ability areas that serve as the foundation for giftedness in science (Peperkorn & Wegner, 2023a).

Prototype Design and Development

All three subtests were designed to be administered as a test battery or distributed over several lessons. Each test is designed to be completed within 20 minutes. This should allow flexible use in educational practice. The diagnostic instruments are not intended to determine status like gifted or non-gifted but to provide an individual impression of various relevant areas of giftedness in science, which helps teachers make adequate educational decisions. The aim is not to identify deficits but to make differentiated teaching offers based on the existing talents within the student group. For example, the tests might indicate that several students exhibit proficiency in experiment planning. Implementing experimentation phases, where students plan and decide independently, can benefit their development. Rost (2013) describes a fundamental problem in the measurement of multifactorial models of giftedness when the objective is to assign a status:

> If, as in Renzulli's model, three characteristics are defined as necessary components of giftedness and it is assumed that a gifted person must achieve a minimum percentile rank of PR = 90 in each characteristic (which, in the case of cognitive performance, corresponds to a measurement value of IQ = 90, for example), then the gifted person must be able to demonstrate a high level of ability (e.g., for cognitive ability, this corresponds to a measurement of IQ = 120), then with a moderate variable intercorrelation of r = .30, one needs an initial sample to be tested of at least 15,396 individuals to compose a group of 100 gifted individuals. (Rost, 2013, p. 238)

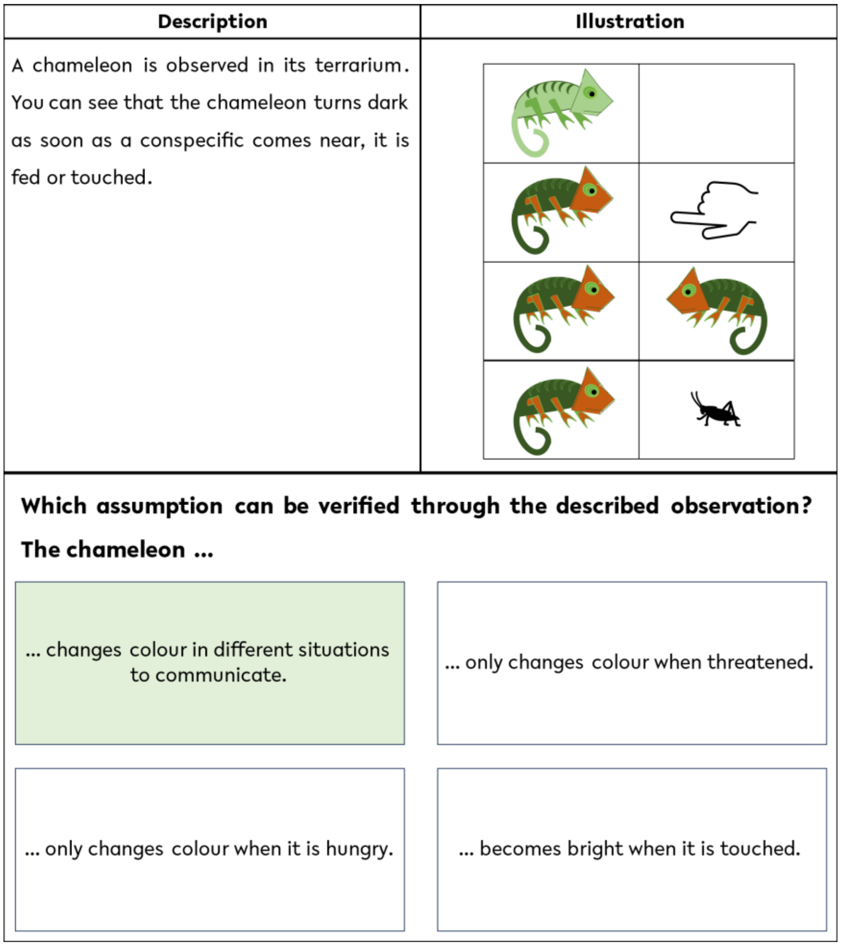
Therefore, the test instruments must be understood as a pedagogical tool to improve domain-specific diagnostics in school. In the following, the subtests and their conceptions are presented.

Scientific Inquiry

Scientific Inquiry "refers to the activities of students in which they develop knowledge and understanding of scientific ideas, as well as an understanding of how scientists study the natural world" (National Research Council, 1996, p. 23). Nowak et al. (2013) developed a model of S*cientific Inquiry* combining *Scientific Reasoning* and *Inquiry Methods*. *Scientific Reasoning* describes the skills *formulate hypotheses*, *plan and perform inquiries,* and *analyze and reflect* them. *Inquiry Methods* describes the skills *modelling*, *experimenting*, and *observing*, *comparing*, *and arranging*. By combining the two dimensions and their sub-skills, the authors defined nine "epistemological acts" (Nowak et al., 2013, p. 184). The test on *Scientific Inquiry* in the present project was derived from the broad ability domains G_l, G_{kn}, and G_f (Peperkorn & Wegner, 2023a). The central broad ability is G_{kn}, as the scientific method must be learned and is based on subject-specific knowledge. In addition, the ability area G_f was assigned because in scientific inquiry, "problems that cannot be solved by using previously learned habits, schema, and scripts" (Schneider & McGrew, 2018, p. 93) must also be solved. Finally, associative skills are needed while transferring learned scientific procedures to new situations or research areas, summarized under the broad ability G_l in CHC theory. It is plausible that the practice-oriented skills of scientific inquiry can be accurately measured by performance-based assessments (e.g., Alfaiz et al., 2020; Zimmerman et al., 2020). However, the daily school routine and the capacities of teachers do not offer the possibility to conduct these time-consuming tests. As the test instrument is intended to be time-efficient and usable in everyday school life, a digital multiple-choice format was chosen. The prototype consists of a maximum of 18 items. The instrument is planned to enable adaptive testing, which is why the item pool is continuously expanded. To minimize the influence of prior knowledge, each item begins with an introduction of an experiment, observation, or scientific model. The items aim to determine which hypothesis can be tested, suggest how an experiment should be set up, advise on how an observation should be structured, recommend how a model can be used to test a given hypothesis or suggest what valid conclusion can be drawn from given results (see Figure 2).

Several studies show that the skills of *Scientific Inquiry* can be measured sufficiently by this test format (Opitz et al., 2017). Based on Nowak et al. (2013), items were developed for each of the nine epistemological acts, which measure the skills of *Scientific Reasoning* and *Inquiry Methods* in combination. To account for variations in student reading proficiency, as the tests were developed for students from 3rd to 5th grade, every introduction information was recorded, and each student was given a chance to listen to it individually while taking the examination. In preliminary studies, this system has proven to be very helpful. The usage of the service differs among students, which suggests that it balances accessibility and increases validity. In contrast to the original instrument (Nowak et al., 2013), topics apart from the school curriculum were chosen to minimize the influence of different educational levels. Digital implementation offers the possibility to automatically evaluate results and save time.

Figure 2: Sample Item for measuring skills in Scientific Inquiry processes (observation / hypotheses)**.**



Description	**Illustration**
A chameleon is observed in its terrarium. You can see that the chameleon turns dark as soon as a conspecific comes near, it is fed or touched.	

Which assumption can be verified through the described observation?

The chameleon ...

... changes colour in different situations to communicate.

... only changes colour when threatened.

... only changes colour when it is hungry.

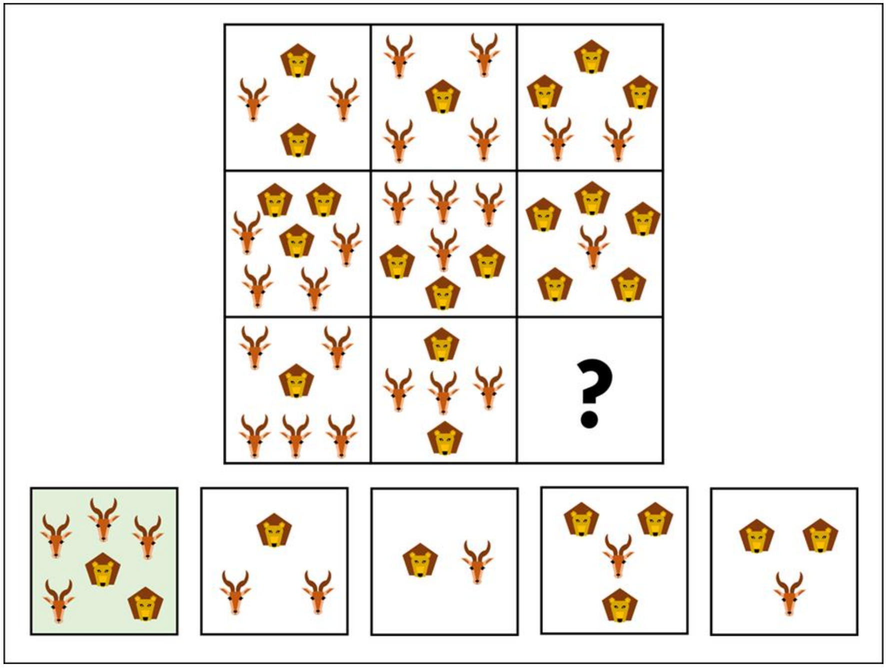
... becomes bright when it is touched.

Note. Items are carried out in a digital format. Items are carried out in a digital format. To select an answer, test takers need to click or tap on it. Correct answer is colored.

Scientific Inductive Thinking

Scientific Inductive Thinking describes a subtest that assesses abilities distinct from the learnable skills of *Scientific Inquiry*. Instead, the test is designed to measure cognitive abilities such as Fluid Reasoning (G_f), Working Memory Capacity (G_{wm}), Processing speed (G_s), Visual Processing (G_v), or Retrieval Fluency (G_r). These abilities were included in the diagnostic instrument as they are seen as a basis for expressing gifted behavior in STEM (Peperkorn & Wegner, 2023a). A matrices test tailored to science contexts was developed to measure these abilities. Matrices are a time-efficient way to measure fluid reasoning (Gf; Alfonso et al., 2005). Their capacity to evaluate the remaining broad ability factors is questionable (Ackerman et al., 2005). This question must be explored through comparative studies with other existing test instruments. When designing matrices, it is crucial to adhere to essential criteria such as creating distractors (Mittring & Rost, 2008) to ensure that psychometric quality standards are met. The subject-specific contextualization is expected to address gifted students in STEM particularly. For this purpose, scientific illustrations were chosen in the matrices (see Figure 3).

Figure 3 : Sample Item for measuring skills in Scientific-inductive Thinking



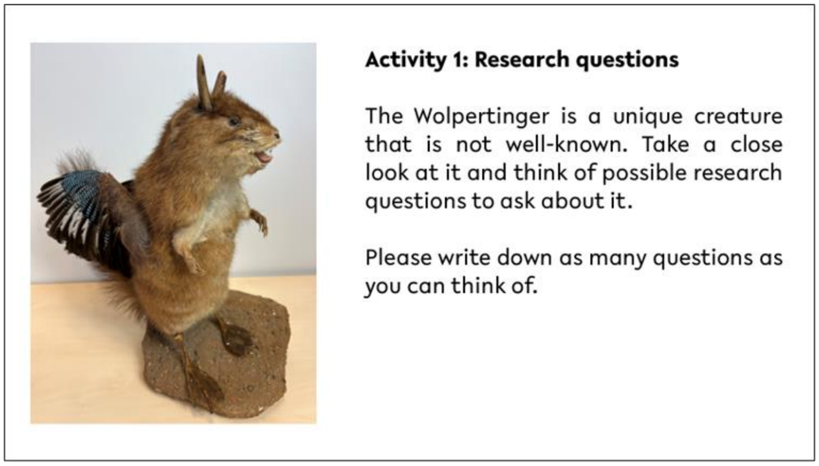
Note. Items are carried out in a digital format. To select an answer, test takers need to click or tap on it. Correct answer is colored.

Benit & Soellner (2012) followed a similar approach by adapting the matrices of an intelligence test to a specific professional field. They were able to achieve a higher face validity for their test. According to this, the student's motivation and interest might be implicitly measured, although this needs additional investigations. The subtest was also designed digitally, enabling easy administration and automated evaluation. The current prototype consists of 24 items. As the test is also planned to enable adaptive testing, its item pool is continuously expanded. Our preliminary studies involved over 200 students from 3rd to 5th grade. The results show that the instruments used were highly accessible and reliable. However, more items should be developed, leading to more detailed assessments of scientific inductive thinking skills. Further steps of validation are planned.

Scientific Divergent Thinking

The analysis of existing test instruments has confirmed that creativity and divergent thinking are essential factors in expressing gifted behavior (Peperkorn & Wegner, 2023a). The traits of *Fluency*, *Flexibility*, and *Originality* play a central role in the manifestation of creativity (Torrance, 1990). *Fluency* is the ability to produce ideas on a specific topic quickly. Fluency can be measured quantitatively by counting these ideas (Torrance, 1990). *Flexibility* refers to the production of ideas that show various options or ways of thinking. It includes the ability to look at things from different angles and to think of many different approaches or strategies. *Flexibility* can be measured by forming response categories (Torrance, 1990). *Originality* describes the ability to produce ideas that are unique and unusual. This involves combining known information on a particular topic in new ways. The O*riginality* of test takers can be measured by comparisons with the total sample (Torrance, 1990). Scientific Creativity is defined "as a kind of intellectual trait or ability producing or potentially producing a certain product that is original and has social or personal value, designed with a certain purpose in mind, using given information" (Hu & Adey, 2002, p. 392). The subtest on *Scientific Divergent Thinking* was derived from the broad ability areas G_l, G_r, and G_v, which were extracted as diagnostically relevant ability areas in scientific giftedness (Avitia & Kaufmann, 2014). Hu & Adey (2002) combined the three essential traits of creativity with two different dimensions in their *Scientific Structure Creativity Model* (SSCM). In the *Product* dimension, traits are focused on the genesis of scientific products of different kinds. These include *technical products*, *scientific knowledge*, *scientific phenomena*, or *science problems*. Within the *Process* dimension, *creative imagination* and c*reative thinking* are differentiated. When looking at the previously developed test instrument (Hu & Adey, 2002), it becomes apparent that the tasks worked with very different stimuli and were located in many different STEM fields. Nevertheless, some of the tasks are very dependent on prior knowledge. Moreover, the test structure does not initiate a creative process that could lead to a higher degree of differentiability. Therefore, the Scientific Divergent Thinking subtest was structured according to the *Incubation Model of Teaching* (IMT; Torrance, 1993), which describes a didactic approach to fostering creativity. The tasks were designed according to the three stages of the model. First, expectations and motivation are raised by providing an exciting stimulus to heighten subjects' curiosity and expectations. This is realized by an imaginal creature composed of different animals, which offers the advantage that influences due to prior knowledge can be largely excluded (see Figure 4). The stimulus piques the curiosity of individuals, leading them to explore more about the imaginary creature and igniting their imagination. In the second stage, additional information is provided through the task format, which enables the subjects to delve deeper into the subject matter. The additional information about the mythical creature should allow the subjects to continue their creative thinking process, which describes an essential condition in this phase (Torrance, 1993). In the third stage, the thinking process is supposed to be extended beyond the information provided. For this purpose, tasks were developed to assess the ability to transfer the stimulus to broader scientific contexts. The subtest on Scientific Divergent Thinking consists of five items evaluated based on fluency, flexibility, and originality. In our pre-studies, we found that constructing the instrument based on the three stages of the IMT (Torrance, 1993) works well and highly motivates the students. However, we noticed that the prototype, which consists of five items, was too lengthy for 3rd to 5th grade students. This led to difficulties in concentration towards the end of the test. In contrast to the previous subtests, evaluating the results is somewhat more complicated since an automatic evaluation of handwritten results requires optical character recognition (OCR) software, which is hardly available to the public (Memon et al., 2020). However, during the conception, an evaluation scheme was created that should enable the evaluation of the results in a time-efficient manner.

Figure 4: Sample Item for measuring Scientific Divergent Thinking (Stage 1).



Activity 1: Research questions

The Wolpertinger is a unique creature that is not well-known. Take a close look at it and think of possible research questions to ask about it.

Please write down as many questions as you can think of.

Formative Evaluation and Re-Design

The designed test instruments will undergo pilot testing in various study settings during the trial phase to ensure proper test development. This will follow the principles of test development (e.g., Irwing & Hughes, 2018) and DBR (e.g., Design-Based Research Collective, 2003). The subtests will be used and tested in clinical settings within a test battery and as multiple short tests in school settings. Within clinical settings, larger samples can be generated, allowing for a more comprehensive analysis of the subtests. During this process, classical test development components will be conducted (Irwing & Hughes, 2018; Lane et al., 2015). It is possible to assess concurrent and divergent validity in clinical settings by comparing existing test instruments. The comparison to IQ tests would be particularly interesting to substantiate the worth of the developed subtests. The planned field studies in school aim to optimally adapt the design of the subtests to practical conditions and ensure that they can be used in schools.

Summative Evaluation

Edelson (2002) describes three types of design-oriented theories that can be formed through summative evaluation in the context of DBR. *Domain theories* are generalized theories of specific problems of practice (Edelson, 2002). In *context theories*, the presented project can observe problems in test administration in classrooms and identify possible solutions. Problems could arise, for example, through digital implementation, the integration into regular lessons or the evaluation of the developed tests alongside the daily school routine. Solutions can be gradually developed through the recurring micro-cycles within the DBR (McKenney & Reeves, 2018). *Outcomes theories* formed through the project could, for example, relate to the general impact of short tests in science giftedness diagnostics. The developed test instruments are intended to facilitate giftedness diagnostics for teachers and provide them with information about the individual giftedness of students in larger groups. However, it is questionable whether teachers will accept integrating such instruments into their lessons and whether implementing the subtests will reduce their workload. To overcome implementation hurdles, teachers should be involved in developing test instruments as expert consultants. Supplementary materials like detailed user manuals, tutorials, and teaching material could be created. In addition, a teacher training course could be developed to train teachers on using test instruments and appropriate fostering methods. In this way, it would be possible to transfer scientific findings directly to schools. Next to *domain theories*, theories of *design frameworks* can also be formed through DBR (Edelson, 2002). These generalized design guidelines can refer to different levels. In the project presented, guidelines on the use of short tests in giftedness diagnostics in general, on the accessibility of digital test formats, or the meaningful integration of short tests in teaching phases might be obtained. *Design methodologies* summarize guidelines concerning the research process (Edelson, 2002). The present project aims to combine criteria from test development and DBR. In both research directions, comprehensive guidelines already exist that describe the respective scientific process.

However, new insights into test development for educational practice could emerge from combining the two approaches, and additional steps could be described, enabling a combination of psychological and pedagogical approaches. Suggestions for successfully recruiting randomized samples from practice and communication with schools would be conceivable. In addition to design-oriented theory, the project presented aims to gain insights into the theoretical foundation of scientific giftedness. By empirically testing the measurement model, conclusions about the correlations between different ability areas and their effect on the expression of gifted behavior in STEM might be drawn. In this way, giftedness diagnostics can be further individualized, enabling adequate educational decisions (Fischer & Fischer-Ontrup, 2022). Besides, the results of the test development can potentially provide further insights into the connections between intelligence and giftedness measurement and provide a knowledge transfer between psychology and pedagogy (Warne, 2016). Initial results from a sample of N = 207 3rd to 5th-grade students show that the scores for scientific inductive thinking and scientific inquiry process skills correlate more highly with each other than with the scores on an IQ test (Peperkorn & Wegner, 2023c). This may indicate that conventional IQ tests cannot adequately diagnose subject-specific aptitudes. Similar results have already been found for university students (Sternberg et al., 2019). In addition to validating the instruments, further studies are needed to test their usability in school settings and meet additional quality criteria of DBR, like reference to the future, openness, context sensitivity, saturation, knowledge diversity, and normativity (Reinmann, 2022).

Conclusion and Outlook

The present article showed that the DBR approach is suitable not only for developing didactic interventions but also for test development and even supplementation with further quality criteria. In the context of giftedness research, the DBR functions as a link between theory and practice and can initiate a transfer between psychology and pedagogy. In addition, scientific findings can be directly implemented in schools through additional perspectives of the DBR and continuously adapted. Within the problem analysis and initial prototype development, an attempt was made to combine approaches from psychology and pedagogy by conducting a CHC-based analysis of diagnostically relevant abilities of scientific giftedness. It should be noted that it is impossible to make final assignments of the diagnostically relevant abilities to the broad ability areas of CHC theory, as there is great potential for discussion and a lack of appropriate data. Nevertheless, the hypothetical measurement model could be used to derive and design subtests to diagnose different facets of scientific giftedness. The CHC theory could be successfully used as a framework model for planning subject-specific giftedness diagnostics and as a communication basis for knowledge transfer (Mickley & Renner, 2019; Warne, 2016). As DBR follows scientific criteria (Fischer et al., 2003; Shavelson et al., 2003; Reinmann, 2005;

2022), psychometric principles could be successfully included in the research process. Placing test development in DBR to develop practice-oriented instruments seems to be a reasonable approach. Nevertheless, it must be noted that obtaining samples and conducting field studies in everyday school life entails some hurdles that can significantly slow down the validation process of the planned instrument. Implementing test development within educational practice is more labor-intensive and requires considerably more commitment. Incorporating psychometric test developments into the DBR approach could create funding and research continuation challenges due to their lengthy design process (Anderson & Shattuck, 2012). To tackle these challenges, the project includes a clinical setting that allows for the rapid acquisition of larger sample sizes. Additionally, obtaining samples outside the school setting allows comparisons between clinical and school practice settings from which further (design-oriented) insights into measuring scientific giftedness can be gained. The presented project aims to empirically test *a priori* models of scientific giftedness (e.g., Wegner, 2014) by recording the structures of the developed test instruments and comparing them with existing instruments. It is questionable to what extent the theoretical assumptions on scientific giftedness are reflected in the measured data and how these can be used to sharpen the concept of scientific giftedness further. Overall, the intended research project can lead to innovations and insights into giftedness research by combining different research approaches and offering opportunities for a scientific transfer.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*(1), 30–60. https://doi.org/10.1037/0033-2909.131.1.30

Alfaiz, F. S., Pease, R., & Maker, C. J. (2020). Culturally Responsive Assessment of Physical Science Skills and Abilities: Development, Field Testing, Implementation, and Results. *Journal of Advanced Academics*, *31*(3), 298–328. https://doi.org/10.1177/1932202X20920572

Alfonso, V. C., Flanagan, D. P., & Radwan, S. (2005). The impact of the Cattell-Horn-Carroll theory on test development and interpretation of cognitive and academic abilities. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and academic abilities* (2nd ed., pp. 185-202). Guilford Publications.

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational researcher*, *41*(1), 16–25. https://doi.org/10.3102/0013189x11428813

Anger, C., Betz, J., Plünnecke, A. (2023). *MINT-Frühjahrsreport 2023 - MINT-Bildung stärken, Potenziale von Frauen, Älteren und Zuwandernden heben* [Enhancing STEM education, maximising the potential of women, older people and immigrants]. Gutachten für BDA, Gesamtmetall und MINT Zukunft schaffen. Köln. https://www.iwkoeln.de/fileadmin/user_upload/Studien/Gutachten/PDF/2023/MINT-Fr%C3%BChjahrsreport_2023.pdf

Avitia, M. J., & Kaufman, J. C. (2014). Beyond *g* and *c*: The relationship of rated creativity to long-term storage and retrieval (*Glr*). *Psychology of Aesthetics, Creativity, and the Arts, 8*(3), 293–302. https://doi.org/10.1037/a0036772

Benit, N., & Soellner, R. (2012). Misst gut, ist gut? Vergleich eines abstrakten und eines berufsbezogenen Matrizentests [Measures well, is good? Comparison of an abstract and a job-related matrix test]. *Journal of Business and Media Psychology*, *3*(1), 22-29. https://journal-bmp.de/wp-content/uploads/22-29_Benit_final.pdf

Bereiter, C. (2002). Design research for sustained innovation. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, *9*(3), 321–327. https://doi.org/10.11225/jcss.9.321

Bergold, S. (2014). Zur diagnostischen Kompetenz von Lehrkräften bei der Identifikation begabter Schülerinnen und Schüler [Teachers' diagnostic competence in identifying gifted students]. *Bildung und Erziehung*, *67*(2), 219-236. https://doi.org/10.7788/bue-2014-0208

Bock, R. D., & Gibbons, R. D. (2021). *Item Response Theory*. John Wiley & Sons. https://doi.org/10.1002/9781119716723

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The journal of the learning sciences*, *2*(2), 141–178. https://doi.org/10.1207/s15327809jls0202_2

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Publications. https://doi.org/10.1017/9781108750561.006

Cobb, P., Confrey, J., DiSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational researcher*, *32*(1), 9-13. https://doi.org/10.3102/0013189x032001009

Demary, V., Matthes, J., Plünnecke, A., & Schaefer, T. (2021). *Gleichzeitig: Wie vier Disruptionen die deutsche Wirtschaft verändern - Herausforderungen und Lösungen* [At the same time: How four disruptions are changing the German economy - challenges and solutions]. IW-Studien: Schriften zur Wirtschaftspolitik aus dem Institut der deutschen Wirtschaft. https://www.iwkoeln.de/fileadmin/user_upload/Studien/IW-Studien/IW-Studie_2021/IW_Studie_2021.pdf

Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational researcher*, *32*(1), 5–8. https://doi.org/10.3102/0013189x032001005

Edelson, D. C. (2002). Design research: What we learn when we engage in design. *The Journal of the Learning sciences*, *11*(1), 105-121. https://doi.org/10.1207/s15327809jls1101_4

Euler, D. (2014). Design-research – a paradigm under development. In D. Euler & P. F. E. Sloane (Eds.), Design-Based Research (Zeitschrift für Berufs- und Wirtschaftspädagogik (Vol. 27, pp. 15–41). Steiner. https://e-books.ciando.com/img/books/extract/3515108416_lp.pdf

Fischer, C., Fischer-Ontrup, C. (2022). Begabungs- und Begabtenförderung im schulischen Kontext [Gifted and Talented Education in the School Context]. In Reinders, H., Bergs-Winkels, D., Prochnow, A., Post, I. (Eds.), *Empirische Bildungsforschung* (S. 1221–1239). Springer VS. https://doi.org/10.1007/978-3-658-27277-7_63

Fischer, F., Bouillion, L., Mandl, H. & Gomez, L. (2003). Bridging theory and practice in learning environment research – Scientific principles in pasteur's quadrant. *International Journal of Educational Policy, Research & Practice, 4*(1), 147-170. https://doi.org/10.1007/978-94-017-0195-2_59

Flanagan, D. P., & Dixon, S. D. (2013). The Cattell-Horn-Carroll Theory of Cognitive Abilities. In C. R. Reynolds, K. J. Vannest, & E. Fletcher-Janzen (Eds.), *Encyclopedia of Special Education*. John Wiley & Sons. https://doi.org/10.1002/9781118660584.ese0431

Gebhard, U., Höttecke, D., & Rehm, M. (2017). Interesse an Naturwissenschaft [Interest in natural sciences]. In U. Gebhard, D. Höttecke, & M. Rehm (Eds.), *Pädagogik der Naturwissenschaften* (pp. 125 – 142). Springer VS. https://doi.org/10.1007/978-3-531-19546-9_8

Heller, K. (1976). Intelligenz und Begabung [Intelligence and giftedness]. In H.-R. Lückert, H. Nickel, & A.-M. Tausch (Eds.), *Studienheft Psychologie in Erziehung und Unterricht*. Reinhardt.

Heller, K. A., Perleth, C., & Lim, T. K. (2005). The Munich model of giftedness designed to identify and promote gifted students. *Conceptions of giftedness*, *2*, 147-170. https://doi.org/10.1017/cbo9780511610455.010

Hu, W., & Adey, P. (2002). A Scientific Creativity Test for Secondary School Students. *International Journal of Science Education*, 24, 389-403.

http://dx.doi.org/10.1080/09500690110098912

iPEGE (2009). *Professionelle Begabtenförderung. Empfehlungen zu Qualifizierung von Fachkräften in der Begabtenförderung* [Professional gifted education. Recommendations for the qualification of professionals in gifted education]. Österreichisches Zentrum für Begabtenförderung und Begabungsforschung (ÖZBF). https://www.oezbf.at/wp-content/uploads/2017/09/iPEGE_1_web.pdf

iPEGE (2014). *Professionelle Begabtenförderung. Fachdidaktik und Begabtenförderung* [Professional gifted education. Didactics and gifted education]. Österreichisches Zentrum für Begabtenförderung und Begabungsforschung (ÖZBF). https://www.oezbf.at/wp-content/uploads/2017/12/iPEGE_4-14-komplett-Standard_01.pdf

Irwing, P., & Hughes, D. (2018). Test development. In P. Irwing, T. Booth, & D. J. Hughes (Eds.), *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development*. John Wiley & Sons Ltd. https://doi.org/10.1002/9781118489772.ch1

Jarosewich, T., Pfeiffer, S. I., & Morris, J. (2002). Identifying gifted students using teacher rating scales: A review of existing instruments. *Journal of Psychoeducational Assessment, 20*(4), 322-336. https://doi.org/10.1177/073428290202000401

Lane, S., Raymond, M. R., Haladyna, T. M., & Downing, S. M. (2015). Test development process. In E. Roeber, & E. R. Trent (Eds.), *Handbook of test development* (pp. 19–34). Routledge.

Machts, N., Kaiser, J., Schmidt, F. T. C., & Möller, J. (2016). Accuracy of teachers' judgments of students' cognitive abilities: A meta-analysis. *Educational Research Review, 19*, 85–103. https://doi.org/10.1016/j.edurev.2016.06.003

McKenney, S., & Reeves, T. C. (2018). *Conducting educational design research* (2nd ed.). Routledge. https://doi.org/10.4324/9781315105642

Memon, J., Sami, M., Khan, R. A., & Uddin, M. (2020). Handwritten optical character recognition (OCR): A comprehensive systematic literature review (SLR). *IEEE Access*, *8*, 142642-142668. https://doi.org/10.1109/access.2020.3012542

Mickley, M., & Renner, G. (2019). Auswahl, Anwendung und Interpretation deutschsprachiger Intelligenztests für Kinder und Jugendliche auf Grundlage der CHC-Theorie: Update, Erweiterung und kritische Bewertung [Selection, application and interpretation of German-language intelligence tests for children and adolescents based on CHC theory: update, extension and critical evaluation]. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, *68*(4), 323-343. https://doi.org/10.13109/prkk.2019.68.4.323

Mittring, G., & Rost, D. H. (2008). Die verflixten Distraktoren: Über den Nutzen einer theoretischen Distraktorenanalyse bei Matrizentests (für besser Begabte und Hochbegabte) [The darned distractors: On the usefulness of a theoretical distractor analysis in matrix tests (for the more gifted and highly gifted)]. *Diagnostica*, *54*(4), 193-201. https://doi.org/10.1026/0012-1924.54.4.193

National Research Council. (1996). *The National Science Education Standards*. National Academy Press. https://doi.org/10.17226/4962

Nowak, K. H., Nehring, A., Tiemann, R., & Upmeier zu Belzen, A. (2013). Assessing students' abilities in processes of scientific inquiry in biology using a paper-and-pencil test. *Journal of Biological Education, 47*(3), 182–188. https://doi.org/10.1080/00219266.2013.822747

Opitz, A., Heene, M., & Fischer, F. (2017). Measuring scientific reasoning–a review of test instruments. *Educational Research and Evaluation*, *23*(3-4), 78-101. https://doi.org/10.1080/13803611.2017.1338586

Peperkorn, C., & Wegner, C. (2020). The Big-Five-Personality and Academic Self-Concept in Gifted and Non-Gifted Students. A Systematic Review of Literature. *International Journal of Research in Education and Science*, *6*(4), 649-667. https://doi.org/10.46328/ijres.v6i4.1290.

Peperkorn, C., & Wegner, C. (2021). Examining Personality Differences between Scientifically Gifted and Nongifted Students: Indications for Gifted Education and Teacher Trainings. *International Journal of Research in Education and Science*, *7*(4), 1245-1262. https://doi.org/10.46328/ijres.2420

Peperkorn, C. & Wegner, C. (2023a). *Measuring scientific giftedness – a systematic review and CHC-based analysis of test instruments* [Manuscript submitted for publication]. Faculty of Biology, Bielefeld University.

Peperkorn, C., & Wegner, C. (2023b). Academic self-concept of gifted and nongifted biology students: Indications for domain-specific gifted education. *Progress in Science Education (PriSE)*, *6*(2), 6-19. https://doi.org/10.25321/prise.2023.1326

Peperkorn, C., & Wegner, C. (2023c, September 18-21). *What is scientific giftedness? - Conception and validation of a subject-*

specific diagnostic instrument [Poster Presentation]. Internationale Tagung der Fachsektion Didaktik der Biologie (FDdB im VBIO) 2023, Ludwigsburg, Germany.

Reinmann, G. (2005). Innovation ohne Forschung? Ein Plädoyer für den Design-Based Research-Ansatz in der Lehr-Lernforschung [Innovation without research? A call for the Design-Based Research in teaching-learning research]. *Unterrichtswissenschaft*, *1*, 52-69.

Reinmann, G. (2020). Ein holistischer Design-Based Research-Modellentwurf für die Hochschuldidaktik [A holistic design-based research model for higher education didactics]. *Educational Design Research, 4*(2), Article 30. https://journals.sub.uni-hamburg.de/EDeR/article/view/1554/1370

Reinmann, G. (2022). Was macht Design-Based Research zu Forschung? Die Debatte um Standards und die vernachlässigte Rolle des Designs [What makes Design-Based Research research? The debate about standards and the neglected role of design]. *Educational Design Research, 6* (2), Article 48. https://doi.org/10.15460/eder.6.2.1909

Renzulli, J.S. (1978). What Makes Giftedness? Reexamining a Definition. *Phi Delta Kappan, 60*(3), 180-184, 261. https://doi.org/10.1177/003172171109200821

Renzulli, J. S., Del Siegle, Reis, S. M., Gavin, M. K., & Reed, R. E. S. (2009). An Investigation of the Reliability and Factor Structure of Four New Scales for Rating the Behavioral Characteristics of Superior Students. *Journal of Advanced Academics*, *21*(1), 84–108. https://doi.org/10.1177/1932202X0902100105

Renzulli J.S., & Reis S.M. (2021). The three-ring conception of giftedness: A change in direction from being gifted to the development of gifted behaviours. In R.J. Sternberg, D. Ambrose (Eds.), *Conceptions of Giftedness and Talent*. Cham. https://doi.org/10.1007/978-3-030-56869-6_19

Rost, D. H. (2013). *Handbuch Intelligenz [Handbook intelligence]*. Beltz.

Rost, D. H. (2015). Das Konstrukt der Intelligenz [The construct of intelligence]. In D. H. Rost (Eds.), *Intelligenz und Begabung, Unterricht und Klassenführung* (pp. 11-45). Waxmann.

Rost, D. H., & Hanses, P. (1997). Wer nichts leistet, ist nicht begabt? Zur Identifikation hochbegabter Underachiever durch Lehrkräfte [Not achieving, not gifted? About the identification of gifted underachievers by teacher-ratings. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *29*(2), 167-177.

Rost, D.H., & Sparfeldt, J.R. (2017). Intelligenz und Hochbegabung [Intelligence and Giftedness]. In Schweer, M. (Ed.), *Lehrer-Schüler-Interaktion* Schule und Gesellschaft (vol. 24, pp. 315 – 346). Springer VS. https://doi.org/10.1007/978-3-658-15083-9_14

Schneider, W. J., & McGrew, K. S. (2018). The Cattell–Horn–Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 73–163). The Guilford Press.

Shavelson, R. J., Phillips, D. C., Towne, L., & Feuer, M. J. (2003). On the science of education design studies. *Educational researcher*, *32*(1), 25–28. https://doi.org/10.3102/0013189x032001025

Spinath, B. (2005). Accuracy of teachers' judgment of pupils' characteristics and the

construct of diagnostic competence. *German Journal of Educational Psychology, 19*,

85-95. https://doi.org/10.1024/1010-0652.19.12.85

Sternberg, R. J. (2018). Direct measurement of scientific giftedness. *Roeper Review*, *40*(2), 78-85. https://doi.org/10.1080/02783193.2018.1434715

Sternberg, R. J., Wong, C. H., & Sternberg, K. (2019). The Relation of Tests of Scientific Reasoning to Each Other and to Tests of General Intelligence. *Journal of Intelligence, 7*(3), 20. http://dx.doi.org/10.3390/jintelligence7030020

Torrance, E. P. (1990). *Torrance Tests of Creative Thinking: Manual for scoring and interpreting results*. Scholastic Testing Service.

Torrance, E. (1993). Understanding creativity. Where to start? *Psychological Inquiry. An International Journal for the Advancement of Psychological Theory*, 4(3), 232–234. https://doi.org/10.1207/s15327965pli0403_17

Warne, R. T. (2016). Five reasons to put the g back into giftedness: An argument for applying the Cattell–Horn–Carroll theory of intelligence to gifted education research and practice. *Gifted Child Quarterly*, *60*(1), 3–15. https://doi.org/10.1177/0016986215605360

Wegner, C. (2014). Fachdidaktik Biologie/Naturwissenschaften [Biology / Science Didactics]. In C. Fischer, E. Hany, A. Holzinger, U. Greiner, V. Müller-Opplinger, C. Perleth, F. Preckel, et al. (Eds.), *4: Vol. 1. Professionelle Begabtenförderung*. (pp. 215–130). IPEGE.

Wetterich, C., & Plänitz, E. (2021). *Systematische Literaturanalysen in den Sozialwissenschaften: Eine praxisorientierte Einführung*

[Systematic literature analyses in the social sciences: a practice-oriented introduction]. Verlag Barbara Budrich.

Zimmerman, R. H., Maker, C. J., & Alfaiz, F. (2020). Culturally Responsive Assessment of Life Science Skills and Abilities: Development, Field Testing, Implementation, and Results. *Journal of Advanced Academics*, *31*(3), 329-366. https://doi.org/10.1177/1932202X2092398