

ISSN 2405-6057

Progress in Science Education

open access journal

research articles &
researched based
reports of practice

Volume **4**

Number **1**

Year **2021**



Editor in-Chief *Nicolas Robin, St. Gallen University of Teacher Education, Switzerland*

Managing Editor *Sanja Atanasova, St. Gallen University of Teacher Education, Switzerland*

Editorial Board *Katrin Boelsterli Bardy, University of Teacher Education Lucerne, Switzerland | Markus Emden, University of Teacher Education Zurich, Switzerland | Urs Kocher, University of Teacher Education Locarno, Switzerland | Andreas Mueller, University of Geneva, Switzerland | Andreas Nehring, Leibniz University Hannover, Germany | Sascha Marc Schmeling, CERN International Relations*

Editorial Associates *Catherine Diederich, University of Teacher Education St. Gallen, Switzerland | Eva Steingruber, University of Teacher Education St. Gallen, Switzerland*

PriSE – Editorial Office *PHSG Institut Fachdidaktik Naturwissenschaften, St. Gallen University of Teacher Education, Notkerstrasse 27, 9000 St. Gallen, Switzerland
Email: PriSE-editors@cern.ch*

Aims and Scope

Science education is a highly dynamic field of applied and basic research and of research-based development. Its ideas and problems arise at the intersection of theoretical and empirical research and of educational practice in science classrooms, informal learning and teacher education, of the important and manifold relations of modern societies with science and education, and of a scientific, evidence-based approach to science teaching and learning and science literacy.

*In this framework, the open access journal *Progress in Science Education* (PriSE) aims at stimulating exchange between researchers, teachers, and other stakeholders in the field, trying to investigate their ideas and visions, and to suggest approaches for an effective and sustainable development of science education in and out of school.*

CERN has always had a strong commitment to science education. As an open institution, making scientific results publically available is an integral part of CERN's mandate. By collaborating with PriSE, CERN can help to strengthen open access to science education. In a sector where many scientists and users – teachers and school students – are not connected to universities and libraries and need access to this information, this collaboration will help to make science education more reachable for everyone.

Progress in Science Education and CERN as publisher propose a new dynamic platform, offering the possibility of peer reviewed, timely publication of high quality research papers in four languages (English, French, German and Italian). With its online open access format it will be accessible for free to a large European and overseas public, including teachers. PriSE is addressing in particular young researchers with the intention to publish their first scientific results. It welcomes quantitative and qualitative empirical research, as well as theoretical, philosophical, programmatic, sociological and historical work, coming from all areas of science education (life, physical, earth, environmental and integrated science), from the intersection with neighboring fields (mathematics, engineering, computer education), and intended for all age groups of learners. To emphasize the importance of the interplay between research and practice, in addition to research articles, there is also a special format for papers in the „Research Based Reports of Practice“ category. In addition, PriSE offers reviews of published books in the field of science education, nature of science and technology.

Submission instructions

All manuscripts must be submitted electronically via

https://e-publishing.cern.ch/index.php/prise/pages/view/submit_your_contribution.

PriSE authors are advised to follow the detailed instructions in the system. Authors should contact the Managing Editor if they have any questions or encounter any problems in the system.

Copyright Notice

Authors who publish with this journal agree to the following terms:

Authors retain copyright and grant the journal right of first publication with the work simultaneously licensed under a Creative Commons Attribution License that allows others to share the work with an acknowledgement of the work's authorship and initial publication in this journal. The applicable licence is <https://creativecommons.org/licenses/by-nc-sa/4.0/>

Authors are able to enter into separate, additional contractual arrangements for the non-exclusive distribution of the journal's published version of the work (e.g., post it to an institutional repository or publish it in a book), with an acknowledgement of its initial publication in this journal.

Authors are permitted and encouraged to post their work online (e.g., in institutional repositories or on their website) prior to and during the submission process, as it can lead to productive exchanges, as well as earlier and greater citation of published work.

TABLE OF CONTENTS

ARTICLES

Choosing an elective - What impact do scientific profile classes have?	5
<small>ALENA SCHULTE, CLAAS WEGNER</small>	
KTSA-A: Konzepttest-Strahlenoptik - Abbildungen	11
Entwicklung eines Konzepttests zur Erfassung von Konzepten der Lichtausbreitung, Streuung und der Entstehung reeller Bilder im Bereich der Strahlenoptik	
<small>ROSA HETTMANNSPERGER, ANDREAS MÜLLER, JOCHEN SCHEID, JOCHEN KUHN, PATRIK VOGT</small>	
Kontext und Problemlösen - Eine Prozessanalyse	36
<small>PATRICK LÖFFLER, ALEXANDER KAUERTZ</small>	
Démarche d'investigation dans l'enseignement des sciences de la nature à l'école primaire	46
Effetes sur la motivation et les représentations des élèves et conséquences sur la formation des enseignants	
<small>LAURENT DUBOIS, ANDREAS MÜLLER, MARINE DELAVAL</small>	
An instrument to measure students' perception of the authenticity of an out-of-school learning place	66
<small>DANIELA SCHRIEBL, ANDREAS MÜLLER, NICOLAS ROBIN, BEAT HENRICH</small>	

Nicolas Robin¹, Editor in-Chief

¹University of Teacher Education St.Gallen, Switzerland

PREFACE

The first issue of *Progress in Science Education* (PriSE) of the year 2021 is not devoted, in contrast to many other specialist journals, to the effects of the pandemic on processes of knowledge transfer and acquisition in science and technology. PriSE continues its declared mission to support researchers in the field of science and technology education by publishing their work at the early stages of their careers. The current edition contains five scientific articles based on doctoral or postdoctoral work. The editorial team is pleased about the growing interest of the scientific community, beyond the borders of Switzerland, in the platform which PriSE offers to young researchers.

The first group of articles deals with the development and validation processes of survey instruments. Rosa Hettmannsperger et al. propose a concept test in the field of radiation optics for secondary school students. The proposed instrument has a high degree of content and curricular validity. In the area of informal scientific and technical education, Daniela Schriebl et al. explore the question of how pupils at lower secondary level perceive the authenticity of learning environments. The instrument developed to measure the perceived authenticity of out-of-school learning places has very good test characteristics. It can therefore be used in broader studies of informal spaces with regard to authentic experiences of science and technology.

In direct connection with the concept of authenticity, the question arises as to the effects of contextualization on learning processes and especially on problem-solving ability. Patrick Löffler and Alexander Kauertz identify the need to examine the contextualization of the problem-solving process more closely as part of an experimental design. This edition of PriSE is complemented by a study on the effects of integrating a research-based approach into science education in primary schools by Laurent Dubois et al. An exploratory study by Alena Schulte and Claas Wegner deals with the factors that motivate secondary school students when choosing an elective and raises the question of the influence of science profile classes.

VORWORT

Die erste Ausgabe von *Progress in Science Education* (PriSE) für das Jahr 2021 widmet sich, im Gegensatz zu vielen anderen Fachzeitschriften, nicht den Auswirkungen der Pandemie auf die Prozesse der Wissensvermittlung und -aneignung in Naturwissenschaft und Technik. PriSE setzt seine erklärte Mission fort, Forscher*innen im Bereich der Fachdidaktik Naturwissenschaft und Technik zu unterstützen, indem es ihre Arbeiten am Anfang ihrer Karriere veröffentlicht. Die aktuelle Ausgabe beinhaltet fünf wissenschaftliche Beiträge, die auf Promotions- oder Postdoc-Arbeiten basieren. Das Redaktionsteam freut sich über das wachsende Interesse der wissenschaftlichen Gemeinschaft, über die Grenzen der Schweiz hinaus, an der Plattform, die PriSE dem wissenschaftlichen Nachwuchs bietet.

Eine erste Gruppe von Artikeln beschäftigt sich mit den Entwicklungs- und Validierungsprozessen von Erhebungsinstrumenten. Rosa Hettmannsperger et al. schlagen einen Konzepttest im Bereich der Strahlenoptik für Sekundarschüler*innen vor. Das vorgeschlagene Instrument hat einen hohen Grad an inhaltlicher und curricularer Validität. Im Bereich der informellen naturwissenschaftlichen und technischen Bildung gehen Daniela Schriebl et al. der Frage nach, wie Schüler*innen der Sekundarstufe I die Authentizität von Lernumgebungen wahrnehmen. Das entwickelte Instrument zur Messung der wahrgenommenen Authentizität des Ortes hat sehr gute Testcharakteristiken. Es kann daher in breiteren Studien über informelle Räume hinsichtlich authentische Erfahrungen von Wissenschaft und Technologie verwendet werden.

In direktem Zusammenhang mit dem Konzept der Authentizität stellt sich die Frage nach den Auswirkungen der Kontextualisierung auf die Lernprozesse und insbesondere auf die Problemlösefähigkeit. Patrick Löffler und Alexander Kauertz identifizieren die Notwendigkeit, die Kontextualisierung im Problemlöseprozess im Rahmen eines experimentellen Designs genauer zu untersuchen. Diese Ausgabe von PriSE wird ergänzt durch eine Studie über die Auswirkungen der Integration des forschenden Ansatzes in den naturwissenschaftlichen Unterricht in der Grundschule

von Laurent Dubois et al. Eine explorative Studie von Alena Schulte und Claas Wegner befasst sich mit den Faktoren, die Schüler*innen der Sekundarstufe in ihrer Berufswahl motivieren und wirft die Frage nach der Wirkung von Klassen mit einem starken naturwissenschaftlichen und technischen Profil auf.

AVANT-PROPOS

Le premier numéro de *Progress in Science Education* (PriSE) de l'année 2021 ne se consacre pas comme beaucoup d'autres revues scientifiques aux effets de la crise sanitaire sur les processus de transmission et d'acquisition des savoirs en sciences et techniques. PriSE poursuit sa mission annoncée de soutien des chercheuses et chercheurs en didactique des sciences et des techniques en publiant leurs travaux au début de leur carrière. Le numéro propose cinq nouvelles contributions scientifiques basées sur des travaux de doctorat ou de post-doctorat. L'équipe éditoriale se réjouit de l'intérêt grandissant de la communauté scientifique, au-delà des frontières de la Suisse, pour la plateforme que propose PriSE aux jeunes chercheuses et chercheurs.

Un premier groupe d'articles se concentre sur le développement et le processus de validation d'instruments de mesure. Rosa Hettmannsperger et al. proposent un test des concepts dans le domaine de l'optique géométrique destiné aux élèves du secondaire I. L'instrument proposé présente une haute validité par rapport aux programmes et aux contenus testés. Dans le domaine de l'enseignement informel aux sciences et techniques, Daniela Schriebl et al. se posent la question de la perception de l'authenticité des lieux d'apprentissage par les élèves du secondaire I. L'instrument développé pour mesurer cette perception de l'authenticité du lieu présente de très bonnes caractéristiques. Il pourra être ainsi utilisé dans des études plus larges sur les espaces informels d'expérience authentique des sciences et des techniques.

En lien direct avec le concept d'authenticité se pose naturellement la question des effets de la contextualisation sur les processus d'apprentissage et notamment sur les capacités à résoudre des problèmes. Patrick Löffler et Alexander Kauertz identifient dans le cadre d'une intervention la nécessité d'étudier plus précisément la contextualisation dans le processus de résolution de problèmes. Ce numéro de PriSE est augmenté par une étude sur les effets de l'intégration de la démarche d'investigation dans l'enseignement des sciences de la nature à l'école primaire par Laurent Dubois et al. Une étude exploratoire d'Alena Schulte et de Claas Wegner s'intéresse aux facteurs qui motivent les élèves du secondaire dans leurs choix professionnels et posent notamment la question de l'effet discuté de classes avec un profil marqué en sciences et techniques.

CHOOSING AN ELECTIVE - WHAT IMPACT DO SCIENTIFIC PROFILE CLASSES HAVE?

Alena Schulte¹, Claas Wegner¹

¹Fakultät für Biologie, Biologiedidaktik, Universität Bielefeld, Germany

*Please address all correspondence to Alena Schulte, alena.schulte@uni-bielefeld.de

STRUCTURED ABSTRACT

Background: With 263,000 vacant jobs in the STEM sector, there is currently a shortage of specialists in Germany. Concurrently, schools report a decrease in scientific interest, particularly noticeable in lower secondary school (Grades 5 – 6), as many students choose to take non-scientific subjects. However, there are a few teaching models that can promote scientific interest and choice: Scientific profile classes (5th – 7th grade) are one possibility to compensate.

Purpose: Previous studies focus on choices made by high school students in 10th grade. We aim to examine the effectiveness of a teaching model that contains profile classes. This will be assessed by investigating which elective students chose after 7th grade, accompanied by exploring influencing factors, such as gender, grade, interest, ability-self-concept and motives for choice.

Sample/Setting: The sample consists of a total of 83 students in 7th grade, where students either attended a scientific profile class or a regular class (n = 55, 29 male/25 female/1 non-binary; n = 28, 15 male/13 female, respectively) at two grammar schools in North Rhine-Westphalia, Germany.

Design Methods: At the end of 7th grade, students answered a questionnaire including items on scientific interest, ability self-concept and motives for choosing an elective.

Results: There were no significant effects on elective choice between profile classes and regular classes ($\chi^2(1)=0.508$, $p = 0.476$) nor between genders ($\chi^2(1)=0.163$, $p = 0.687$). However, the factors utility value and ability self-concept have a significant influence on elective choice.

Conclusions/Implications for classroom practice and future research: This study provides the first attempt to identify what motivates secondary school students to choose an elective. Although we have determined a target group of students to promote science, our study concluded that science profile classes have no effect on elective choice. Future studies should further develop the curriculum and teaching method to effectively promote students' interests.

Keywords: Electives, Profile Classes, Secondary School, Scientific Interest, Motives for Choice

Received: April 2020. **Accepted:** October 2020.

1 INTRODUCTION

In Germany, there are currently 263,000 vacant jobs in the STEM (Science, Technology, Engineering, and Mathematics) industry (MINT-Herbstreport¹, 2019). One explanation could be an observable decline in scientific interest in students throughout their school career (Ferdinand, 2014). Interest in a subject becomes crucial when they are faced with the option of choosing electives or advanced courses, and much later when making career decisions (Hülsmann, 2015). Students begin thinking about their future at the age of 12 (Lindahl, 2003; Maltese &

Tai, 2010): If they are less interested in science, they are less likely to choose a scientific career.

Currently, many teaching models do not exclusively aim to foster scientific interest and ultimately fail at assisting students when they transition to the job market. An intervention, such as vocational training, in Germany begins in middle school (Bundesministerium für Bildung und Forschung & Bundesministerium für Arbeit und Soziales, 2016). However, scientific interest is already very low by this point (Gebhard, Höttecke & Rehm, 2017), so it hardly encourages scientific career decisions.

Research indicates that, in addition to interest, students' choice of school courses is a decisive factor in their

¹ Scientists examine the scientific labour market twice a year



subsequent scientific career decisions. Recent studies focus on the advanced course choice in 11th and 12th grade high school students. However, students can already make initial decisions in the 7th grade when choosing an elective. This field of research has hardly been investigated so far. Depending on the school, a scientific profile can already be formed at this stage, which could also have an influence on a later scientific career. This study examines factors which influence elective choice in 7th grade.

2 RESEARCH BACKGROUND

2.1 Profile Classes

Profile classes offer a way for schools in Germany to develop a distinct profile. Schools have the opportunity to provide their students with individual support in various fields, such as music, languages or social sciences. In this study, scientific profile classes are examined (Schulte & Wegner, 2020). Prospective students are admitted to the profile classes based on interviews and the previous math and science grades. These classes are specifically intended to promote scientific interest and self-concept. Furthermore, they advocate students to continue science courses, which could later impact their career choice.

Implementing profile classes between 5th and 7th grade provides students with more time and flexibility, as it offers additional teaching time (1 extra lesson per week) as well as a wide range of extracurricular after-school support. It is mandatory to participate in scientific study groups and take part on numerous excursions, which has also been shown to trigger interest (Henriksen, Dillon & Ryder, 2014).

The scientific profile classes examined follow an alternative curriculum with context-oriented lessons and cover topics relevant to everyday life. The teaching material focuses on the scientific methods through observation, experimentation, and result comparison with discussions. Students can develop their own ideas, research questions and individually conduct experiments. This enabling them to choose topics in science class (Lindahl, 2003).

Teaching science in school is essential to maintain and promote interest, yet it is not always effective. Students often criticize the transmissive teaching as well as decontextualized and difficult topics, which subsequently have a negative effect on their interest and engagement with the subject (Lyons, 2006).

Recent research indicates that context-oriented teaching units are particularly beneficial to increase interest (Habig, Blankenburg, van Vorst, Fechner, Parchmann & Sumfleth, 2018).

2.2 Electives in Middle School

In Germany, educational policy is determined by each federal state. In North Rhine-Westphalia, the mandatory curriculum is supplemented by electives which start in 8th and continue until 9th grade. This is the initial opportunity for students to individualize their school profile and choose their specializations. Whereas students are used to

having all subjects with the same classmates, the electives are arranged in courses mixing students from an entire grade and, therefore, breaking up the normal “class system”. Student performance in electives is graded and credited for graduation (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, n.d._a). At the school of this study, the following electives are offered: Spanish, biology/chemistry; mathematics/informatics; social sciences; theatre/media/performance and physics/technology. However, other schools may vary their electives depending on their educational resources or school profiles (e.g. bilingual history, biology/geography, and other foreign languages).

Further specialization occurs in the last two years of high school. Students are required to choose four “Abitur²” subjects to graduate: two advanced and two basic courses. Advanced courses are taught with more lessons per week and cover content in greater depth. Finally, students complete the “Abitur” with tests in their four selected subjects. Depending on which subjects were chosen, a scientific focus is possible (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen, n.d._b).

2.3 Scientific Interest

In scientific profile classes, the subject of interest is formed by the different teaching content in the subjects biology, chemistry and physics as well as extracurricular topics in the additional lessons.

Several studies indicate that scientific interest declines over the course of the school career (e.g. Gardner, 1987; Merzyn, 2008). The declining interest could possibly be explained by the theoretical and complex teaching of scientific knowledge. The inability to connect to science education starting from primary school on contributes to the loss of interest among students from 5th to 10th grade (Brüggemeyer, 2018).

It is accompanied by a decline in the continuation of science subjects during school, which is a global phenomenon that can be observed in Australia, Canada, Japan, and the EU (Lyons, 2006). According to Maltese and Tai (2010), students who are interested in scientific topics report that they were excited in early childhood. The authors state that scientific interest is already detected in primary school, or in the first years of lower secondary school. They explain that an origin of interest can be classified in three ways: intrinsic self-interest, school/education-based experience, and interest encouraged by a family member, with the first two having the most influence. Students who indicate at the beginning of their education that they aim to pursue a career in science, may change their minds during their school career. Therefore, schools play a decisive aspect in influencing interest in science, and should aspire to maintain and promote existing interest at an early stage.

Many studies on course selection focus on advanced course choice during the last few years of high school, which repeatedly has been demonstrated to correlate with interest (Abels, 2002; Merzyn, 2010; Pohlmann & Streblov, 2017). Hülsmann (2015) illustrated that the proba-

² Students take their “Abitur” examination in four subjects. However, during the final qualification phase between 11th and

12th/13th grade, students take between 11 and 13 subjects altogether. (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen,, n.d._c).

bility of choosing chemistry as a subject in upper secondary school increased if students participated in a science elective in lower secondary school. We infer that students choosing a science elective could have been impacted by participating in profile classes beforehand. If they are particularly interested in the natural sciences, it is more likely that they will choose a scientific elective and later on, an advanced science course. Ultimately, choosing an advanced science course is one factor towards pursuing a scientific career (Merzyn, 2010).

Students who choose to continue their school career in fields other than science are characterised by lower scientific competency and scientific interest. Both factors can be traced to negative experiences in science lessons throughout grades 5 and 6 and the perceived irrelevance of the content (Merzyn, 2010; Cleaves, 2005). Therefore, profile classes are intended to prevent negative experiences and instead promote interest, self-concept and competences in the sciences. Furthermore, personal attributes such as ability self-concept, grades and prospects of success also play an important role in course selection (Abels, 2002; Cleaves, 2005).

As there are few studies about the selection of science electives in relation to interest in secondary schools, our study aims to bridge this gap. Additionally, it examines whether a teaching model involving profile classes is effective at maintaining and fostering scientific interest, which is hoped to support the choice of continuing science subjects.

2.4 Research Hypotheses

This study examines whether profile classes promote scientific interest and explores motives that are decisive when choosing an elective:

- (i) Which electives do students choose and are there differences between students in scientific profile and regular classes?
 - Students in profile classes tend to choose a science elective (see e.g. Malteste & Tai, 2010; Hülsmann, 2015).
- (ii) Which factors have an influence on choosing a science elective?
 - Interest, ability self-concept and grades in biology, chemistry and physics have an influence on elective choice (see e.g. Hülsmann, 2015; Abels, 2002).

3 METHODS

The sample consisted of three 7th grade classes (n = 83) which were split by two profile classes and one regular class (n = 55 and 28, respectively). All students completed a questionnaire at the end of 7th grade. The instrument contained 32 closed items, using a 6-point rating scale from "strongly agree" to "strongly disagree" (scientific interest, motives for choice) and "very good" to "very bad" (ability self-concept), as well as one open item (see Tab. 1). Questionnaire scales are taken from already used and proven test instruments.

Items were divided into motives which *directly* influence choice, such as interest-enjoyment, attainment and utility, and *indirect* motives which are assumed to have an influence (scientific interest, ability self-concept and grades). The scale motives for choice was selected from Hülsmann's (2015) questionnaire.

Finally, students indicated which elective they chose. Additional data such as gender, age and grades in biology, chemistry and physics were collected.

Tab. 1. Test instrument scales: reference, number of items, an example item and Cronbach's alpha are given. The items were translated from German for the purpose of the publication.

	Scale (Reference)	No. of items	Item Example	Cronbach's α
Indirect motives	Scientific Interest ³ (Ferdinand, 2014; Schreiner & Sjøberg, 2004; Frey, 2009)	13	I enjoy studying scientific topics.	.95
	Ability Self-Concept (Hofman, Häußler & Lehrke, 1998)	7	According to my own assessment, my performance in the natural sciences is94
Direct motives	Motives for Choice (Hülsmann, 2015)		I have chosen this specialty for the next two school years ...	
	Interest-Enjoyment	3	... because I am interested in the elective and its topics.	.88
	Attainment value	2	... because it is important that I learn a lot about this subject.	.78
	Utility value	3	... because I expect that I will require knowledge in this area in the future.	.67
	Elective	1	Which elective did you choose?	-

³ We performed an exploratory factor analysis. Variables were suitable (Bartlett test (Chi-square(78) = 837.59, $p < .000$; KMO = .930). A principal component analysis with varimax rotation

indicates a single factor-solution that accounts for 64.5% of the variance.

4 RESULTS

To summarize our data, the electives biology/chemistry, mathematics/informatics and physics/technology were assigned as science electives and elective selection was analyzed as a dichotomous variable (yes/no). A chi-square test was conducted to explore differences in choosing an elective based on class type. All expected cell frequencies were greater than 5. There was no difference between chosen elective and class type or gender ($\chi^2(1)=.508$, $p=.476$; $\chi^2(1)=.163$, $p=.687$, respectively, see Fig. 1).

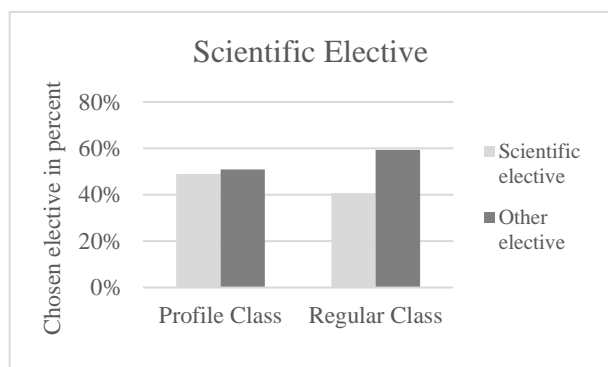


Fig. 1. Distribution of selected electives from students in regular (n=28) and profile classes (n=55).

Separate logistic regression analyses were calculated for direct and indirect motives to investigate influential factors for elective choice.

The model for direct motives ($\chi^2(2)=11,199$, $p=.004$) and the z-test for the regression coefficient utility value are significant (Wald(1) =7,518, $p=.006$). (Nagelkerke R²Quadrat =.194, corresponds to a strong effect according to Cohen (1992), see Tab. 2).

Additionally, we examined the influence of the indirect motives such as scientific interest and ability self-concept using a logistic regression analysis

The model for indirect motives ($\chi^2(3)=12,868$, $p=.005$) and the z-test for the regression coefficient ability self-concept are significant (Wald(1)=4.455, $p=.035$). (Nagelkerke R²Quadrat =.171, corresponds to a strong effect according to Cohen (1992), see Tab. 2).

The factors utility value and self-ability concept are significant predictors of interest for future elective choice. Grades in biology, chemistry and physics, as well as interest-enjoyment and attainment value have no significant influence on elective choice.

Tab. 1. Logistic regression analyses predicting elective selection from direct (above) and indirect (below) factors.

Predictor	β	Wald χ^2	p	Odds Ratio
Interest-Enjoyment	.708	1.77	.184	2.03
Attainment value	-.065	0.04	.848	.94
Utility value	.702	7.52	.006	2.02
Scientific Interest	-.051	0.02	.880	.951
Ability Self-Concept	.991	4.46	.035	2.69

Students in profile classes (M=3,59, SD=1,04) show a significantly higher scientific interest ($t(81)=2,457$, $p=.016$) than students of the regular class (M=2,96, SD=1,12). Furthermore profile class students have significantly better grades in chemistry ($t(80)=-4,479$, $p=.000$) and physics ($t(80)=-3,595$, $p=.001$). There is no significant difference in the subject biology.

5 DISCUSSION AND CONCLUSION

Profile classes are intended to build interest as they implement a curriculum with specifically designed context-oriented units. This provides flexibility while adapting to students' previous knowledge and interests. However, we found that attending the science profile class did not encourage students to choose a scientific elective more than those attending the regular class.

Both direct and indirect motives are assumed to influence elective choice. However, the factors utility value and ability self-concept were the only significant predictors (Cleaves, 2005). When students were asked about their reasons for choosing an elective, they justified their decision solely based on how useful they find the subject for their future. Utility-value has been shown to increase in importance with advanced courses (Hülsmann, 2015), and now this is seen in younger students in profile classes. We also observed an influence of self-concept on elective choice. This may be partially explained by perceived positive self-closeness, which is defined as the extent to which a person uses the elective to define themselves as well as their confidence in ability (Hannover & Kessels, 2004).

Interest in the elective and a general scientific interest did not seem to have a significant influence on the decision to choose an elective. In contrast to other studies, interest here does not significantly influence elective choice (e.g. Busch, 2016). Despite a comparatively higher level of interest, students in the profile class do not choose a scientific elective. One reason for this could be that new areas of electives like theatre or media were introduced in schools at the time of the survey. Many students chose these electives due to a novelty effect; those in the profile class may have preferred other electives because they had already acquired "sufficient" knowledge in science. In order to verify this assumption, interviews would have to be conducted with students to analyse reasons to choose or avoid certain electives.

Another influencing factor could be that students have to decide to attend a profile class at the beginning of high school. Therefore, it cannot be ruled out that there may also be students interested in science in the regular classes. These students might also choose science electives.

As attainment-value did not have an influence on elective choice, it seems that students do not mention the electives' relevance for their own learning as a reason for elective choice. Finally, grades in biology, chemistry and physics had no influence on elective choice, which could indicate that students choose an elective regardless of subject performance. This aspect is supported by the fact that students in profile classes have significantly better grades in chemistry and physics.

To examine if grades have an influence, all grades would have to be recorded and analysed to look for correlations with their chosen elective.

These results illustrate necessary adaptations for the existing profile classes in this study. Vocational training (such as internships or inviting experts and scientists) should be offered earlier to provide students with the opportunity to be exposed to scientific professions, as students in 7th grade already make choices with their future career in mind (Lindahl, 2003; Lyons, 2006). This adaptation is supported by the items exploring construct utility value, which inquire the professional relevance of the area of elective. Although we did not see an impact of interest in our study, numerous studies show the importance of interest for a later scientific career (e.g. Hülsmann, 2015; Lyons, 2006; Henriksen, Dillon & Ryder, 2014; Abels, 2002). A school-related approach to improve profile classes would be to adjust the subject matter and improve context-oriented teaching units to benefit students on an individual level (van Vorst et al. 2015; Sennebogen, 2013).

Our study adds to the growing body of evidence that helps identify factors that encourage students to choose a scientific elective. Although profile classes are not yet ideal, we suggest further changes in the curriculum to promote scientific self-concept, a greater focus on student interest and starting vocational training earlier. If students are confident in their own abilities, students might choose another elective area. Furthermore, if students tend to be interested in all fields of science, they will choose more science courses. Ultimately, this could affect future career choices and hopefully might affect the shortage of STEM workers.

6 LIMITATIONS

One limitation of this study is the small sample size. Our results can only be interpreted in relation to the profile classes mentioned here and do not allow general conclusions to be drawn about the voting behavior of 7th graders. The control group only consists of one regular class; this must be enlarged in following studies to support our results. Subsequently, a cross-sectional study on elective choice in middle schools within all grammar schools in Nordrhein-Westfalen would be a useful approach to study elective choice in profile class students.

In order to obtain reliable scales, items that were used in the study by Hülsmann (2015) had to be excluded. Additionally, only a small number of variables are included in the calculation for attainment and utility value. In a subsequent survey, the questionnaire will be supplemented by further items.

Further studies should look at the extent to which profile classes already pursue context-oriented teaching, such as investigating teaching methods using video analysis or teacher and student interviews. We should also aim to increase our sample size, as it would be useful to survey groups at different schools for an entire year to exclude external influences, such as teaching styles, previous educational experiences and enrichment programmes. Additionally, most comparative studies look at students choosing advanced courses and not electives,

therefore future studies should aim to look at studying elective selection, as career choices are already considered during lower secondary school (Lindahl, 2003 & Lyons, 2006). Based on Lindahl (2003), expanding the study by interviewing students about their decisions for elective choice would be also beneficial.

This study provides valuable information about choosing electives in profile class students and offers methods to promote scientifically interested students.

ACKNOWLEDGEMENT

The authors would like to thank Vivienne Litzke for constructive criticism and language editing of the manuscript.

REFERENCES

- Abels, J. (2002). Kurswahlen aus Interesse? *Die deutsche Schule*, 94(2), 192-203.
- Brüggemeyer, M. (2018). Die Chemie im naturwissenschaftlichen Unterricht der Klassen 5 und 6. Dissertation Universität Bielefeld. Hamburg: Dr. Kovač.
- Bundesministerium für Bildung und Forschung & Bundesministerium für Arbeit und Soziales (2016). Bildungskettenvereinbarung Nordrhein-Westfalen. Retrieved from: <https://www.bildungsketten.de/de/604.php>
- Busch, M. (2016). Empirische Studien zum fächerübergreifenden naturwissenschaftlichen Unterricht. Universität Jena.
- Cleaves, A. (2005). The formation of science choices in secondary school. *International Journal of Science Education*, 27(4), 471-486.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 122(1), 155-159.
- Ferdinand, H. (2014). Entwicklung von Fachinteresse: Längsschnittstudie zu Interessenverläufen und Determinanten positiver Entwicklung in der Schule. Münster [u.a.] : Waxmann.
- Frey, A. (Ed). (2009). PISA-2006-Skalenhandbuch. Münster [u.a.] : Waxmann.
- Gardner, P. (1987). Schülerinteressen an Naturwissenschaften und Technik. M. Lehrke, L. Hoffmann & M.-L. Brüggemann (Eds.), *Schülerinteressen am naturwissenschaftlichen Unterricht. Untersuchungen und Erklärungen* (Beiträge zum 12. IPN-Symposion). Köln: Aulis Verlag Deubner, 13-49.
- Gebhard, U., Höttecke, D., & Rehm, M. (2017). Pädagogik der Naturwissenschaften: Ein Studienbuch. Lehrbuch. Wiesbaden: Springer VS.

Habig, S., Blankenburg, J., van Vorst, H., Fechner, S., Parchmann, I., & Sumfleth, E. (2018). Context characteristics and their effects on students' situational interest in chemistry. *International Journal of Science Education*, 40(10), 1154-1175.

Hannover, B., & Kessels, U. (2004). Empfundene „Selbstnähe“ als Mediator zwischen Fähigkeitsselbstkonzept und Leistungskurswahlintentionen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 36(3), 130-138.

Henriksen, E.K., Dillon, J., & Ryder, J. (Eds). (2014). Understanding student participation and choice in science and technology education. Dordrecht: Springer.

Hofmann, L., Häußler, P., & Lehrke, M. (1998). IPN-Interessenstudie Physik. Kiel : IPN.

Hülsmann, C. (2015). Kurswahlmotive im Fach Chemie. Berlin: Logos

Institut der deutschen Wirtschaft (2019). MINT-Herbstreport 2019. Retrieved from: <https://www.iwkoeln.de/studien/gutachten/beitrag/christina-anger-oliver-koppel-axel-pluennecke-enno-roeben-ruth-maria-schueler-basis-zur-zukunftssicherung-durch-forschung-und-digitalisierung.html>

Lindahl, B. (2003). Pupils' responses to school science and technology? A longitudinal study of pathways to upper secondary school. English summary of a dissertation. Retrieved From: https://www.researchgate.net/publication/237722627_Pupils'_responses_to_school_science_and_technology_A_longitudinal_study_of_pathways_to_upper_secondary_school

Lyons, T. (2006). Different Countries, Same Science Classes: Students' experiences of school science in their own words. *International Journal of Science Education*, 28(6), 591-613.

Maltese A., & Tai, R. (2010). Eyeballs in the Fridge: Sources of early interest in science. *International Journal of Science Education*, 32(5), 669-685.

Merzyn, G. (2008). Naturwissenschaften, Mathematik, Technik – immer unbeliebter? Die Konkurrenz von Schulfächern um das Interesse der Jugend im Spiegel vielfältiger Untersuchungen. Baltmannsweiler: Schneider Verlag Hohengehren.

Merzyn, G. (2010). Kurswahlen in der gymnasialen Oberstufe. Beiträge zur DPG-Frühjahrstagung.

Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen. (n.d.a) Mittelstufe. Retrieved from: <https://www.schulministerium.nrw.de/docs/Schulsystem/Schulformen/Gymnasium/Sek-I-G9/Mittelstufe/index.html>

Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen . (n.d.b) Sekundarstufe II. Retrieved from: <https://www.schulministerium.nrw.de/docs/Schulsystem/Schulformen/Gymnasium/Sek-II/index.html>

Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen. (n.d.c) Schulformen NRW. Retrieved from: <https://www.schulministerium.nrw.de/docs/Schulsystem/Schulformen/index.html>

Pohlmann, B., & Streblow, L. (2017). Vorhersagen von Kurswahl und Leistung in der gymnasialen Oberstufe: Fächerübergreifende Zusammenhänge. Retelsdorf, J. (Eds). *Im Blickpunkt pädagogisch-psychologischer Forschung. Selbstbezogene Kognitionen, sprachliche Kompetenzen und Professionalisierung von Lehrkräften*. Münster: Waxmann, 47-67.

Sennebogen, S. (2013). Kooperatives Lernen mit Wettbewerb im Biologieunterricht. Berlin: Logos.

Schulte, A. & Wegner, C. (2020). Science-Klassen – Wie gelangt naturwissenschaftliche Förderung in die Schulen? *MNU Journal*, 73(4), 309-312.

Schreiner C., & Sjøberg, S. (2004). Sowing the seeds of ROSE. Background, Rationale, Questionnaire Development and Data Collection for ROSE (The Relevance of Science Education) – a comparative study of students' views of science and science education (pdf) (Acta Didactica 4/2004). Oslo: Dept. of Teacher Education and School Development, University of Oslo.

van Vorst, H., Dorschu, A., Fechner, S., Kauertz, A., Krabbe, H., & Sumfleth, E. (2015). Charakterisierung und Strukturierung von Kontexten im naturwissenschaftlichen Unterricht – Vorschlag einer theoretischen Modellierung. *Zeitschrift für Didaktik der Naturwissenschaften*, 21, 29-39.

KTSO-A: Konzepttest-Strahlenoptik – Abbildungen. Entwicklung eines Konzepttests zur Erfassung von Konzepten der Lichtausbreitung, Streuung und der Entstehung reeller Bilder im Bereich der Strahlenoptik

Rosa Hettmannsperger¹, Andreas Müller², Jochen Scheid³, Jochen Kuhn⁴ und Patrik Vogt⁵

¹Hessische Lehrkräfteakademie, Wiesbaden

²Section Physique, Faculté des Sciences, Université de Genève

³Institut für naturwissenschaftliche Bildung, AG Physikdidaktik, Universität Koblenz-Landau, Campus Landau

⁴Fachbereich Physik/Didaktik der Physik, Technische Universität Kaiserslautern

⁵Institut für Lehrerfort- und -weiterbildung, Mainz

*Please address all correspondence to Rosa Hettmannsperger, rosa.hettmannsperger-lippolt@kultus.hessen.de

STRUCTURED ABSTRACT

Hintergrund: Die Erfassung von konzeptuellem Wissen und Verständnis ist ein Thema, das sowohl in zahlreichen physikdidaktischen Forschungsarbeiten als auch für die Diagnose des Kenntnisstands von Schülerinnen und Schülern im Unterrichtskontext relevant ist. Für den Bereich der Strahlenoptik liegt zwar seit Jahrzehnten eine breite Wissensbasis zu konzeptionellen Lernschwierigkeiten und Schülervorstellungen und darauf basierenden Items vor, sowie auch verschiedene Tests, die jedoch nur ansatzweise psychometrisch charakterisiert und validiert sind.

Ziel: Aus diesem Grund wurde im Rahmen einer Interventionsstudie im Physikunterricht der Sekundarstufe I ein Konzepttest zu den Teilbereichen Lichtausbreitung, Bildentstehung und Strahlenmodell/Bildkonstruktion im Kontext des Lerngegenstandes „Abbildung bei Sammellinse“ entwickelt und empirisch auf das Vorliegen akzeptabler Test-Charakteristika überprüft (Konzept-Test Strahlenoptik: Optische Abbildungen – KTSO-A).

Stichprobe: Das Testinstrument (bestehend aus 10 Items) wurde in zwei Teilstichproben von Schülerinnen und Schülern ($n = 389$ und $n = 480$) der Sekundarstufe I erprobt.

Methode: Die Testkonstruktion basiert auf Forschungsergebnissen empirischer Studien zu konzeptuellen Schwierigkeiten in dem betrachteten Teilbereich der Strahlenoptik und folgt der Literatur zum Vorgehen bei der Entwicklung von Konzepttests. Neben der Berechnung üblicher Test- und Itemkennwerte (Itemschwierigkeiten, Trennschärfen, Reliabilität) wurden im Rahmen einer Kreuzvalidierung zwischen den beiden Teilstichproben die Methoden der exploratorischen und konfirmatorischen Faktorenanalysen kombiniert und zudem ein Expertenrating umgesetzt.

Ergebnisse: Die Kennwerte des Tests (Itemschwierigkeiten, Trennschärfen, Reliabilität) liegen für die beiden Stichproben durchwegs bzw. größtenteils in den in der Testtheorie üblichen Akzeptanzbereichen. Abweichungen davon, Unterschiede zwischen den Stichproben und Details auf Item- und Distraktorebene, lassen sich und unter Bezug auf die vorliegende Forschung zu Konzepttests sinnvoll interpretieren. Die inhaltliche und curriculare Validität des Tests wurde in einem Expertenrating als hoch eingestuft. Mittels des Verfahrens der Kreuzvalidierung wurde die Dimensionalität des Tests untersucht. Die Analyse zeigt eine dreidimensionale Struktur mit akzeptablen bis guten Kennwerten auf (gute Fit-Indizes: $\chi^2_{(32)} 44.13$, $p = 0.075$; $CFI = 0.98$; $TLI = 0.98$, jedoch teilweise niedrige Faktorladungen) und ermöglicht eine plausible inhaltliche Deutung in folgenden Teildimensionen: Bildentstehung, geradlinige Lichtausbreitung und Streuung sowie Verständnis des Strahlenmodells und der Bildkonstruktion.

Diskussion und Relevanz: Die Gesamteinschätzung des vorgestellten Tests wird im Hinblick auf den Stand der Entwicklung von Konzept-Tests in anderen Bereichen und dabei bekannte Begrenzungen und Schwierigkeiten bezüglich Validität, Reliabilität und Strukturaufklärung ausführlich diskutiert. Die psychometrischen Eigenschaften des KTSO-A können im Vergleich mit dem Entwicklungsstand von Konzepttests im Allgemeinen als akzeptabel bis gut eingeschätzt werden. Darüber hinaus werden einige mögliche Perspektiven für künftige Weiterentwicklungen diskutiert. Der Wert des vorgestellten Konzept-Tests sehen wir darin, dass er für einen inhaltlich und curricular wohldefinierten Bereich eine praktikable und nach gängiger Forschungspraxis im Feld psychometrisch adäquate Testfassung bereitstellt, die für Forschungszwecke und zur unterrichtlichen Diagnostik eingesetzt werden kann.

Keywords: Konzepttest, Strahlenoptik, optische Abbildungen, Faktorenanalyse



STRUCTURED ABSTRACT

Background: A number of several concept inventories have been developed for various domains of physics, such as mechanics, heat, electricity or astronomy. There can be obtained test instruments for the domain of ray optics, as well. However, these inventories were only partially psychometrically validated.

Purpose: For this reason, an inventory (KTSO-A Konzepttest Strahlenoptik-Abbildungen / Concept Inventory Ray Optics – Imaging), was created detecting conceptions in the domains of light propagation, image formation and understanding ray diagrams in the context of imaging via convex lens. An empirical investigation of item statistics and test parameters was carried out to check if all parameters were within acceptable range.

Sample: The developed inventory consisting of 10 items was investigated based on two samples of students in the 7th and 8th grade at secondary schools in Rhineland-Palatinate ($n = 389$ and $n = 480$).

Methods: The outlined instrument is based on a well-known body of results on students' difficulties with learning optics and related to intuitive conceptions reported for this domain. The method of test construction followed the procedure of creating inventories in physics education known in notable literature.

Results: An item analysis revealed that the inventory tasks allow us to gather students' knowledge and understanding with a satisfactory discriminatory power and reliability. In order to investigate the dimensionality of the test, exploratory and confirmatory factor analyses were carried out. Results ($\chi^2_{(32)} 44.13, p = 0.075$; $CFI = 0.98$; $TLI = 0.98$, $RMSEA = 0.05$; $SRMR = 0.03$) indicate that there are three clearly interpretable subdimensions: understanding image formation, conceptions of linear light propagation and scattering as well as understanding ray diagrams.

Conclusion: In accordance with psychometric standards, the presented test version can be already used both in science education research and classrooms at schools. Finally, some prospects for future improvements of the presented inventory are outlined.

Keywords: *concept inventory, ray optics, imaging, factor analysis*

Received: May 2020. **Accepted:** August 2020.

1 EINLEITUNG

Forschungsergebnisse zum Thema Schülervorstellungen und Lernschwierigkeiten zeigen, dass das Denken vieler Schülerinnen und Schüler von inkonsistenten und fachlich unangemessenen Vorstellungen geprägt ist (Duit & Treagust, 2003; Duit, 2009; Scott, Adams & Leach, 2007; Shtulman & Lombrozo, 2016). Viele der fehlerbehafteten Vorstellungen der Schüler sind tief in den Alltagserfahrungen verankert, andere werden erst im Unterricht selbst gebildet. Klassische Lehrstrategien erweisen sich oft als wenig effektiv darin, die Vorstellungen der Lernenden zu verändern (Andersson & Kärrqvist, 1983; White & Gunstone, 1989; Fetherstonhaugh & Treagust, 1992; Galili, 1996; Langley, Ronen & Eylon, 1997; Limon, 2001; Heywood, 2005). Lernprozesse, die eine Wissensveränderung beinhalten, sind in der Regel zwar recht (zeit-)aufwändig, aber dennoch realisierbar. So belegen Guzzetti, Snyder, Glass und Gamas (1993) in einer Metaanalyse, dass Vermittlungsstrategien, in denen Konflikte zwischen den vorunterrichtlichen Vorstellungen und den zu lernenden wissenschaftlichen Vorstellungen herausgearbeitet worden sind, solchen Strategien statistisch signifikant überlegen waren, in denen das nicht der Fall war.

Die Feststellung, dass das konzeptuelle Verständnis von Lernenden verändert werden kann, ist jedoch daran geknüpft, dieses Verständnis erfassen zu können. Hier setzt die Entwicklung von Konzepttests an, wie sie für

viele Gebiete der Naturwissenschaften bereits entwickelt wurden (Liu, 2012; Madsen, McKagan & Sayre, 2017). Auch im Bereich der (Strahlen-) Optik liegen eine Reihe von Wissens- und Verstehensfragen sowie einige Tests vor, die auf die Erfassung von konzeptuellem Wissen und Verständnis in der (Strahlen-) Optik zielen und hierbei weitverbreitete Schülervorstellungen und Lernschwierigkeiten einbeziehen (siehe Abschnitt 2.2). Bislang fehlen jedoch publizierte, psychometrische charakterisierte Testinstrumente, mit denen verschiedene Wissensfacetten im Bereich der Strahlenoptik in praktisch und methodologisch zufriedenstellender Weise geprüft werden können. Die vorliegende Arbeit soll einen Beitrag zur Schließung dieser Lücke zu den Teilbereichen Lichtausbreitung, Bildentstehung und Strahlenmodell/Bildkonstruktion leisten. Es wird zunächst ein Forschungsüberblick zu konzeptionellen Schwierigkeiten gegeben und dann die Entwicklung und psychometrische Charakterisierung eines einschlägigen Testinstrumentes vorgestellt.

2 FORSCHUNGSHINTERGRUND

2.1 Definitionen und Vorgehen

Konzeptuelles Wissen definiert sich nach (Byrnes & Wasik, 1991, S. 777) als Wissen über Kernkonzepte einer Domäne und die Beziehung dieser Kernkonzepte untereinander. Es zeichnet sich gemäß Byrnes und Wasik durch

die kompetente Verwendung unterschiedlicher Konstrukte aus, welche aus semantischen Netzwerken, Hierarchien und mentalen Modellen bestehen können. Baumert et al. (2001) betonen in Hinblick auf die PISA-Studien, dass Schülerinnen und Schüler konzeptuelles Verständnis benötigen, „um Phänomene der natürlichen und der von Menschen geschaffenen Welt zu verstehen“. Deshalb wird hier weniger Wert auf die Reproduktion von Faktenwissen gelegt, hingegen geht es darum, „ein konzeptuelles Verständnis“ zu erfassen, das mit der Anwendung von Alltagskonzepten beginnt und bis zum Arbeiten mit naturwissenschaftlichen Modellvorstellungen reicht.

Unter didaktischer Perspektive kann die Erfassung des Wissensstandes helfen, das Denken der Lernenden kennenzulernen, um den Unterricht entsprechend anpassen und abstimmen zu können (vgl. Tyson, Venville, Harrison & Treagust, 1997; Özdemir & Clark, 2007). Methoden, die einen Einblick in das Denken der Schülerinnen und Schüler und ihre Lernschwierigkeiten erlauben, wie beispielsweise offene Interviews oder halbstandardisierte Fragebögen, sind in der Forschungspraxis und im Schulalltag sowohl bei der Durchführung als auch bei der Auswertung mit einem hohen zeitlichen Aufwand verbunden. Tests im Multiple-Choice Format weisen den Vorteil auf, dass sie bei der Durchführung und Auswertung relativ wenig Zeit beanspruchen und dennoch einen Einblick in die Wissensstruktur der Lernenden ermöglichen, sofern sie entsprechend konzipiert sind und den Standards der Testkonstruktion entsprechen.

Üblicherweise bestehen Konzepttests aus Single- oder Multiple-Choice-Items, bei denen die Teilnehmer aus mehreren vorgegebenen Antwortmöglichkeiten diejenige(n) auswählen, die sie als zutreffend befinden (siehe Lindell, Peak & Foster, 2007; Liu, 2012 im Überblick). Neben der korrekten Antwortalternative erfordert die Entwicklung der Items also auch die Formulierung geeigneter, d. h. „plausibler“ Distraktoren. Diese Distraktoren basieren in der Regel auf weitverbreiteten fehlerhaften Schülervorstellungen (vgl. ebd.). Gemeint sind hierbei Vorstellungen, „die einem wissenschaftlichen Verständnis entgegenstehen oder so unklar und vage sind, dass sie einer fachlichen Präzisierung bedürfen“ (Hettmannsperger, 2015, S. 67).

Für einige Themengebiete der Physik existieren bereits Konzepttests. Der wohl bekannteste und auch am häufigsten verwendete Konzepttest ist das Force Concept Inventory (FCI) von Hestenes, Wells & Swackhamer (1992), welches das Verständnis von Konzepten der klassischen Newtonschen Mechanik erfasst. Für den FCI besteht auch eine Fassung für den deutschen Sprachraum (Gerdes & Schecker, 1999). Andere Testinstrumente für den Bereich der Mechanik und Kinematik sind die Force and Motion Conceptual Evaluation (FMCE) und der Energy and Motion Conceptual Survey (EMCS). Yeo und Zadnick (2001) entwickelten für den Bereich der Wärmelehre den Introductory Thermal Concept Inventory (ITCI), der von Engelke ins Deutsche übersetzt wurde. Des Weiteren existieren im englischsprachigen Raum Konzepttests für die folgenden Bereiche (zit. n. Lindell, Peak & Foster, 2007): Astronomie (Astronomy Diagnostic Test, ADT), Elektrizität und Magnetismus (Brief

Electricity and Magnetism Assessment, BEMA; Conceptual Survey in Electricity and Magnetism, CSEM; Diagnostic Exam Electricity and Magnetism, DEEM; Determining and Interpreting Resistive Electric Circuits Test, DIRECT) sowie für Konzepte im Bereich der Wellenlehre (Wave Concept Inventory, kurz WCI).

Auch im Bereich der (Strahlen-) Optik liegen eine Reihe von Wissens- und Verstehensfragen sowie einige Tests vor, die auf die Erfassung von konzeptuellem Wissen und Verständnis in der (Strahlen-) Optik zielen und hierbei weitverbreitete Schülervorstellungen und Lernschwierigkeiten einbeziehen. Für die folgende Diskussion und den darauf aufbauenden Konzepttest wurde eine systematische Recherche mit Fokus auf das Thema Lichtausbreitung, Bildentstehung und Strahlenmodell/Bildkonstruktion durchgeführt (Bibliographie von Duit (2009), Recherche in ERIC, Google Scholar und andere Datenbanken).

Wiesner (1986; 1992a, 1992b) entwickelte mehrere Items zur gerichteten Reflexion, die auf ein adäquates Verständnis der geradlinigen Lichtausbreitung und der physikalischen Sehvorstellung zielen, nach der Licht vom Gegenstand in das Auge des Betrachters fallen muss.

Eine Zusammenstellung von Testaufgaben auf Basis der Arbeiten von Jung (1981) und Wiesner (a. a. O.) findet sich in der Arbeit von Herdt (1990), allerdings ohne weitergehende psychometrische Charakterisierung.

Ebenfalls kann eine solche Zusammenstellung, ergänzt um einige weitere Aufgaben zum Thema gerichtete Reflexion, geradlinige Lichtausbreitung und Streuung von Schecker (ohne Jahresangabe) auf der Homepage des Instituts für Didaktik der Naturwissenschaften der Universität Bremen abgerufen werden (<http://www.idn.uni-bremen.de/schuelervorstellungen/>).

Fetherstonhaugh und Treagust (1992) entwickelten einen Fragebogen über „Licht und seine Eigenschaften“ (light and its properties), der aus 16 Items besteht, wobei 12 Items im Multiple-Choice-Format mit drei oder vier Distraktoren vorliegen. Die übrigen Items nutzen offene Antwortformate (vgl. ebd., S. 657). Die Autoren berichten Ergebnisse von 83 Mittelstufenschüler/-innen aus städtischen und ländlichen Gebieten Australiens, wobei die Itemschwierigkeiten und Mittelwertvergleiche in Form von t-Tests zwischen den aus der Stadt und vom Land stammenden Stichproben berichtet werden. Die Angabe psychometrischer Eigenschaften wie weiterer Itemstatistiken oder Angaben zur Prüfung der Dimensionalität des Testinstruments bleiben jedoch aus.

Sokoloff (2006) erstellte im Rahmen des Projekts „Active Learning in Optics and Photonics“ (ALOP) Multiple-Choice-Items zur Strahlen- und Wellenoptik. Das Manual, in dem die Items publiziert sind, enthält zwar eine Anleitung zum Einsatz des Tests in Schule und Unterricht (Sokoloff, 2006, S. 227) jedoch keine Angaben zu empirischen Kennwerten der Items.

Bardar, Prather, Brecher und Slater (2007) entwickelten ein Testinstrument inklusive psychometrischer Angaben zu den entwickelten Aufgaben zur Erfassung von Konzepten im Bereich von Licht und Spektroskopie, den „Spectroscopy Concept Inventory (LSCI)“. Dieser Test wurde mit 548 Studienanfängern im Bereich der Astronomie erprobt und erfragt optische Konzepte, die weit über

die Inhalte der Mittelstufe hinausgehen, wie das Verständnis von Licht als elektromagnetischer Welle und seinen diesbezüglichen Eigenschaften, schwarzen Strahlern und der Dopplerverschiebung.

Chu, Treagust und Chandrasegaran (2009) entwickelten ebenfalls einen Multiple-Choice-Test, der darauf zielt Grundkenntnisse zur Lichtausbreitung bei Tag und Nacht, Lichtquellen, Wahrnehmungsprozesse und Sehen zu erheben, wobei der Test mit zweiteiligen Items arbeitet („two-tier instrument“). Dabei wird außer nach dem Vorliegen bestimmter Vorstellungen auch nach einer Begründung gefragt (mit aus der Literatur bekannten Prä- und Fehlkonzepten als Distraktoren). Die Autoren berichten einen generellen Wert der internen Konsistenz (Cronbachs α) von $\alpha_c = .65$ und geben den jeweiligen prozentuellen Anteil von insgesamt 1775 koreanischen Schüler/-innen für die Jahrgangsstufen 8 bis 10 an, welche die Items korrekt beantwortet haben (ebd., S. 258). Des Weiteren untersucht die Studie mittels multivariater Varianzanalyse und Korrelationen den Zusammenhang zwischen den Testwerten und der Einstellung zu naturwissenschaftlichen Konzepten, welche durch den Science Attitude Questionnaire (SAQ) (ebd., S. 262) erfasst wurden. Die Autoren berichten einen mittleren Zusammenhang von $r = .29$. Angaben zu den korrigierten Trennschärfen und eine Analyse der Dimensionalität erfolgen nicht.

Haagen-Schützenhöfer und Hopf (2018) schreiben ebenfalls, dass für den Bereich der Optik kein weitverbreitetes Standardinstrument publiziert wurde. Die genannten Autoren berichten auch über die Entwicklung eines Testinstrumentes mit zweiteiligen Aufgaben (Haagen-Schützenhöfer und Hopf 2012, 2014a-c). Die Gesamtreliabilität des Instrumentes wird als zufriedenstellend berichtet ($\alpha_c = 0.77$; Haagen-Schützenhöfer und Hopf, 2014a), allerdings ist das Instrument bislang nicht publiziert, sondern nur einzelne Items (z.B. Haagen-Schützenhöfer und Hopf, 2014b zu Lichtausbreitung und Sehvorgang). Auch eine Item-Analyse liegt in den veröffentlichten Arbeiten nicht vor.

Zusammenfassend kann festgehalten werden, dass etliche Autoren intensive Arbeiten für die Entwicklung von Items im Bereich der Strahlenoptik geleistet haben. Bislang fehlt jedoch ein publiziertes Testinstrument mit psychometrischer Charakterisierung, mit dem verschiedene Wissensfacetten im Bereich der Strahlenoptik in ökonomischer und psychometrisch zufriedenstellender Weise geprüft werden können. Das hier vorgestellte Testinstrument „KTSO-A“ (Konzepttest Strahlenoptik – optische Abbildungen) zielt darauf, dass konzeptuelle Wissen und Verständnis von Mittelstufenschülerinnen und -schülern im Bereich der Strahlenoptik mit dem Fokus optische Abbildungen in ökonomischer Weise zu erheben und soll einen Beitrag dazu leisten, diese Lücke zu schließen. Die Testdurchführung erfordert maximal 15 Minuten, die Auswertung erfolgt auf Basis dokumentierter Lösungen; eine Lösungsschablone kann auf Anfrage zur Verfügung gestellt werden. Die Formulierung der Distraktoren basiert hierbei auf Schülervorstellungen (vgl. Liu, 2012; Lindell, Peak und Foster, 2007, S. 3), die in empirischen Studien berichtet wurden. Wir stellen deshalb im Folgenden Ergebnisse aus Forschungsarbeiten für den hier interessierenden Inhaltsbereich zusammen.

2.2 Empirische Befunde und Recherche

Fehlerhafte Schülervorstellungen im Kontext der Strahlenoptik und Mittelstufenphysik sind unter anderem für folgende Wissensbereiche dokumentiert: Vorstellungen zum physikalischen Sehvorgang, Entstehung von Licht und Schatten, Farben, Entstehung von Spiegelbildern und Entstehung reeller und virtueller Bilder bei der Sammellinse sowie allgemeine ontologischen Grundannahmen, dazu was „Licht“ unter physikalischer Perspektive sein könnte (vgl. Hettmannsperger, 2015, S. 107 f.). Eine der ersten und maßgeblichen Studien zu naiven Vorstellungen über Licht stammt von Guesne (1985). Sie interviewte 30 Jugendliche im Alter von 13 und 14 Jahren, die zuvor keinen Optikerunterricht in der Schule erhalten hatten, in standardisierten Interviews. Im Ergebnis zeigte sich, dass die befragten Jugendlichen Licht häufig mit seiner Quelle oder seinen Wirkungen gleichsetzten. So lokalisierten die Teilnehmenden Licht in der Quelle, z. B. in der Glühlampe. Die Gleichsetzung mit einer Wirkung fand sich beispielsweise in der Vorstellung, man könne Licht nur an Stellen sehen, auf denen das Licht auf der Wand helle Flecken erzeugt (z. B. durch Sonnenlicht oder durch die Reflexion eines Spiegels). Als zentrales Forschungsergebnis hob die Autorin hervor, dass kaum einer der Befragten erkannte, dass Gegenstände, die nicht selbst leuchten, unter der Beleuchtung durch Tageslicht oder Lampen Licht zurückwerfen. Daraus resultieren Defizite beim Verständnis des Sehvorgangs. Da die befragten Kinder im Alltag Licht nur erkennen, wenn es einen deutlich wahrnehmbaren Effekt hervorbringt, glaubten sie nicht, dass auch bei nicht-selbst-leuchtenden Körpern Licht ins Auge gelangt (vgl. Guesne, 1985, S. 91).

Auf Basis dieser und ähnlicher Beobachtungen stellte Wiesner in einem mehrjährigen Forschungs- und Entwicklungsprogramm weitere zentrale Lernschwierigkeiten in der Strahlenoptik zusammen (Wiesner, 1988; 1992a; 1992b; 1994). Dabei wird insbesondere der enge Zusammenhang zwischen einer fehlerhaften physikalischen Sehvorstellung und fehlerhaften Konzepten zur Streuung deutlich (Wiesner, 1992a): So gehen viele Schülerinnen und Schüler (wie in der Studie von Guesne, s. o.) davon aus, dass beleuchtete Gegenstände wie Tische, Bücher oder Bilder kein Licht abstrahlen (Wiesner, 1986). Auch hier fehlt ein klares Gesamtbild von der Lichtausbreitung von der Quelle über die Streuung an beleuchteten (und daher sichtbaren) Gegenständen bis zum Auge des Beobachters (von Wiesner „Sender-Empfängervorstellung“ genannt). In den genannten Arbeiten wird insbesondere herausgearbeitet, dass ein solches Gesamtbild notwendige Voraussetzung für das Verständnis von Strahlengangkonstruktionen ist, wie sie z. B. bei der in dieser Arbeit interessierenden optischen Abbildung zugrunde liegen. Die genannte Gruppe von Lernschwierigkeiten schließt das Fehlkonzept ein, das auftreffende Licht mache die Gegenstände hell, bleibe auf diesen liegen oder verschwinde allmählich (ebd.). Ebenfalls damit verbunden ist die Vorstellung, dass von Lichtquellen mit geringer Intensität, wie Räucherstäbchen oder weit entfernte beleuchtete Fenster, kein Licht ins Auge gelangt (ebd.). Spätere Studien bestätigen, dass die physikalische Erklärung des Sehvorgangs vielen Schüler/-innen Schwierigkeiten bereitet (Selley, 1996; Chu et al., 2009).

Die Schwierigkeiten mit der physikalischen Sehvorstellung schlagen sich nach Wiesner (1992a) auch auf Lernschwierigkeiten mit dem Spiegelbild nieder. Auch hier erkennen viele Schülerinnen und Schüler nicht, dass das Licht aus der Richtung des Spiegels ins Auge fallen muss, damit das Spiegelbild wahrgenommen werden kann. Schülerinnen und Schülern bereitet aber nicht nur die Erklärung der Wahrnehmung, sondern auch die Lage des Spiegelbildes große Probleme: So werden die strahlengeometrische Konstruktion des Spiegelbildortes von den meisten Schülerinnen und Schülern als nicht nachvollziehbar eingestuft. Die Schülerinnen und Schüler gehen oft davon aus, dass das Spiegelbild auf der Spiegeloberfläche liegt (vgl. Wiesner, 1986, S. 26, S. 27.; 1992a; 1992b, S. 288). Der Spiegel wird auch oft als ein Gegenstand aufgefasst, der das Spiegelbild zum Betrachter zurückwirft. Auch mit der Entstehung von reellen Bildern durch die Sammellinse ist gemäß Wiesner (1994, S. 8) eine Reihe von Fehlvorstellungen verbunden: Viele Schülerinnen und Schüler nutzen zur Erklärung der Entstehung des reellen Bildes bei der Sammellinse nicht das Konzept einer Punkt-zu-Punkt-Abbildung. Zu den gängigen Vorstellungen zählt: Das Bild geht als Ganzes durch die Linse zum Schirm und wird dabei in der Linse umgedreht (Wiesner, 1994, S. 8). Wiesner bezeichnet dieses weitverbreitete Konzept als holistische Erklärung des Abbildungsvorgangs. Dass Lernende auf eine solche holistische Erklärung des Abbildungsvorgangs zurückgreifen, wird insbesondere bei Abdeckaufgaben deutlich. Unter der Annahme, dass das Bild als Ganzes vom Gegenstand aus durch die Linse auf den Schirm transportiert wird, ist es nur konsequent anzunehmen, dass ein Teil des Bildes abgeschnitten wird, wenn man eine Blende vor die Linse hält (vgl. Wiesner, 1992b, S. 288). Hält man eine ringförmige Blende vor die Linse, glauben viele Lernende entsprechend, dass das Bild kreisförmig am äußeren Rand abgeschnitten wird. Wird die Linse zur Hälfte abgedeckt, gehen viele Schülerinnen und Schüler aber auch Studierende davon aus, dass auch das reelle Bild zur Hälfte abgeschnitten wird, einige überlegen sich sogar, welche Hälfte des Bildes (obere versus untere Hälfte) betroffen ist (vgl. Goldberg & McDermott, 1987, S. 112; Wiesner, 1994, S. 8).

Weitere Vorstellungen, die ebenfalls das Konzept der Punkt-zu-Punkt-Abbildung außer Acht lassen, bestehen in der Idee, die Linse konzentriert das Licht, oder hinter der Linse seien mehr Licht bzw. mehr Strahlen vorhanden als vor der Linse (vgl. Wiesner, 1986). Häufig wird die Entstehung reeller Bilder durch Spiegelung und Reflexion erklärt, dabei wird einem Gegenstandspunkt in der Regel nur ein Strahl zugeordnet und nicht ein divergierendes Strahlenbündel (vgl. ebd.).

Goldberg und McDermott (1987) berichten über das Verständnis der Entstehung reeller Bilder durch Sammellinse und Hohlspiegel von der folgenden weiteren Verständnisschwierigkeit bei Collegestudent/-innen (80 Studierende aus einem Einführungskurs zur Physik). Diesen wurde ein Versuchsaufbau gezeigt, bei dem eine Glühlampe, eine Linse und ein Schirm hintereinander auf einer

optischen Bank montiert sind. Im Verlauf des Interviews wurden die Studierenden gefragt, wo das Bild wäre, wenn man den Schirm entfernt und sie frei um den Versuchsaufbau im Raum herumgehen können (natürlich wird hier von den Autoren der physikalische Bildbegriff als Schnittpunkt von Lichtstrahlen und als Ort des Luftbildes bei der reellen Abbildung vorausgesetzt, siehe z. B. Meschede (2006, S. 484)). Nur wenige Studierende waren in der Lage zu erkennen, dass sich das Bild an der gleichen Position befindet wie zuvor der Schirm. Die übrigen Studierenden gaben eine Erklärung ab wie etwa, dass sich das Bild auf oder in der Linse befindet. Insbesondere war die Vorstellung verbreitet, dass ein Bild nur mit Hilfe eines Schirms gesehen werden kann und dass die Linse das Bild quasi einrahmt (vgl. Goldberg & McDermott, 1987, S. 114). Auch Guesne (1985) hatte bereits in ihrer oben genannten Studie Kinder zur Rolle der Sammellinse in der Funktion als Lupe und Brennglas befragt. Dabei konnte sie zwei Antworttypen auffinden: Die erste Hälfte der Antworttypen beinhaltet die Ansicht, das Vergrößerungsglas „vermehrte“ das Licht, während die andere Hälfte der Antworten die Überzeugung umfasst, dass die Sammellinse das Licht konzentriert. Kinder, welche das Konzept der Lichtkonzentration vertraten, waren der Ansicht, die gesamte Lichtmenge, die durch das Vergrößerungsglas hindurchgehe, bleibt hinter der Linse erhalten, was wissenschaftlich korrekt ist. Dass auch diese Gruppe von Kindern nicht notwendigerweise eine physikalisch angemessene Vorstellung von der Funktionsweise hat, zeigte sich in den Zeichnungen der Kinder, mit denen sie verdeutlichen sollten, wie die Linse das Licht konzentriert (vgl. Guesne, 1985, S. 87). Kinder, die der Ansicht waren, das Vergrößerungsglas „vermehrte“ das Licht, gaben entweder an, dass hinter der Lupe mehr Licht ist als vor der Lupe oder dass das Licht hinter der Lupe verstärkt bzw. vermehrt wird.

Die Literatur zeigt demnach, dass Lernende Vorstellungen von Licht besitzen, welche sich grundlegend von einem wissenschaftlichen Verständnis unterscheiden. Dies offenbart sich insbesondere auch in substanzbasierten Vorstellungen von Licht. Reiner et al. (2000, S. 14, 15) beschreiben eine solche Konzeption wie folgt: Wenn Lernende gefragt werden, wie Sehen funktioniert, gaben sie an, dass Moleküle, im Sinn von Partikeln, zwischen dem gesehenen Gegenstand z. B. einem „Buch“ und dem „Auge“ vorhanden seien. Der Schvorgang wird demgemäß als das Ergebnis beweglicher Lichtpartikel interpretiert. Das scheint zwar im Einklang mit dem Konzept von Photonen zu sein, Licht wird dabei aber oft als Flüssigkeitsstrom beschrieben, der in Bewegung ist, sich aber auch in Ruhelage befinden kann. Damit gehen Schwierigkeiten einher, Lichtstrahlen als Modellvorstellung zu erkennen und zwischen den Begriffen Lichtbündel und Lichtstrahl differenzieren zu können. Der Vollständigkeit halber sei hier auch auf neuere Arbeiten zum konzeptuellen Verständnis von Farben hingewiesen (Martinez-Borreguero et al., 2013; Naranjo-Correa et al., 2015); dieses Gebiet liegt jedoch außerhalb der Domäne des hier vorgestellten Tests.

Tab. 1. Angaben zu Items des KTSO-A

Konzept	Distraktoren: Schülervorstellungen mit Beispielen	Item/s	Quellen
<i>Sekundäre Lichtquellen und physikalische Sehvorstellung, Streuung</i>	Unterscheidung primäre und sekundäre Lichtquellen In einem vollständig abgedunkelten Raum sind helle Gegenstände sichtbar. Bei Lichtquellen mit geringer Intensität gelangt kein Licht mehr ins Auge.	1	Wiesner, 1992a, S. 16
<i>Lichtausbreitung</i>	Streuung und (diffuse) Reflexion an Oberflächen Beleuchtete Gegenstände wie Tische, Bücher oder Bilder strahlen kein Licht ab / beleuchtete Gegenstände werfen kein Licht zurück. Licht, das auf eine Oberfläche fällt, bleibt auf dieser liegen und macht sie hell.	2, 4, 5	Wiesner, 1986, S. 26, 27
<i>Ontologisches Verständnis von Licht (-strahlen) und Modell-Kompetenz</i>	Licht und Lichtstrahlen als ontologische Kategorie Licht wird als Substanz aufgefasst: Licht wird als Flüssigkeitsstrom beschrieben, der in Bewegung ist, sich aber auch in Ruhelage befinden kann. Verwechslung von Lichtbündel und Lichtstrahl	3	Reiner et al., 2000, S. 15ff.
<i>Entstehung reeller Bilder bei der Sammellinse</i>	Linsenabbildung / Bildentstehung Das reelle Bild bei der Sammellinse entsteht durch Spiegelung oder Reflexion. Holistische Konzeption des Abbildungsvorgangs: das Bild geht als Ganzes durch die Linse zum Schirm und wird dabei in der Linse umgedreht. Die Linse konzentriert oder „vermehrt“ Licht	6,7	Abbildungen zu Item 6 und 7 in Anlehnung an Wiesner, 1986, S. 28
<i>Entstehung reeller Bilder bei der Sammellinse</i>	Fehlvorstellung „Blende schneidet Bild ab“ Bei Abdeckung mit einer ringförmigen Blende werden die Bildränder kreisförmig abgeschnitten. Bei Abdeckung der oberen bzw. unteren Hälfte der Linse wird entsprechend die obere bzw. untere Hälfte des Bildes abgeschnitten.	8a, 8b, 8c	weitere Beschreibung der Schülervorstellungen Wiesner, 1994, S. 8

2.3 Rahmen und inhaltliche Spezifikation

Auf der Basis der zuvor dargestellten Arbeiten zu Schülervorstellungen im Bereich der Strahlenoptik zielt das hier vorgestellte Testinstrument darauf, das konzeptuelle Verständnis im Bereich der Strahlenoptik mit dem Schwerpunkt optische Abbildungen bei der Sammellinse und hiermit verbundener physikalischer Grundkonzepte zu erfassen. Der Test wurde im Rahmen des Projekts „Das Experiment als Mittel zur Entwicklung von Repräsentationskompetenz im Rahmen einer problemorientierten Aufgabenkultur des Physikunterrichts“ (2009 bis 2012) des DFG-Graduiertenkollegs „Unterrichtsprozesse“ (GK 1561) der Universität Koblenz-Landau am Campus Landau entwickelt, um mögliche Lernfortschritte in Bezug auf Grundkonzepte in der Strahlenoptik in zwei aufeinander bezogenen Interventionsstudien Scheid (2013) und Hettmannsperger (2015) analysieren zu können. Die verwendeten Items erfassen Grundkonzepte aus den Bereichen Lichtausbreitung, Streuung, Reflexion und Linsenabbildung (vgl. Tab. 1). Andere Bereiche, wie z. B. Licht und Farben oder Licht und Schatten sowie Spiegelbilder, welche für das Verständnis optischer Abbildungen keine oder eine vergleichsweise untergeordnete Rolle spielen, wurden nicht in den Test aufgenommen. Wie die Literatur und die nachfolgende vorgestellte empirische Analyse (siehe Kapitel 3 und 4) zeigen, umfasst der Phänomenbereich *optische Abbildung bei der Sammellinse* (also das Zustandekommen eines reellen Bildes) einen hinreichend großen und inhaltlich zusammenhängenden Teil der Strahlenoptik, um die Entwicklung eines Konzept-Tests für diesen Bereich zu rechtfertigen (siehe auch Hettmannsperger, 2015, S. 129-131).

Das Vorgehen der Itemkonstruktion entsprach im ersten Schritt der rationalen Testkonstruktion (vgl. Bühner,

2011, S. 93) und der Literatur zur Standards bei der Entwicklung von Konzepttests und anderen diagnostischen Tests (Adams & Wiemann, 2011, Madsen et al, 2017): Auf Basis vorhandener Literatur zu Schülervorstellungen und Lernschwierigkeiten und in der Literatur beschriebener Aufgaben wurden zu jedem jeweilig gewünschten wissenschaftlichen Konzept, eine korrekte Antwortalternative und Distraktoren entwickelt (vgl. Tabelle 1). Die Konstruktion der Distraktoren folgt daher forschungsgeleitet (vgl. Haladyna & Downing, 1989, S. 68 zur methodischen Begründung des Vorgehens und Liu, 2012; Lindell, Peak und Foster, 2007, S. 3 zum Vergleich mit Arbeiten anderer Autoren). Aus diesem Arbeitsschritt resultierten zunächst 21 Items, die einer empirischen Überprüfung (Itemanalyse und Exploratorische Faktorenanalyse) unterzogen und eliminiert wurden, wenn die Kennwerte außerhalb der etablierten Akzeptanzbereiche lagen (Ding & Beichner, 2009). Detaillierte Angaben zu diesem Überarbeitungsprozess finden sich in Hettmannsperger (2015, S. 132 f.). Die resultierende Testfassung enthält 10 Items mit zufriedenstellenden bis guten psychometrischen Eigenschaften, die im Folgenden berichtet werden.

3 STICHPROBE UND METHODEN

3.1 Stichprobe

Der Test wurde mit $n = 869$ Schüler/-innen der Klassenstufen 7 und 8 aus 38 Klassen an 15 Gymnasien und 3 Gesamtschulen erprobt (vgl. Tabelle 2); die Schüler/-innen waren im Schnitt 13 Jahre alt ($M = 13.4$ Jahre, $SD = 0.7$). Alle Schüler/-innen erhielten erstmals Unterricht zum Thema „Strahlenoptik“. Zur Itemanalyse wurden Daten aus den zwei oben genannten Studien genutzt (Hettmannsperger, 2015; Scheid, 2013). Dabei war mit Blick auf die bekanntermaßen hohen kognitiven Anforderungen

der Nutzung von multiplen Repräsentationen eine Entscheidung für eine mehrheitlich aus dem Gymnasium stammenden Stichprobe getroffen worden, um zunächst bei Schülern mit günstigen Lernvoraussetzungen die Wirksamkeit der untersuchten Interventionen zu erheben. Weitergehende Angaben zu der Art der Interventionen (die hier nicht Gegenstand sind) und zur Stichprobe finden sich in Hettmannsperger (2015, S. 143 f.) und Scheid (2013, S. 93 f.; Scheid et al., 2019).

Die Teilstichproben werden in Abhängigkeit davon, ob eine gezielte Intervention für die Förderung des konzeptuellen Verständnisses und eine Überwindung von Lernschwierigkeiten und Fehlvorstellungen im Themenbereich des Tests stattfand, wie folgt unterschieden: Nur in Teilstichprobe A fand die vorgenannte Intervention statt, in Teilstichprobe B fand eine andere Intervention statt (vgl. Scheid et al., 2019). Gemäß Bühner (2011, S. 81) kann die Größe der Stichproben als gut bis sehr gut angesehen werden.

Tab. 2. Stichprobenangaben

Level	N Gesamt	n Studie A	n Studie B
Schulen	18	10	8
Klassen	38	21	17
Schüler/-innen	869	480	389

Studie A – Hettmannsperger, 2015; Studie B - Scheid, 2013

Stichprobe A (Hettmannsperger, 2015): Der Unterricht zielt in beiden Bedingungen (Treatment- und Kontrollgruppe) auf die Förderung des konzeptuellen Verständnisses und auf eine Überwindung von Lernschwierigkeiten und Fehlvorstellungen im Themenbereich dieses Artikels. Treatment- und Kontrollgruppe (TG und KG) dieser Studie unterschieden sich dahingehend, ob die Schülerinnen und Schüler kognitiv aktiviert wurden sich mit multiplen Repräsentationen auseinander zu setzen (TG) oder keine solche auf multiple Repräsentationen gezielte kognitive Aktivierung stattfand (KG), während sie in beiden Gruppen Aufgaben bearbeiteten, die Schülervorstellungen thematisierten. Da zwischen den beiden Bedingungen keine signifikanten Unterschiede bestanden (Hettmannsperger, 2015, S. 146 f.), werden im Folgenden die Werte der Stichprobe A insgesamt berichtet und nicht getrennt nach Bedingungen.

Teilstichprobe A kann also exemplarisch zeigen, ob der Test in der Lage ist, die Wirksamkeit einer auf die entsprechenden Lernschwierigkeiten/Fehlkonzepte zielenden Intervention nachzuweisen. Außerdem dient diese Stichprobe dazu, eine etwaige Substruktur des Tests per Kreuzvalidierung zu prüfen (s. Abschnitt 3.4).

Stichprobe B (Scheid, 2013): Der Unterricht in dieser Stichprobe enthielt keine gezielte Intervention zur Überwindung von Lernschwierigkeiten und Fehlvorstellungen im Themenbereich dieses Artikels. Gegenstand des Vergleichs zwischen TG und KG war vielmehr die Förderung der Fähigkeit, Kohärenz zwischen mehreren Repräsentationen herzustellen (siehe auch Scheid et al., 2019). Da zwischen den beiden Bedingungen bezüglich des Konzeptverständnisses keine signifikanten Unterschiede bestanden (Scheid, 2013) werden im Folgenden die Werte

der Stichprobe B ebenfalls gesamt berichtet und nicht getrennt nach Bedingungen.

Teilstichprobe B entspricht also der üblichen Situation für eine Vergleichsstichprobe für Konzept-Tests: nach Unterricht, aber ohne gezielte Maßnahmen hinsichtlich der fraglichen Lernschwierigkeiten und Schülervorstellungen.

3.2 Durchführung und empirisches Vorgehen

Die vorgestellten Ergebnisse basieren auf den Daten von Schülerinnen und Schülern, die erstmals in Klasse 7 oder zu Beginn von Klasse 8 Optik Unterricht zum Thema Strahlenoptik erhalten hatten (s. o.). Die Darstellung der Ergebnisse dieses Messzeitpunkts wurde gewählt, da sie dem Kenntnisstand der Schüler/-innen der Mittelstufe nach Unterricht (laut Lehrplan des entsprechenden Schultyps) entspricht. Die vollständige Testfassung ist in Appendix 2 dieses Artikels abgedruckt. Die Schülerinnen und Schüler benötigten in der Regel nicht mehr als 10 – 15 Minuten, um den Test zu bearbeiten. Die Schüler/-innen wurden in einer standardisierten kurzen Testinstruktion darauf hingewiesen, dass etwaige Abbildungen Bestandteil der Aufgaben seien, und für eine Lösung sorgfältig zu beachten seien, und dass manchmal entweder nur eine Antwort oder manchmal mehrere Antworten richtig sein können.

Die Bewertung der Aufgaben wurde wie folgt vorgenommen: Die Antworten wurden mit 2 Punkten bewertet, wenn ausschließlich die korrekte Alternative gewählt wurde, ein Punkt wurde vergeben, wenn zusätzlich zu der richtigen Antwort auch ein Distraktor angekreuzt wurde (2 Punkte für die Lösung weniger 1 Punkt Abzug). In allen anderen Fällen wurde die jeweilige Antwort mit 0 Punkten bewertet.

Im Sinn der klassischen Testtheorie (vgl. Lienert & Raatz, 1998; Rost, 2004; Bühner, 2011) wurden Itemkennwerte wie Lösungswahrscheinlichkeit und korrigierte Trennschärfe sowie interne Konsistenz (α_c) als Schätzer für die Reliabilität berechnet. Des Weiteren wurde eine deskriptive Analyse der Distraktoren durchgeführt. In Anbetracht der Anzahl der Distraktoren ($k = 3$) (Wilcox, 1981) und dem Messzeitpunkt (das Verständnis der Schülerinnen und Schüler wurde nach dem Unterricht erfasst), kann für einen gegebenen Distraktor eine Untergrenze von mindestens 5 % an Antworten als zufriedenstellend betrachtet werden (vgl. auch Wakefield, 1958 zit. n. Haladyna & Downing, 1989, S. 58; Kline, 2005, S. 57).

Unter Berücksichtigung des Skalenniveaus wurde für die Berechnung der Trennschärfen und α_c polychorische bzw. polyseriale Korrelationen verwendet (vgl. Eid, Gollwitzer & Schmitt, 2011, S. 538). Die Berechnung erfolgte mit Hilfe des R-Pakets „polycor“ (Fox, 2010). Polyseriale Korrelationen sind für die Berechnung der Korrelation zwischen ordinalskalierten und intervallskalierten Daten die Methode der Wahl (vgl. Eid et al., 2011, S. 538), polychorische Korrelationen für die Berechnung von Korrelationen zwischen ordinalskalierten Daten untereinander (vgl. ebd., 2011, S. 515).

Zur Schätzung, wie gut sich die jeweiligen Items dazu eignen, zwischen Personen mit unterschiedlichen Fähigkeiten zu differenzieren, wurden die korrigierten Eigen-trennschärfen, d. h. die Korrelation des jeweiligen Items

mit der Gesamtpunktzahl des Tests (ausgenommen des jeweiligen Items selbst) berechnet. Da Varianzen bei Daten mit drei Abstufungen nur begrenzt interpretierbar sind, wurde das standardisierte Cronbach's Alpha verwendet, welches auf den durchschnittlichen Korrelationen zwischen den Items basiert (Bonanomi, Nai Ruscone & Osmetti, 2013).

Aus zwei Gründen wurde darauf verzichtet, den Test mittels eines Rasch-Modells zu skalieren: Zum ersten ist der Test so konzipiert, dass die Aufgaben der Skala Abdeckaufgaben sich auf den gleichen Experimentalaufbau beziehen, bei dem die Linse durch eine Blende (teilweise) abgedeckt wird. Aus diesem Grund ist die Voraussetzung der lokalen stochastischen Unabhängigkeit der Aufgaben nicht erfüllt (vgl. Strobl, 2012, S. 17). Gegen ein klassisches eindimensionales Raschmodell spricht des Weiteren zudem, dass der Test nicht eindimensional ist, da mindestens die Aufgaben zur Abdeckung der Sammellinse eine Subdimension des Tests bilden sollten.

Es wurde zunächst auch geprüft, ob der Test tatsächlich wie erwartet mehrdimensional ist, anschließend wurde das Instrument mittels exploratorischer und konfirmatorischer Faktorenanalyse kreuzvalidiert (Bühner, 2011; Anwendungen auf Konzepttests: Ramlo, 2008; Scott, Schuhmayer & Gray, 2012).

Wie bekannt, ermöglicht eine exploratorische Faktorenanalyse eine Datenstruktur zu finden (Bühner, 2011, S. 254). Deshalb wurde zunächst eine exploratorische Faktorenanalyse (EFA) auf Basis der Stichprobe B durchgeführt. Das Modell, welches sich aus der exploratorischen Faktorenanalyse ergab, wurde dann mittels der Methode der konfirmatorischen Faktorenanalyse anhand der Stichprobe A überprüft (ebd.). Abschließend wurden diese Schritte vice versa wiederholt, um sicherzustellen, dass sich für beide Teilstichproben die gleiche Teststruktur findet. Dies entspricht dem Verfahren der Kreuzvalidierung (ebd.).

Die exploratorische Faktorenanalyse wurde mit der freien Statistiksoftware R unter Verwendung der Pakete psych (Revelle, 2013), lavaan (Rosseel, 2012) und qgraph (Epskamp et al., 2012) durchgeführt. Der Vorteil der Statistiksoftware R besteht darin, dass sie die Möglichkeit bietet, der Analyse eine polychorische Korrelationsmatrix für ordinalskalierte Daten zugrunde zu legen. Eine Normalverteilung der Daten ist hierbei keine zwingende Voraussetzung (Bühner, 2011).

Abschließend wurde ein Expertenrating mit 11 Experten aus dem Bereich Physikdidaktik zur Beurteilung der internen Validität des Tests durchgeführt und die Korrelationen mit den Schulnoten in Mathematik und Physik angegeben. Weitere Angaben zur Expertenstichprobe und den Ratingfragen finden sich in Abschnitt 4.3.

4 ERGEBNISSE

4.1 Item- und Teststatistiken

Die Analyse der Itemschwierigkeiten (vgl. Abb. 1 und Tab. 3) ergab, dass für die Gesamtstichprobe und die

Stichprobe A alle Items innerhalb des Toleranzbereichs von $0.20 \leq P_i \leq 0.80$ liegen (Kline, 2015). In der Stichprobe B bestehen kleinere Abweichungen für Item 2, das Item ist etwas zu „leicht“ sowie für die Items 8a-c, diese Items sind etwas zu schwer. Für die Trennschärfen ergeben sich in der Gesamtstichprobe Werte in einem Bereich zwischen $.23 \leq r_{ii} \leq .43$ (vgl. Abb. 2 und Tab. 3). Auch diese Werte liegen innerhalb des Toleranzbereichs von $r_{ii} \geq 0.2$ (Kline, 2015). Für A sind die Werte durchwegs höher, für B liegen die Trennschärfen bei etlichen Items nur bei knapp 0.20 (s. Diskussion).

Auch die Werte für interne Konsistenz (α_c ; Tab. 3) als Schätzer für die Reliabilität des Tests können gemäß gängiger Richtlinien (COTAN-System, Evers, 2001; darauf beruhend: EFPA, 2013) als ausreichend für Gruppenvergleiche angesehen werden, der Wert für die Teilstichprobe A ist nahe dem Schwellenwert für Individualdiagnostik (Bühner, 2011; Ding & Beichner, 2009; Doran, 1980). Die Werte in Tabelle 3 belegen zudem, dass die interne Konsistenz weder für die Teilstichproben noch für die Gesamtstichprobe durch den Ausschluss weiterer Items angehoben werden könnte.

Die Distraktorenanalyse, welche in Tabelle 7 berichtet wird, zeigt auf, dass für jedes Item mehr als 5 % der Antworten auf mindestens einen der Distraktoren fällt. Unter Berücksichtigung der o. g. Schwelle für 5 % bei 3 Distraktoren ($k = 3$) (Wilcox, 1981; Haladyna & Downing, 1989, S. 58; Kline, 2005, S. 57) und dem Messzeitpunkt (das Verständnis der Schülerinnen und Schüler wurde nach dem Unterricht erfasst), muss keines der Items aufgrund eines zu selten gewählten Distraktors ausgeschlossen werden.

4.2 Analyse der Dimensionalität

Zur Analyse der Dimensionalität wurde zunächst faktorenanalytisch geprüft, ob alle Items auf einen Generalfaktor laden. Erwartungskonform zeigte sich, dass das Modell mit einem Generalfaktor verworfen werden musste ($\chi^2(119) = 1390$, $p < 0.001$; $CFI = 0.60$; $TLI = 0.54$; $RMSEA = 0.11$; $SRMR = 0.10$). Übereinstimmend mit den theoretischen Vorüberlegungen ist daher von einem mehrdimensionalen Konstrukt auszugehen. So war anzunehmen, dass mindestens die Aufgaben zur Abdeckung der Sammellinse eine eigene Skala bilden.

Um die Struktur des Tests genauer zu analysieren, wurde der Test nun einer Kreuzvalidierung unterzogen. Hierbei sollten sich für beide Stichproben A und B die gleichen Facetten bzw. Subdimensionen ergeben. Das Vorgehen der Kreuzvalidierung umfasste drei Schritte:

- 1) Es wurde eine exploratorische Faktorenanalyse mit Stichprobe B umgesetzt.
- 2) Das ermittelte exploratorische Modell wurde mittels einer konfirmatorischen Faktorenanalyse an Stichprobe A überprüft.
- 3) Die Schritte 1) und 2) wurden vice versa wiederholt.

Tab. 3. Itemkennwerte

Item		Schwierigkeit			Korrigierte Trennschärfe			Interne Konsistenz α_c unter Ausschluss des jeweiligen Items		
		A <i>n</i> = 480	B <i>n</i> = 389	Gesamt <i>N</i> = 869	A <i>n</i> = 480	B <i>n</i> = 389	Gesamt <i>N</i> = 869	A <i>n</i> = 480	B <i>n</i> = 389	Gesamt <i>N</i> = 869
y1	Sichtbare Gegenstände in dunklem Raum	0.74	0.81	0.77	0.28	0.24	0.23	.78	.62	.74
y2	Helligkeit eines Zimmers - Farbe der Tapete	0.70	0.63	0.67	0.28	0.18	0.25	.78	.63	.74
y3	Lichtstrahlen sind ...	0.42	0.39	0.41	0.30	0.20	0.27	.78	.63	.74
y4	Lichtstrahl in dunklem Raum	0.35	0.34	0.35	0.29	0.15	0.23	.78	.63	.74
y5	Lichtstrahl: Kreidestaub	0.68	0.76	0.71	0.41	0.30	0.33	.76	.58	.72
y6	Entstehung reeller Bilder an Sammellinse	0.43	0.35	0.39	0.26	0.19	0.25	.78	.62	.74
y7	Bildkonstruktion: Bildentstehung	0.40	0.29	0.35	0.34	0.22	0.32	.77	.62	.73
y8a	Abdeckung einer Linsenhälfte	0.55	0.18	0.38	0.47	0.17	0.40	.76	.62	.72
y8b	Ringförmige Abdeckung der Linse	0.40	0.18	0.30	0.48	0.23	0.43	.76	.60	.71
y8c	Enge ringförmige Abdeckung der Linse	0.39	0.17	0.29	0.48	0.19	0.42	.75	.61	.71

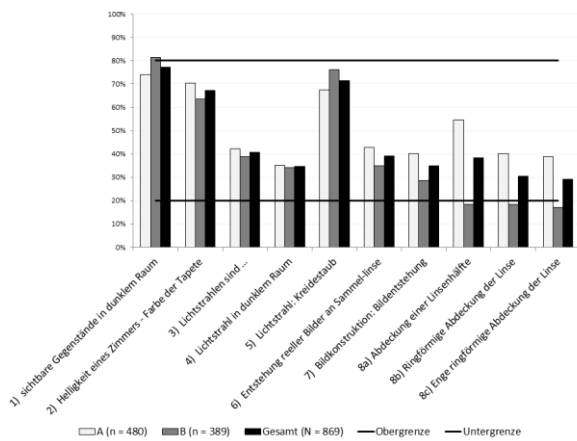


Abb. 1. Itemschwierigkeiten des KTSO-A je Stichprobe

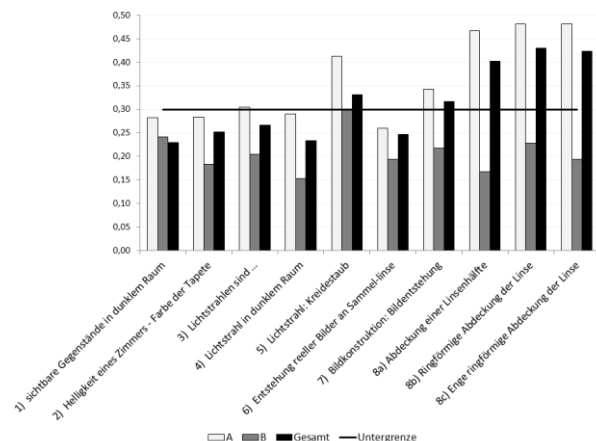


Abb. 2. Korrigierte Trennschärfen des KTSO-A je Stichprobe

Diese Abfolge sichert die Vorgehensweise insofern ab, als eine etwaige Teststruktur exploratorisch in der Teilstichprobe ohne gezielte Intervention (Stichprobe B, weniger ausgeprägte Wissensstruktur) gesucht wird und erst anschließend in Stichprobe A überprüft wird. Bei der Analyse wurde die Annahme zugrunde gelegt, dass die Faktoren untereinander korrelieren können, d.h. dass die Beantwortung von bestimmten Itemgruppen nicht notwendigerweise unabhängig von anderen Itemgruppen erfolgt. Daher wurde in der exploratorischen Faktorenanalyse gemäß der Empfehlung von Bühner (2011, S. 349) das oblique Rotationsverfahren „Promax“ verwendet. Die Prüfung der Voraussetzungen ergab, dass die Bewertungskriterien zur Durchführbarkeit der Analyse vollständig erfüllt sind (vgl. Bühner, 2011, S. 346). So belegt der Bartlett-Test, dass die polychorische Korrelationsmatrix signifikant von der Einheitsmatrix verschieden ist ($\chi^2 = 450, df = 45, p < 0.001$) und somit faktorisiert werden kann. Der Kaiser-Meyer-Olkin-Koeffizient weist darauf hin, dass sich die Itemauswahl für die Faktorenanalyse eignet und der empirisch vorgefundene Wert von 0.79 im Sinne der üblichen Qualitätsstandards in einem

guten Bereich liegt (vgl. ebd., S. 347). Auf Itemebene zeigen die MSA-Koeffizienten (Measure of Sample Adequacy) ebenfalls gute Werte $0.73 \leq MSA \leq 0.89$. Der Verlauf der Eigenwerte und das Ergebnis der Parallelanalyse nach Horn (1965) weisen darauf hin, dass die ersten drei Faktoren die Hauptinformationen der Daten repräsentieren (vgl. Abb. 3 und Tab. 5). Die Kommunalitäten liegen mit $0.11 \leq \hat{H} \leq 0.86$ in einem für die Stichprobe von $n = 389$ und drei bis vier Items pro Faktor ebenfalls in einem akzeptablen bis guten Bereich (vgl. Bühner, 2011, S. 345). Es zeigt sich, dass die Items 8 a-c einen mittleren, die Items 1, 4, 5, 7 einen mittleren bis hohen und die Items 2, 3 und 6 den geringsten Varianzanteil aufklären (siehe Tab. 4 und Tab. 7). Von den Faktorladungen nach Schritt 1) sind vier Werte gut oder besser, drei angemessen, zwei akzeptabel, und zwei zu klein (Tabachnick & Fidell, 2007). Es ergeben sich demnach drei Dimensionen für den untersuchten Test. In Abbildung 4 wird eine Darstellung der durch die Faktorenanalyse gefundenen Ergebnisse gege-

ben, die an die Darstellungen von Strukturgleichungsmodellen angelehnt ist, um den Bezug zur konfirmatorischen Faktorenanalyse in Schritt 2) darstellen zu können. Alle drei genannten Dimensionen tragen zum Verständnis der Bildentstehung mittels Sammellinse bei, was sich in den wechselseitigen mittleren Korrelationen ($.26 \leq r \leq .42$) zeigt (siehe Abb. 4).

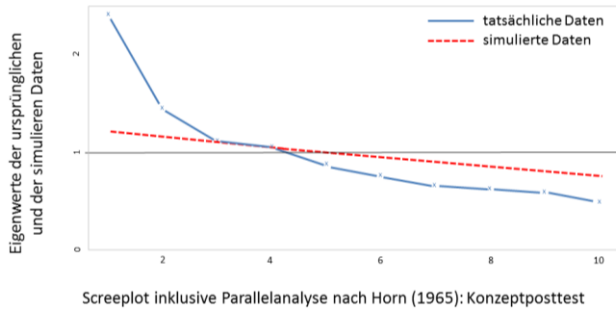


Abb. 3. Screeplot inklusive Parallelanalyse nach Horn (1965)

Insgesamt werden 33 % der Gesamtvarianz durch die drei Faktoren aufgeklärt. Im zweiten Schritt wurden die Ergebnisse der exploratorischen Faktorenanalyse mit der Stichprobe A ($n = 480$) Kreuzvalidiert.

Da es sich um ordinale Daten handelt, wurde zur Durchführung der konfirmatorischen Faktorenanalyse die ADF-Methode verwendet, welche keine multivariat normalverteilten Daten auf Itemebene voraussetzt (vgl. Bühner, 2011, S. 432). Des Weiteren wurde geprüft, dass keine Ausreißer vorliegen und keine Kollinearität der Items ($r_{in} = 389) < 0.76$) besteht (vgl. ebd.).

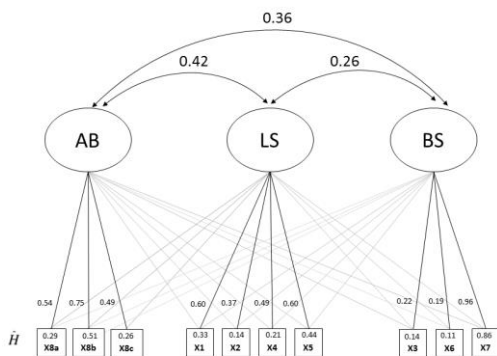


Abb. 4. Exploratorische Faktorenanalyse. AB = Bildentstehung inklusive Abdeckaufgaben, LS = geradlinige Lichtausbreitung und Streuung, BS: Bildkonstruktion / Strahlenmodell

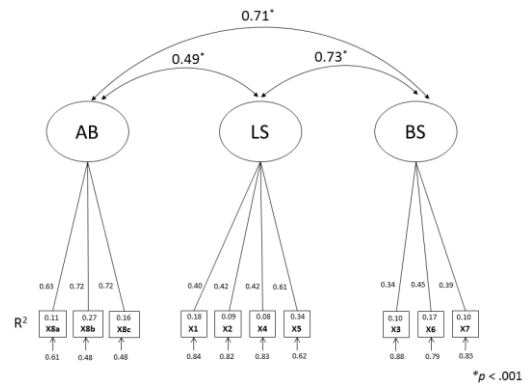


Abb. 5. Konfirmatorische Faktorenanalyse: AB = Verständnisfragen zur Bildentstehung inklusive Abdeckaufgaben, LS = geradlinige Lichtausbreitung und Streuung, BS: Verständnis Bildkonstruktion / Strahlenmodell.

Auf Basis des Modells, das sich an der exploratorischen Faktorenanalyse orientierte, werden die Items wie folgt zugeordnet (siehe Abb. 5): AB: Abdeckaufgaben mit Verständnisfragen zur Bildentstehung: y8a, y8b, y8c; LS: geradlinige Lichtausbreitung und Streuung: y1, y2, y4, y5 und BS: Verständnis zur Bildkonstruktion, zu Lichtquellen und zum Strahlenmodell: y3, y6, y7. Die Ergebnisse der konfirmatorischen Faktorenanalyse auf Basis von Stichprobe A bestätigen die vorgefundene Struktur, welche die exploratorische Faktorenanalyse aufzeigte.

Insgesamt ergaben sich gute globale Fit-Indizes ($\chi^2_{(32)} 44, p = 0.075; CFI = 0.98; TLI = 0.98, RMSEA = 0.05; SRMR = 0.03$). Wie bei den Ergebnissen der exploratorischen Faktorenanalyse weisen die Korrelationen der Faktoren untereinander darauf hin, dass alle drei Dimensionen zum Verständnis der Bildentstehung mittels Sammellinse beitragen und wechselseitig inhaltlich zusammenhängen ($0.49 \leq r \leq 0.73$). In Bezug auf die lokalen Gütekriterien zeigte sich, dass jeder der Pfadkoeffizienten signifikant von Null verschieden war.

In Schritt drei der Analyse wurden beide Analysen exploratorische und konfirmatorische Faktorenanalyse mit der jeweils anderen Teilstichprobe durchgeführt. In beiden Fällen konnten die Ergebnisse jeweils mit geringfügigen Abweichungen repliziert werden, daher werden die Ergebnisse hier lediglich im Überblick dargestellt.

Insgesamt werden in Stichprobe A in der exploratorischen Faktorenanalyse 44 % der Gesamtvarianz durch drei Faktoren mit Eigenwerten > 1 aufgeklärt und in Stichprobe B 33 % der Gesamtvarianz. In Stichprobe A ergab sich für den Faktor AB ein Eigenwert von 2.00 (20 % aufgeklärte Varianz), für den Faktor BS ein Eigenwert von 1,34 (13 % aufgeklärte Varianz) und für Faktor LS ein Eigenwert von 1.09 (11 % aufgeklärte Varianz). Die Kommunalitäten lagen mit Werten Bereich von $0.17 \leq \hat{h} \leq 0.75$ in einem für die Stichprobengröße zufriedenstellendem bis sehr guten Bereich. Auch in Teilstichprobe A wurden mittlere bis hohe Korrelationen zwischen den drei Faktoren gefunden, was erneut bestätigt, dass alle drei Dimensionen inhaltlich eng miteinander verknüpft sind. So korrelieren die Skalen AB und LS mit $r = .41$, die Faktoren AB und BS mit $r = .53$ sowie die Faktoren LS und BS mit $r = .62$.

Für die konfirmatorische Faktorenanalyse in Stichprobe B ergab sich die gleiche Struktur wie in der exploratorischen und konfirmatorischen Faktorenanalyse der Stichprobe A, wiederum mit mittleren bis hohe Korrelationen zwischen

den drei Faktoren ($.33 \leq r \leq .53$). Die Annahme des Modells wird hierbei durch gute globale Fit-Indizes gestützt ($\chi^2_{(32)} 30.12, p = 0.562; CFI \approx 1.00; TLI \approx 1.00, RMSEA = 0.05; SRMR = 0.03$).

Tab. 4. Eigenwerte der Faktoren und Anteil aufgeklärter Varianz der resultierenden Skalen für die Faktorenanalyse auf Basis der Daten von Stichprobe B, interne Konsistenz je Skala für A ($n = 480$) / B ($n = 389$) / Gesamtstichprobe A+B ($N = 869$)

Faktor	Interpretation	Items y1-8c	α_c A/B/A+B	Eigenwert	Anteil erklärter Varianz in %	Anteil kumulierter Varianz in %
1	Verständnisfragen zur Bildentstehung inklusive Abdeckaufgaben (AB)	y8a, y8b, y8c	.85/.59/.83	2,41	12	12
2	Geradlinige Lichtausbreitung und Streuung (LS)	y1, y2, y4, y5	.63/.57/.61	1,45	11	23
3	Verständnis Bildkonstruktion / Strahlenmodell (BS)	y3, y6, y7	.54/.43/.45	1,12	10	33

Tab. 5. Antwortverteilung und Ergebnisse der exploratorischen Faktorenanalyse des KTSO-A (Stichprobe B, $n = 389$)

Variable		Anteil Antworten in %, Mehrfachauswahl möglich, Gesamtsumme kann > 100 % sein	Promax Rotation Mustermatrix			
y _i	Itemformulierung		F1 ^a : AB	F2 ^a : BS	F3 ^a : LS	Ĥ (y _i) ^b
y1	Welche der folgenden Gegenstände / Tiere kann man in einem völlig abgedunkelten Raum sehen?		-0.09	-0.01	0.60	0.33
	ein leuchtendes Glühwürmchen	92 %				
	ein weißes Blatt Papier	7 %				
	einen Fahrrad-Reflektor	6 %				
	die Augen einer Katze	29 %				
y2	Hat es einen Einfluss auf die Helligkeit in einem Zimmer, ob es helle oder dunkle Tapeten hat?		0.09	0.13	0.37	0.15
	ja, weil die helle Tapete mehr Licht streut, das ins Auge fällt, als eine dunkle Tapete.	77 %				
	nein, weil dunkle Tapeten nichts an der Menge des Lichtes im Raum ändern.	6 %				
	ja, weil auf der hellen Tapete mehr Licht liegen bleibt.	18 %				
	nein, es kommt auf die Lampe in dem Zimmer an oder das Sonnenlicht, das durch das Fenster fällt und nicht auf die Helligkeit der Tapete.	13 %				
y3	Was ist richtig? Lichtstrahlen sind ...		0.07	0.22	0.20	0.14
	...etwas Wirkliches, so wie dünne Wasserstrahlen aus einer Spritzpistole.	19 %				
	...etwas Gedachtes, so wie Konstruktionen in der Geometrie, um z. B. Dreiecks-Probleme lösen zu können.	53 %				
	... exakt das Gleiche wie Lichtbündel.	17 %				
	... Lichtbündel sind etwas Gedachtes, z. B. um die Bildgröße bestimmen zu können.	38 %				
y4	In einem abgedunkelten Raum ist der helle Fleck einer Taschenlampe an der Wand zu sehen, nicht aber der Lichtstrahl zwischen Taschenlampe und Wand. Warum?		-0.05	-0.09	0.49	0.21
	Erst das an Gegenständen gestreute Licht trifft ins Auge und ist sichtbar.	36 %				
	In dem dunklen Raum wird das Licht absorbiert (verschluckt), daher ist es nicht zu sehen.	29 %				
	Das Licht erhellt die Wand, weil es auf ihr liegen bleibt.	52 %				

Variable		Anteil Antworten in %,	Promax Rotation Mustermatrix			
	<i>Das Licht der Taschenlampe entfernt sich vom Beobachter, erst durch die Wand wird es umgedreht und geht auf den Beobachter zu.</i>	22 %				
y5	Was passiert, wenn man in dem Lichtstrahl einen Tafellappen aufschüttelt?		-0.06	0.09	0.60	0.44
	<i>Die Staubteilchen wirken wie kleine Linsen, die das Licht auf der Wand bündeln.</i>	7 %				
	<i>Der feine Kreidestaub sammelt das Licht und dadurch sieht man den hellen Fleck auf der Wand nicht mehr.</i>	7 %				
	<i>Die Staubteilchen werden durch das auftreffende Licht durcheinandergewirbelt.</i>	19 %				
	<i>Die Staubteilchen streuen das Licht in alle Richtungen, dadurch trifft es ins Auge und wird sichtbar.</i>	72 %				
y6	Wie entsteht durch Verwendung einer Sammellinse ein Bild, das auf einem Schirm aufgefangen werden kann?		0.11	0.19	0.13	0.11
	<i>Solch ein Bild entsteht durch Spiegelung der Lichtstrahlen an der Linse nach dem Reflexionsgesetz.</i>	19 %				
	<i>Eine Sammellinse hat den Effekt, die Lichtstrahlen aufzuhehlen.</i>	6 %				
	<i>Lichtstrahlen, die von einem Gegenstandspunkt ausgehen, werden durch die Sammellinse abgelenkt und treffen sich im Bildpunkt.</i>	49 %				
	<i>Das Bild geht als Ganzes durch die Linse zum Schirm, dabei wird es in der Linse unter Einhaltung der Linsengesetze umgedreht.</i>	30 %				
y7	Welche Aussagen zur Bildkonstruktion und Bildentstehung treffen zu?		-0.08	0.96	0.02	0.86
	<i>Nur die ausgezeichneten Strahlen kann man im Strahlengang zeichnen.</i>	15 %				
	<i>Mit den ausgezeichneten Strahlen kann man den Strahlengang besonders leicht zeichnen.</i>	47 %				
	<i>Die ausgezeichneten Strahlen erschweren die Zeichnung, machen sie dafür aber besonders genau.</i>	19 %				
	<i>Ohne die ausgezeichneten Strahlen (wenn diese z. B. durch dünne Stifte aufgehalten werden) kann es kein Bild geben.</i>	20 %				
y8	In einer Versuchsanordnung sind eine Glühlampe, eine Sammellinse und ein Schirm auf einer optischen Bank so montiert, dass ein vergrößertes, umgekehrtes, scharfes Bild des Glühfadens entsteht:					
y8a	Was passiert, wenn man die untere Hälfte der Linse abgedeckt.		0.54	0.08	-0.15	0.29
	<i>Die obere Hälfte des Bildes wird abgeschnitten.</i>	16 %				
	<i>Die untere Hälfte des Bildes wird abgeschnitten.</i>	20 %				
	<i>Das Bild wird dunkler.</i>	61 %				
	<i>Das Bild wird kleiner.</i>	11 %				
y8b	Was passiert, wenn man einen Karton mit großem Loch (ringförmige Blende) vor die Linse hält?		0.75	-0.10	0.01	0.51
	<i>Das Bild wird kleiner.</i>	23 %				
	<i>Das Bild wird dunkler.</i>	47 %				
	<i>Die Ränder des Bildes werden kreisförmig abgeschnitten.</i>	37 %				
	<i>Das Bild wird heller.</i>	10 %				
y8c	Was passiert, wenn man einen Karton mit einem sehr kleinen Loch 5 mm (ringförmige Blende) vor die Linse hält?		0.49	-0.03	0.08	0.26
	<i>Das Bild wird kleiner.</i>	39 %				
	<i>Das Bild wird dunkler.</i>	48 %				
	<i>Die Ränder des Bildes werden kreisförmig abgeschnitten.</i>	27 %				
	<i>Das Bild wird heller.</i>	10 %				

4.3 Validität

Als ein weiterer Baustein der Validität des KTSO-A (neben der Darstellung von Rechercheergebnissen in den Kapiteln 2.2 und 2.3) wurde ein Expertenrating durchgeführt. An dem Rating nahmen 11 Personen (5 Frauen und 6 Männer) im Alter von 27 bis 71 Jahren ($M = 40.2$ Jahre, $SD = 14.8$) teil. Alle Befragten verfügten über einen Masterabschluss, Diplom oder Staatsexamen in Physik und

hatten sich entweder während des Studiums oder im Anschluss an das Studium auf Fachdidaktik Physik spezialisiert. 8 Befragte haben das Referendariat und das zweite Staatsexamen absolviert und verfügen zwischen 1 bis 30 Jahren Unterrichtserfahrung nach dem Referendariat ($M = 13$ Jahre, $SD = 10.79$), 7 Personen aus diesem Kreis haben auch in den Klassenstufen 7 und 8 unterrichtet, eine Person in Klassenstufe 9 und in der Oberstufe. Bei den drei Befragten ohne 2. Staatsexamen handelt es sich um

Doktorand/-innen in Fachdidaktik Physik mit Unterrichtserfahrung an der Hochschule im Umfang von 1 bis 3 Jahren ($M = 2.05$ Jahre, $SD = 1.23$). Da sich die Einschätzung der 8 erfahrenen Experten von den weniger erfahrenen Personen deskriptiv kaum und jedenfalls nicht signifikant unterscheiden, werden im Folgenden die Werte für die Gesamtstichprobe aller 11 Befragten berichtet.

Die Befragten stuften jede Testaufgabe des KTSO-A unabhängig voneinander schriftlich (postalische Befragung) und anonym auf Basis der folgenden drei Items anhand einer 5-stufigen endpunktbenannten Likert-Skala von „trifft gar nicht zu (1)“ bis „trifft zu“ (5) ein.

- Item 1: Diese Aufgabe erfasst ein relevantes Konzept der Strahlenoptik.
- Item 2: Ich kann mir vorstellen, diese Aufgabe im Unterricht zur Strahlenoptik in Klassenstufe 7 oder 8 zu thematisieren.
- Item 3: Probleme bei der Lösung dieser Aufgabe zeigen grundlegende Verständnisschwierigkeiten in der Strahlenoptik auf.

In Tabelle 6 wird die unjustierte Intraklassenkorrelation (ICC) für die drei Expertenfragen zu allen 10 Items als Maß der Beurteilerreliabilität sowie das Maß r_{WG} zur Beurteilung des Ausmaßes an Beurteilerübereinstimmung auf Itemebene des KTSO-A berichtet (James, Demaree & Wolf, 1984, S. 87; Lindell, Brandt & Whitney, 1999, S. 127; die Formel, welche der Berechnung zugrunde liegt, ist in Appendix 1 angegeben).

Die unjustierte ICC wurde gewählt, da die Experten die 10 Testaufgaben absolut anhand einer Likert Skala bewerten sollten. Hierbei wurde davon ausgegangen, dass die Beurteiler eine zufällige Auswahl aus der Grundgesamtheit möglicher Beurteiler darstellen. (Die Formel zur Berechnung der verwendeten Intraklassenkorrelation ist ebenfalls in Appendix 1 angegeben). Die ICC ist ein Maß für die Interraterreliabilität einer Ratingskala und kein Maß zur Beurteilung der Übereinstimmung eines einzelnen zu beurteilenden Objekts oder Items.

Der Wert fällt hoch aus, wenn die Beurteiler ein gegebenes Item ähnlich, und verschiedene Items entsprechend unterschiedlich bewerten. Das heißt die Höhe der ICC hängt auch von der Größe der Mittelwertunterschiede zwischen den zu ratenden Items ab.

Insofern ist die Höhe der ICC in diesem Fall unter dem Vorbehalt zu interpretieren, dass idealerweise alle Aufgaben als konsistent geeignet eingestuft werden sollen, was

zu geringen Mittelwertunterschieden im Rating der Testaufgaben und damit zu einer geringen Varianz in Bezug auf die Beurteilung führt (Wirtz und Caspar, 2002, S. 161). Vor diesem Hintergrund kann der empirisch vorgefundene Wertebereich von $.50 \leq ICC \leq .62$, wobei alle Werte auf dem 5 % Niveau signifikant sind, als erwartungsgemäß betrachtet werden.

Auf Ebene der Einzelitems des KTSO-A ergeben sich durchweg konsistent Mittelwerte > 4 (Skala 1-5) und bis auf eine Ausnahme (Item 6) verhältnismäßig geringe Standardabweichungen. So wurde Item 6 „Wie entsteht durch Verwendung einer Sammellinse ein Bild, das auf einem Schirm aufgefangen werden kann?“ von einem Teil der Beurteiler kritisch bewertet und Item 8b (ringförmige Abdeckung der Linse) von einem Beurteiler als für diese Altersstufe als schwierig eingestuft. Die konsistenten und bis auf die genannten beiden einzelnen Ausnahmen guten Bewertungen spiegeln sich auch in den Werten der r_{WG} wieder. Im Gegensatz zum ICC, der die Beurteilerreliabilität abschätzt, ist die r_{WG} ein Maß zur Einschätzung der Beurteilerübereinstimmung, welches auf Itemebene angewandt werden kann und gibt an zu welchem Grad unterschiedliche Beurteiler ein Objekt, hier eine Testaufgabe, gleich bewerten (Lindell, Brandt & Whitney, 1999, S. 127). Der Index für Einzelitems vergleicht die beobachtete Varianz der Beurteilerantworten mit der Varianz, welche erwartet wird, wenn sich alle Ratings durch eine gleichverteilte Residualvarianz auszeichnen. Werte $\geq .80$ gelten bei einer fünfstufigen Likert-Skala als ausreichend hoch, um eine konsistente Beurteilerübereinstimmung annehmen zu können (vgl. ebd.). Dieses Kriterium ist für alle Items und jede der drei Expertenfragen mit Ausnahme von Item 6 erfüllt (vgl. Tab. 6).

Als Anhaltspunkt zur Abschätzung der Kriteriumsvalidität wurden des Weiteren Korrelationen zwischen der Gesamtpunktzahl und Zeugnisnoten in Mathematik, Physik und Deutsch berechnet. Im Ergebnis zeigen sich signifikante mittlere Korrelationen zwischen den Fachnoten in Physik ($r_{(869)} = -.34, p < 0.001$) und Mathematik ($r_{(869)} = -.29, p < 0.001$.) und den Gesamtpunktzahlen des KTSO-A. Die negativen Vorzeichen resultieren aus der entgegengesetzten Skalierung von Schulnoten im deutschen Schulsystem und den Testscores des KTSO-A.

Tab. 8. Expertenrating zum KTSO-A, $N = 11$

	Skala	Item	^a Expertenbewertung Frage 1			^a Expertenbewertung Frage 2			^a Expertenbewertung Frage 3		
			^b ICC = .522 * (10 Items)			^b ICC = .619 * (10 Items)			^b ICC = .500 * (10 Items)		
			<i>M</i> (<i>SD</i>)	<i>Median</i> (<i>Range</i>)	<i>r</i> _{WG}	<i>M</i> (<i>SD</i>)	<i>Median</i> (<i>Range</i>)	<i>r</i> _{WG}	<i>M</i> (<i>SD</i>)	<i>Median</i> (<i>Range</i>)	<i>r</i> _{WG}
y_1	LS	Sichtbare Gegenstände / Lebewesen in dunklem Raum	4.82 (0.40)	5 (4 -5)	0.96	4.73 (0.47)	5 (4 -5)	0.95	4.45 (0.69)	5 (4 -5)	0.89
y_2	LS	Helligkeit eines Zimmers - Farbe der Tapete	4.64 (0.50)	5 (4 -5)	0.94	4.55 (0.69)	5 (3 -5)	0.89	4.55 (0.69)	5 (3 -5)	0.89
y_3	BS	Lichtstrahlen sind ...	4.45 (0.82)	5 (4 -5)	0.85	4.36 (0.67)	5 (3 -5)	0.90	4.45 (0.82)	5 (4 -5)	0.89
y_4	LS	Lichtstrahl in dunklem Raum	4.82 (0.40)	5 (4 -5)	0.96	4.82 (0.40)	5 (4 -5)	0.96	4.82 (0.40)	5 (4 -5)	0.95
y_5	LS	Lichtstrahl: Kreidestaub	4.82 (0.40)	5 (4 -5)	0.96	4.73 (0.47)	5 (4 -5)	0.95	4.82 (0.40)	5 (4 -5)	0.96
y_6	BS	Entstehung reeller Bilder an Sammellinse	4.10 (1.45)	5 (1 -5)	0.50	4.22 (1.30)	5 (2 -5)	0.62	4.10 (1.20)	5 (2 -5)	0.64
y_7	BS	Bildkonstruktion: Bildentstehung	4.60 (0.52)	5 (4 -5)	0.94	4.90 (0.32)	5 (4 -5)	0.98	4.90 (0.32)	5 (4 -5)	0.98
y_{8a}	AB	Abdeckung einer Linsenhälfte	4.73 (0.47)	5 (4 -5)	0.95	4.82 (0.40)	5 (4 -5)	0.96	4.91 (0.30)	4 (4 -5)	0.98
y_{8b}	AB	Ringförmige Abdeckung der Linse	4.67 (0.71)	5 (4 -5)	0.89	4.20 (1.32)	5 (1 -5)	0.88	4.56 (0.73)	5 (3 -5)	0.88
y_{8c}	AB	Enge ringförmige Abdeckung der Linse	4.80 (0.42)	5 (4 -5)	0.96	4.50 (0.71)	5 (3 -5)	0.89	4.60 (0.52)	5 (4 -5)	0.94

a 5-stufige Likert Skala von 1 (trifft gar nicht zu) bis 5 (trifft zu)
b Unjustierter Wert, entspricht der Auswahl „absolute agreement“ in SPSS. * $p < 0.05$
c r_{WG} (Definition siehe Appendix 1)

5 DISKUSSION UND AUSBLICK

5.1 Diskussion der Testeigenschaften

Wir diskutieren zunächst die Ergebnisse zu Items, Distraktoren und dem Gesamttest in einem allgemeinen Überblick und dann Grenzen, Mängel und Stärken des Tests im Hinblick auf den Stand der Entwicklung von Konzepttest in anderen Bereichen.

Die Kennwerte des Tests (Itemschwierigkeiten, Trennschärfen, Reliabilität) liegen für Stichprobe A durchwegs und für B großenteils in den in der Literatur angegebenen Akzeptanzbereichen. Abweichungen davon, Unterschiede zwischen den Stichproben (insbesondere die schlechteren Trennschärfen von Stichprobe B im Vergleich zu A) sowie weiterer Entwicklungs- und Verbesserungsbedarf werden unten diskutiert. Die Werte für die Reliabilität (innere Konsistenz) insbesondere sind für beide Stichproben vergleichbar zu denen einer Reihe anderer Konzepttests im naturwissenschaftlich-mathematischen Bereich (Evolution: 0.58, 0.64 - zwei Stichproben), Anderson et al, 2002; Allgemeine Chemie: 0.71, Mulford & Robinson, 2002; elektrische Schaltkreise: 0.7, Engelhardt & Beichner, 2004 ; Elektrizität & Magnetismus: 0.75, Maloney et al., 2001; Statistik: ≈ 0.7

(verschiedene Stichproben), Allen, 2006; Materie-Energie-Wechselwirkung: 0.75, Ding & Beichner, 2009). Der Wert für die Teilstichprobe A ist nahe dem Schwellenwert für Individualdiagnostik (Bühner, 2011; Ding & Beichner, 2009; Doran, 1980).

Die Analyse der Distraktoren zeigt auf, dass alle Distraktoren so plausibel waren, dass sie auch nach dem Unterricht ausreichend häufig gewählt wurden. Niedrige, aber noch ausreichende Werte zwischen 5 % und 10 % ergeben sich bei Items der Skala Lichtausbreitung und Streuung, die vergleichsweise leicht zu lösen waren ($0.67 \leq P_i \leq 0.77$). Dabei zeigt sich, dass auch jeweils zwei der „selteneren“ Distraktoren in gleicher Weise mit einer Häufigkeit über dem Schwellwert gewählt wurden, was als Hinweis gewertet werden kann, dass sie übereinstimmend mit den Ergebnissen von Wiesner (1986; 1992a) tatsächlich relevante Schülervorstellungen erfassen und daher beibehalten werden sollten. Abweichend davon ergibt sich nur bei Item 6, „Wie entsteht durch die Verwendung einer Sammellinse ein Bild, das auf einem Schirm aufgefangen werden kann?“ ein deutlicher Unterschied in der Wahlhäufigkeit für die Option „die Linse hat den Effekt, die Lichtstrahlen aufzuhellen“ im Vergleich zu den anderen Distraktoren (6 % versus, 19 % bis 49 %); das Item wurde auch als einziges in dem Expertenrating nicht zufriedenstellend bewertet ($r_{WG} < 0.70$). Dennoch wurde das Item in der Analyse aus inhaltlichen

Gründen einbezogen, da auch dem selten gewählten Distraktor eine klassische Schülervorstellung zugrunde liegt (vgl. dazu Guesne, 1985; Wiesner, 1994), und er selbst nach Optikunterricht noch von immerhin 6 % der Schülerinnen und Schüler gewählt wird.

Es finden sich in den Kennwerten Unterschiede zwischen den Teilstichproben A und B, die sich sinnvoll interpretieren lassen. Die Lösungswahrscheinlichkeit für A ist höher als die für B, was zeigt, dass der Test grundsätzlich in der Lage ist, den Effekt eines auf die fraglichen Lernschwierigkeiten bzw. Fehlkonzepte zielenden Unterrichts nachzuweisen. Auch die korrigierten Trennschärfen liegen in A generell höher als in B, das gleiche gilt für die interne Konsistenz, ein Sachverhalt, der ähnlich auch bei anderen Konzept-Tests gefunden wurde (Zeilik et al., 1997; Ramlo, 2008). Das lässt sich so deuten, dass mit zunehmendem Wissen und Verständnis in einem Bereich diese auch konsistenter werden; Wissen und Verständnis haben Systemcharakter (Mandl & Spada, 1988; Braisby & Gellatly, 2008), insbesondere in den Naturwissenschaften (Nersessian, 1991; Kircher et al., 2009, Kap., 19.4.4).

Das Ergebnis der Faktorenanalyse (Abschnitt 4.2) kann inhaltlich wie folgt interpretiert werden. Es ergeben sich drei Dimensionen des konzeptuellen Verständnisses im Bereich der Strahlenoptik:

- Abdeckaufgaben mit Verständnisfragen zur Bildentstehung (AB);
- Verständnisfragen zur geradlinigen Lichtausbreitung und Streuung (LS) und
- Verständnisfragen zur Bildkonstruktion und zum Strahlenmodell (BS).

Als Anhaltspunkt für die Bewertung der Konstruktvalidität zeigt die Kreuzvalidierung, dass der KTSO-A mit den drei übergeordneten Verständnisbereichen „Abdeckaufgaben mit Verständnisfragen zur Bildentstehung“, „Verständnisfragen zur geradlinigen Lichtausbreitung und Streuung“ sowie „Verständnisfragen zur Bildkonstruktion und zum Strahlenmodell“ das konzeptuelle Verständnis der Strahlenoptik im Kontext der Bildentstehung an der Sammellinse in inhaltlich sinnvoll interpretierbarer Weise erfasst. Das Modell belegt, dass die drei Verständnisbereiche unterschiedliche Facetten des konzeptuellen Verständnisses erfassen, welche jedoch signifikant untereinander korrelieren. Die Korrelationen weisen einen mittleren bis starken Zusammenhang der drei Facetten untereinander auf. Im Zug der Kreuzvalidierung wurde sowohl in der exploratorischen wie auch der konfirmatorischen Faktorenanalyse die gleiche Struktur aufgefunden und zwar, wie zuvor angesprochen, unabhängig davon, ob die Schülerinnen und Schüler gezielt Unterricht erhalten hatte, welcher Schülervorstellungen behandelt (Stichprobe A) oder nicht (Stichprobe B). Die Einschätzung der Kreuzvalidierung und Strukturaufklärung lassen sich so zusammenfassen: 1) die Anwendungskriterien einer exploratorischen Faktorenanalyse zur Strukturaufklärung sind gut erfüllt; 2) sie führt zu einer dreidimensionalen Substruktur des Tests, die gute Fit-Gütekriterien aufweist; 3) die gefundene Struktur wird durch eine konfirmatorische Faktorenanalyse mit ebenfalls guter Fit-Güte bestätigt; 4) die so psycho-

metrisch gefundene Struktur hat eine plausible inhaltliche Deutung; 5) ein Mangel ist, dass 2 von 10 Faktorladungen unter dem üblichen Cut-off-Wert liegen.

5.2 Einordnung der Ergebnisse und Grenzen

Wir diskutieren die gefundenen Ergebnisse nun im Hinblick auf den Stand der Entwicklung von Konzept-Tests in anderen Bereichen. Der Sachverhalt teilweise geringer Item-Test-Korrelationen bei der Entwicklung von Konzepttests ist wohlbekannt (Evolution (CINS): Anderson et al, 2002; Statistik (SCI): Allen, 2006; statische Kräfte (CATS), Dynamik (DCI): Jorion et al., 2015); der in der Literatur diskutierte Grund ist der gleiche, der oben schon bei der Diskussion der Unterschiede von Trennschärfen und inneren Konsistenzen zwischen den Stichproben A und B genannt wurde: Ohne gezielte Intervention für die Konzepte in einem gegebenen Inhaltsbereich (und erst recht überhaupt vor Unterricht) ist die begriffliche Struktur von Lernenden wenig konsistent (zur Konsistenzthematik s. o. und Savinainen & Viiri (2008); Nieminen et al, 2010 und Lasry et al., 2011).

Auch Schwierigkeiten bei der Strukturaufklärung von Konzepttests durch Faktorenanalyse sind in der Literatur wohlbekannt. Bei den wenigsten Tests ist ein Versuch dieser Art publiziert (eine Übersichtsarbeit von, 2012 zählt unter 15 Tests nur 4 Versuche auf; Liu, 2012). Für den CSEM (Elektrizität und Magnetismus, Maloney et al., 2001) ist dieser Versuch gescheitert und hat trotz klarer a priori definierbarer Inhaltsbereiche auf keine interpretierbare Struktur geführt. Für den FCI war und ist eine interpretierbare Faktorstruktur Gegenstand intensiver Auseinandersetzung (Überblick: Lasry et al., 2011), und es hat rund, 20 Jahre gedauert, bis eine solche vorgeschlagen werden konnte (Scott et al., 2012).

Schließlich werden weder für den FCI (Scott et al., 2012) noch für die wenigen anderen Konzepttests, für die heute Faktorstrukturen bekannt sind, Reliabilitäten für die gefundenen Subskalen angegeben (Evolution/CINS: Anderson et al, 2002; elektrische Schaltkreise/DIRECT: Engelhardt & Beicher, 2004; Bewegung, Kraft/FMCE, Ramlo, 2008; Materie-Energie-Wechselwirkung/MIET: Ding & Beichner, 2009). Bei den wenigen Ausnahmen (tatsächlich den einzigen, die den Autoren bekannt sind) bei den auch die Subskalen-Reliabilitäten genannt sind, liegen diese nochmals deutlich unter der Gesamtreliabilität (Allen, 2006, SCI/Statistik: 0.3 – 0.5; Jorion et al., 2015, DCI/Dynamik: 0.2 – 0.6; CATS/statische Kräfte (das beste Beispiel): 0.5. – 0.7). Für den TUG-K, ein anderes gut etabliertes Instrument (kein Konzepttest im eigentlichen Sinne), ergeben sich für die inhaltlich a priori definierten Subskalen 0.18 - 0.57, für die beiden durch PCA gefundenen Hauptfaktoren 0.56 und 0.78 (Bektasli, 2006).

Jorion et al. (2015) geben einen Überblick über die Problematik der Validierung von Konzepttests und illustrieren die oben skizzierten Begrenzungen und Schwierigkeiten bezüglich Validität, Reliabilität und Strukturaufklärung mit einer psychometrischen Analyse von Tests aus drei Bereichen (s.o.). Als Ergebnis ihrer Analyse schlagen sie ein an Zielsetzung und Entwicklungsstand des Forschungsgebietes angepasstes Raster der Bewertung von Konzepttests vor, welches die Kriterien der

klassischen Itemanalyse, der Item-Response-Theorie und der exploratorischen und konfirmatorischen Faktorenanalyse zusammenfasst (Tab. 7). Für den KTSO-A befinden sich alle Werte in dem Bereich zwischen „average“ und „excellent“.

Vor diesem Hintergrund fassen wir die Bewertung der Testgüte des KTSO-A wie folgt zusammen: Es wurden (anhand zweier Stichproben mit Stichprobengrößen ≥ 400) psychometrische Eigenschaften ermittelt, die – vor allem im Vergleich mit dem Entwicklungsstand von Konzepttests im Allgemeinen – als akzeptabel bis gut eingeschätzt werden können; Unterschiede zwischen

verschiedenen Items und zwischen den Stichproben lassen sich sinnvoll und unter Bezug auf die vorliegende Forschung interpretieren.

Im Rahmen der Empfehlungen für Testkonstruktion im Allgemeinen (Haladyna & Downing, 1989) und Konzepttest im Besonderen (Lindell et al., 2007) sehen wir es daher als einen Beitrag der Arbeit an, den Schritt von einer forschungsbasierten Übersicht von inhaltlich relevanten konzeptionellen Schwierigkeiten (2.2) zu einem Instrument mit akzeptabler bis guter psychometrischer Charakterisierung (4.1) vollzogen zu haben.

Tab. 7. Raster zur Bewertung von Konzepttests (“categorical judgement scheme for evaluating a concept instrument” nach Jorion et al, 2015; wir behalten die Bezeichnungen des Originals bei), sowie Einordnung des KTSO-A (Rahmen)

Analysis Method	criterion	excellent	good	average	poor	not acceptable
<i>Classical test theory, item statistics</i>	difficulty	.2 to .8 ^A	.2 to .8	.1 to .9 ^B	1 to .9	0 to .10
	discrimination	>.2 ^A	>.1 ^B	>.0	>-.2	> -1.0
	α_c with-item-deleted	all	(3)	(6)	(9)	> (9)
	\leq overall α_c	>.9	>.8	>.65 ^{A,B}	>.5	>.0
<i>Item response theory: not applicable, see text</i>						
<i>Structural analysis*</i>	EFA: items conform to interpretable constructs	all	(5)	(10)	(15)	> (15)
	CFA					
	item loadings	>.3	>.3 (3)	>.1	>.1 (3)	> -1.0
	CFI	>.9	>.8	>.7	>.6	>.0
	RMSEA	<.03	<.05	<.10	<.20	>.20

Hinweise:

– Werte in Klammern stehen für Anzahl von Items die das jeweiligen Kriterium verletzen dürfen

– Die Einstufung des KTSO-A in diesem Raster ist in Fettdruck kenntlich gemacht, die Indizes A/B beziehen sich auf die beiden Validierungsstichproben; alle Werte befinden sich in dem Bereich zw. „average“ und „excellent“, s. Rahmen

* Ergebnisse mit EFA zu Stichprobe B, CFA zu Stichprobe A; zusätzlich wurde eine Kreuzvalidierung mit positivem Ergebnis durchgeführt, s. Text.

Darüber hinaus wurde eine mehrdimensionale Struktur gefunden, mit ihrerseits akzeptablen bis z. T. sehr guten Kenngrößen, und einer inhaltlichen Deutung; mit Blick auf die bekannten Schwierigkeiten der Strukturaufklärung bei Konzepttests kann man mit der gebotenen Vorsicht zumindest von starken Hinweisen auf interpretierbare Teilkonstrukte sprechen. In diesem Zusammenhang sei darauf hingewiesen, dass hinsichtlich der üblichen Akzeptanzwerte psychometrischer Indizes durchwegs betont wird, dass diese nicht starre Kriterien darstellen, sondern mit Augenmaß und im Zusammenhang mit dem Stand und Zielsetzungen des Anwendungsbereichs zu sehen sind (Abell et al., 2009; EFPA, 2013; Jorion et al., 2015). Moosbrugger und Kelava (2012) stellen fest, dass „wenn es keine besser geeigneten Testverfahren gibt, der Einsatz eines niedrig reliablen Messinstruments immer noch aufschlussreicher sein kann als der gänzliche Verzicht auf den Einsatz von Tests“. Für den FCI als am intensivsten untersuchten Konzepttest waren Validität, Reliabilität und Strukturaufklärung Gegenstand einer jahrzehntelangen, intensiven wissenschaftlichen Auseinandersetzung (Härtig, 2014; Scott, 2012; Stewart, Griffin & Stewart, 2007; Morris, Branum-martin & Harshman et al., 2006; Rebello & Zollman, 2004; Lasry et al., 2011, auch für einen Überblick über ältere Arbeiten). Diese wichtige Diskussion hätte nicht geführt werden können, ohne dass es den Test in publizierter Form gegeben hätte - mit einer mit Augenmaß geführten Bewertung der Gütekriterien. Für den Bereich

des KTSO-A liegen ebenfalls seit Jahrzehnten isolierte Items vor, aber kein auch nur in Ansätzen validiertes Instrument. Diese Lücke wird in der vorliegenden Arbeit geschlossen, mit einem nach der oben geführten Diskussion durchaus brauchbaren Instrument, das in sinnvoller Weise Ausgangspunkt einer weiteren Entwicklung in der Community sein kann (wie es etwa beim FCI der Fall war).

Künftige Weiterentwicklungen des KTSO-A in diesem Sinne sehen wir entlang folgender Linien. Auch wenn die Kennwerte (bis auf einzelne Ausnahmen) in den laut Literatur akzeptierten Wertebereichen liegen, müssen wir folgende Einschränkungen einräumen: Die Werte einiger Trennschärfen sind im unteren Bereich des Akzeptanzbereiches; auch die Reliabilität des Gesamttests für die Stichprobe ohne gezielte Intervention für konzeptuelles Lernen ist nur akzeptabel. Hier können Versuche der Überarbeitung der Itemformulierungen und Hinzunahme weitere Items eine Verbesserung bringen; andererseits kann es auch sein, dass wie bei anderen Konzepttest (s. o. g. Diskussion und Literatur) dass die geringe Konsistenz des konzeptionellen Wissens der Lernenden selbst die Ursache ist. Auf der Ebene von Einzelitems können des Weiteren einige Distraktoren optimiert werden: So könnte eine künftige Testfassung die beiden am häufigsten gewählten Distraktoren und die korrekte Antwortoption beibehalten werden (vgl. Rodriguez, 2005, zur optimalen Anzahl an Distraktoren in Multiple-Choice-Tests).

Auch wenn Kreuzvalidierung und inhaltliche Diskussion gute Hinweise auf die Existenz sinnvoll interpretierbarer Subskalen geben, sollten die Reliabilitätswerte der LS und insbesondere BS-Skalen verbessert werden (beispielsweise kann letztere um zwei bis drei weitere Aufgaben erweitert werden, welche speziell das Konzept der Punkt-zu-Punkt-Abbildung thematisieren). Weitere Forschung zu diesem und zu anderen Konzepttests muss aber auch auf der allgemeinen Ebene zeigen, wie die bekannten Probleme der psychometrischen Identifikation und niedriger Reliabilitäten (Maloney et al., 2001; Allen, 2006; Jorion et al., 2015) auch inhaltlich sehr gut definierter Teilskalen zu erklären und ggf. zu lösen sind.

Im Hinblick auf die abgedeckten Teilbereiche bietet es sich im Bereich der Strahlenoptik an, die Entwicklung auf weitere Themen wie Licht- und Schatten, gerichteten Reflexion und Spiegelbildern und virtuellen Bildern bei Spiegeln und Linsen zu erweitern. Dies könnte dazu beitragen, die Lücke zwischen grundsätzlich guter Kenntnis einschlägiger Fehlkonzepte und Lernschwierigkeiten und dem Vorliegen guter Einzelitems hierzu (Wiesner, 1986, 1992a, 1992b; Fetherstonhaugh & Treagust, 1992; Chu et al., 2009) und psychometrisch validierten Konzepttests zu schließen. Eine Weiterentwicklung des vorliegenden Tests in diesem Sinn ist in Arbeit.

Schließlich ist eine Begrenzung der vorliegenden Arbeit, dass die Validierung bei einer mehrheitlich aus dem Gymnasium stammenden Stichprobe durchgeführt wurde. Bei anderen Schülergruppen (z. B. in der Realschule oder integrierten Gesamtschule), kann es zu Abweichungen in den Ergebnissen zu den Itemkennwerten, der internen Konsistenz oder der Dimensionalität kommen. Ob Schülervorstellungen gezielt im Unterricht behandelt wurden (Teilstichprobe A) oder nicht (Teilstichprobe B) wirkt sich aber, wie die Kreuzvalidierung belegt, nachweislich nicht auf die Struktur des Tests aus.

5.3 Ausblick und Fazit

Für die o. g. Nachteile und Grenzen des Tests können nur weitere Untersuchungen zu dessen Einsatz Aufschluss geben, bei verschiedenen Lernergruppen, und durch verschiedene Forschergruppen, wie es auch die Entwicklung des FCI über mehr als zwei Jahrzehnte vorgebracht hat. Dazu erscheint es sinnvoll, den Test in der heutigen Fassung in publizierter Form zur Verfügung

zu haben. Bereits auf dem gegenwärtigen Stand kann als Resümee festgehalten werden, dass der KTSO-A mit den vorgestellten 10 Items eine laut Literaturstandards psychometrisch akzeptable bis gute und zugleich praktikable Testfassung für Konzepte zur Bildentstehung an der Sammellinse und der geradlinigen Lichtausbreitung und Streuung darstellt, die für Forschungszwecke und zur unterrichtlichen Diagnostik eingesetzt werden kann.

APPENDIX

Appendix 1:

ICC_{unjustiert/random} (Wirtz & Caspar, 2002, S. 184)

$$ICC_{unjust} = \frac{MS_{zw} - MS_{res}}{MS_{zw} + (k - 1) MS_{res} + \frac{k}{n} (MS_{rat} - MS_{res})}$$

(MS = „mean squares“)

MS_{zw} : Varianz zwischen den zu ratenden Objekten

MS_{res} : Residualvarianz

MS_{rat} : Varianz zwischen den Beurteilern

k : Anzahl der Beurteiler (hier 11)

n : Anzahl der zu beurteilenden Objekte oder Personen (hier 10 Testitems)

r_{wg} Index of Interrater Agreement for Multi-Item Ratings of a single target (James, Demaree, Wolf, 1984, S. 87; Lindell, Brandt & Whitney, 1999, S. 127)

$$r_{wg} = 1 - \frac{S_x^2}{S_{EU}^2}$$

r_{wg} : Beurteilerübereinstimmung innerhalb einer Gruppe von k Beurteilern in Bezug auf ein einzelnes zu beurteilendes Objekt X

s_x^2 : beobachtete Varianz in Bezug auf das beurteilte Objekt X

s_{EU}^2 : Varianz in Bezug auf das beurteilte Objekt X, die zu erwarten wäre, wenn alle Urteile nur auf einem zufälligen Messfehler basieren.

Der Index „EU“ steht für einen erwarteten Fehler E basierend auf einer Gleichverteilung englisch „uniform distribution“ mit A Antwortmöglichkeiten;

bei einer fünf-stufigen Likertskala ergibt

$$\text{sich: } s_{EU}^2 = \frac{A^2 - 1}{12} = 2$$

Appendix 2: Vollständige Testfassung mit Lösungen

Konzepttest Strahlenoptik

Abbildungen

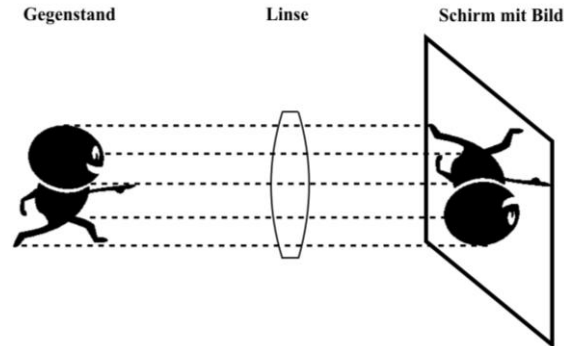
Kreuze jeweils diejenigen Antworten an, die richtig sind.

Es können eine oder mehrere Antworten richtig sein! Beachte die Abbildungen, wenn vorhanden!

Skala	Nr.	Item
LS	1.	<p>Welche der folgenden Gegenstände / Lebewesen kann man in einem völlig abgedunkelten Raum sehen?</p> <p><input type="checkbox"/> ein leuchtendes Glühwürmchen</p> <p><input type="checkbox"/> ein weißes Blatt Papier</p> <p><input type="checkbox"/> einen Fahrrad-Reflektor</p> <p><input type="checkbox"/> die Augen einer Katze</p>
LS	2.	<p>Hat es einen Einfluss auf die Helligkeit in einem Zimmer, ob es helle oder dunkle Tapeten hat?</p> <p><input type="checkbox"/> Ja, weil die helle Tapete mehr Licht streut, das ins Auge fällt, als eine dunkle Tapete.</p> <p><input type="checkbox"/> Nein, weil dunkle Tapeten nichts an der Menge des Lichtes im Raum ändern.</p> <p><input type="checkbox"/> Ja, weil auf der hellen Tapete mehr Licht liegen bleibt.</p> <p><input type="checkbox"/> Nein, es kommt auf die Lampe in dem Zimmer an oder das Sonnenlicht, das durch das Fenster fällt und nicht auf die Helligkeit der Tapete.</p>
BS	3.	<p>Was ist richtig?</p> <p><input type="checkbox"/> Lichtstrahlen sind etwas Wirkliches, so wie dünne Wasserstrahlen aus einer Spritzpistole.</p> <p><input type="checkbox"/> Lichtstrahlen sind etwas Gedachtes, so wie Konstruktionen in der Geometrie, um z. B. Dreiecks-Probleme lösen zu können.</p> <p><input type="checkbox"/> Lichtstrahlen sind exakt das gleiche wie Lichtbündel.</p> <p><input type="checkbox"/> Lichtbündel sind etwas Gedachtes, z.B. um die Bildgröße bestimmen zu können.</p>
LS	4.	<p>In einem abgedunkelten Raum ist der Lichtfleck einer Taschenlampe an der Wand zu sehen, nicht aber der Lichtstrahl von der Taschenlampe zur Wand. Warum?</p> <p><input type="checkbox"/> Erst das an Gegenständen gestreute Licht trifft ins Auge und ist sichtbar.</p> <p><input type="checkbox"/> In dem dunklen Raum wird das Licht absorbiert (verschluckt), daher ist es nicht zu sehen.</p> <p><input type="checkbox"/> Das Licht erhellt die Wand, weil es auf ihr liegen bleibt.</p> <p><input type="checkbox"/> Das Licht der Taschenlampe entfernt sich vom Beobachter, erst durch die Wand wird es umgedreht und geht auf den Beobachter zu.</p>
LS	5.	<p>Was passiert, wenn man in dem Lichtstrahl einen Tafellappen aufschüttelt?</p> <p><input type="checkbox"/> Die Staubteilchen wirken wie kleine Linsen, die das Licht auf der Wand bündeln.</p> <p><input type="checkbox"/> Der feine Kreidestaub sammelt das Licht und dadurch sieht man den hellen Fleck auf der Wand nicht mehr.</p> <p><input type="checkbox"/> Die Staubteilchen werden durch das auftreffende Licht durcheinander gewirbelt.</p> <p><input type="checkbox"/> Die Staubteilchen streuen das Licht in alle Richtungen, dadurch trifft es ins Auge und wird sichtbar.</p>

6. **Wie entsteht durch Verwendung einer Sammellinse ein Bild, das auf einem Schirm aufgefangen werden kann?**¹

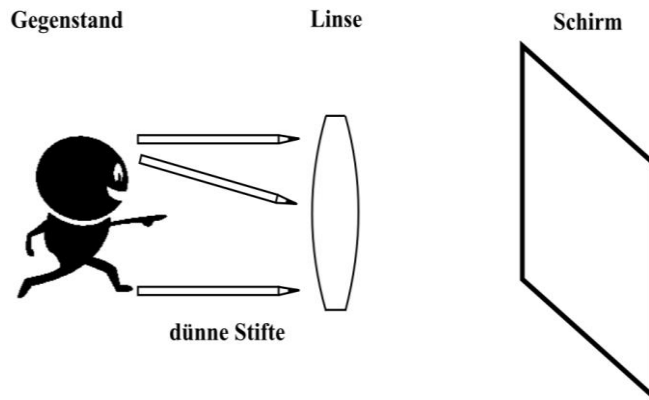
- Solch ein Bild entsteht durch Spiegelung der Lichtstrahlen an der Linse nach dem Reflexionsgesetz.
- Eine Sammellinse hat den Effekt, die Lichtstrahlen aufzuhellen.
- Lichtstrahlen, die von einem Gegenstandspunkt ausgehen, werden durch die Sammellinse abgelenkt und treffen sich im Bildpunkt.
- Das Bild geht als Ganzes durch die Linse zum Schirm, dabei wird es in der Linse unter Einhaltung der Linsengesetze umgedreht (siehe Skizze).



BS

7. **Welche Aussagen zur Bildkonstruktion und Bildentstehung treffen zu?**²

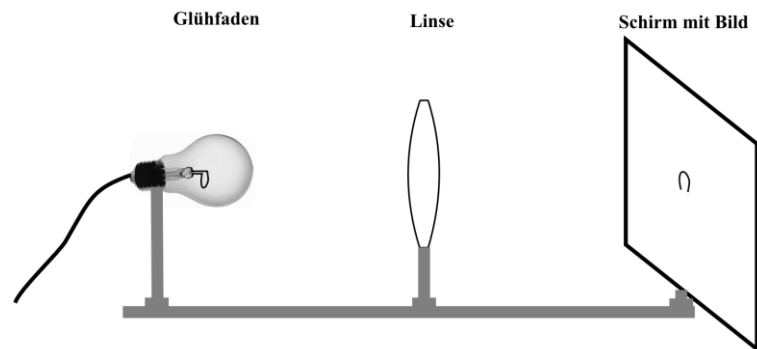
- Nur die ausgezeichneten Strahlen kann man im Strahlengang zeichnen.
- Mit den ausgezeichneten Strahlen kann man den Strahlengang besonders leicht zeichnen.
- Die ausgezeichneten Strahlen erschweren die Zeichnung, machen sie dafür aber besonders genau.
- Ohne die ausgezeichneten Strahlen (wenn diese z. B. durch dünne Stifte aufgehalten werden) kann es kein Bild geben (siehe Abbildung).



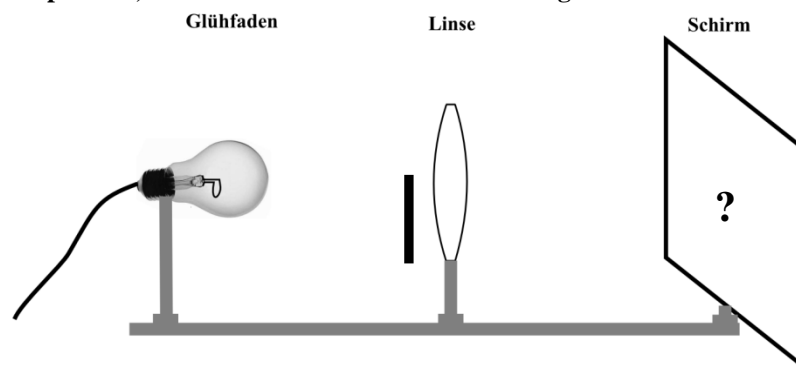
¹ Abbildungen zu Item 6 in Anlehnung an Wiesner, 1986, S. 28

² Abbildungen zu Item 7 in Anlehnung an Wiesner, 1986, S. 28

8. In einer Versuchsanordnung sind eine Glühlampe, eine Sammellinse und ein Schirm so montiert, dass ein vergrößertes, umgekehrtes, scharfes Bild des Glühfadens entsteht:³

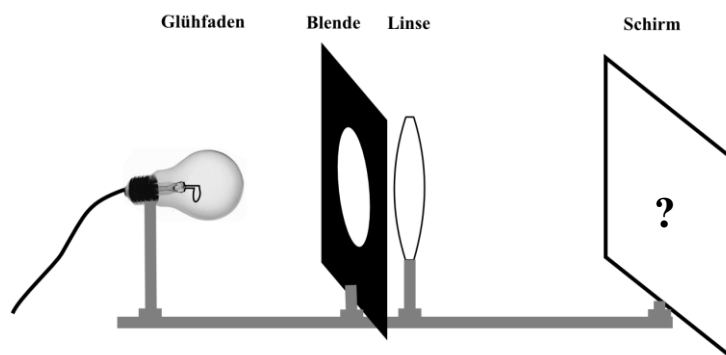


- a) Was passiert, wenn die untere Hälfte der Linse abgedeckt wird?⁴



- Die obere Hälfte des Bildes wird abgeschnitten.
- Die untere Hälfte des Bildes wird abgeschnitten.
- Das Bild wird dunkler.
- Das Bild wird kleiner.

- b) Was passiert, wenn man einen Karton mit großem Loch (ringförmige Blende) vor die Linse hält?⁵



- Das Bild wird kleiner.
- Das Bild wird dunkler.
- Die Ränder des Bildes werden kreisförmig abgeschnitten.
- Das Bild wird heller.

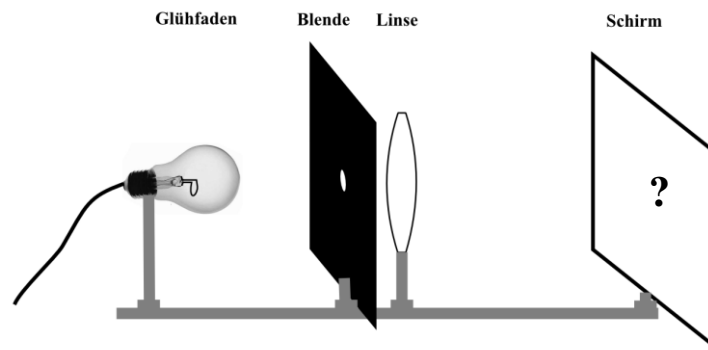
AB

³ Abbildung zu Item 8a-c in Anlehnung an Goldberg und McDermott, 1987, S. 112

⁴ Abbildung zu Item 8a in Anlehnung an Goldberg und McDermott, 1987, S. 112

⁵ Abbildung zu Item 8b Anlehnung an Goldberg und McDermott, 1987, S. 112

- c) Was passiert, wenn man einen Karton mit einem sehr kleinen Loch (5 mm ringförmige Blende) vor die Linse hält? ⁶



- Das Bild wird kleiner.
 Das Bild wird dunkler.
 Die Ränder des Bildes werden kreisförmig abgeschnitten.
 Das Bild wird heller.

Lösungen:

Die Lösungen sind wie folgt angegeben
(Beispiel, hier Item 1)

Welche der folgenden Gegenstände / Lebewesen kann man in einem völlig abgedunkelten Raum?

- x ein leuchtendes Glühwürmchen
 ein weißes Blatt Papier
 einen Fahrrad-Reflektor
 die Augen einer Katze

Lösung: Nr. 1: a)

Tabelle der Lösungen

Item	Antwort
1	a)
2	a)
3	b)
4	a)
5	d)
6	c)
7	b)
8a	c)
8b	b)
8c	b)

REFERENZEN

Abell, N., Springer, D.W. & Kamata, A. (2009). Developing and validating rapid assessment instruments. Oxford: University Press.

Adams, W. K. & Wieman, C. E. (2011). Development and validation of instruments to measure learning of expert-like thinking. *Int. J. Sci. Educ.* 33, 1289-1312.

Allen, K.C. (2006). The Statistic Concept Inventory: Development and Analysis of a Cognitive Assessment Instrument in Statistics, Doctoral Dissertation. Oklahoma: University of Oklahoma.

Anderson, D.L., Fisher, K.M. & Norman, G.J. (2002). Development and Evaluation of the Conceptual

Inventory of Natural Selection. *Journal of Research in Science Teaching.* 39(10), 952-978.

Andersson, B. & Kärrqvist, C. (1983). How Swedish pupils, aged 12-15 years, understand light and its properties. *European Journal of Science Education*, 5(4), 387-402.

Bardar, E. M., Prather, E. E; Brecher, K. & Slater, T. F. (2007). Development and validation of the light and spectroscopy concept inventory. *Astronomy Education Review*, 5(2), 103-113.

Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, & Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (Hrsg.). (2001). PISA, 2000: Schülerleistungen im internationalen Vergleich.

⁶ Abbildung zu Item 8c in Anlehnung an Goldberg und McDermott, 1987, S. 112

- Bektasli, B. (2006). The relationships between spatial ability, logical thinking, mathematics performance and kinematics graph interpretation skills of 12th grade physics students. Doctoral Dissertation. Columbus: Ohio State University.
- Bonanomi, A., Nai Ruscone, M. & Osmetti, S. A. (2013). The Polychoric Ordinal Alpha, measuring the reliability of a set of polytomous ordinal items. SIS, 2013 Statistical Conference. *Vita e Pensiero*. 1-6.
- Braisby, N. & Gellatly, A. (2008). *Cognitive Psychology* Oxford: University Press.
- Bühner, M. (2011). Einführung in die Test- und Fragebogenkonstruktion (3. Aufl.). München, Boston [u.a.]: Pearson Studium.
- Byrnes, J. P., & Wasik, B. A. (1991). Role of conceptual knowledge in mathematical procedural learning. *Developmental Psychology*, 27(5), 777-786.
- Chu, H. E., Treagust, D. & Chandrasegaran, A. L. (2009). A stratified study of students' understanding of basic optics concepts in different contexts using two-tier multiple-choice items. *Research in Science and Technological Education*, 27(3), 253-265.
- Ding, L. & Beichner, R. (2009). Approaches to data analysis of multiple-choice questions. *Physical Review Special Topics-Physics Education Research*, 5(2), 020103.
- Doran, R. L. (1980). *Basic Measurement and Evaluation of Science Instruction*. Washington, DC: National Science Teachers Association.
- Duit, R. & Treagust, D.R. (2003). Conceptual change: a powerful framework for improving science teaching and learning, *International Journal of Science Education*, 25(6), 671-688.
- Duit, R. (2009). Bibliography STCSE. Students and teachers conceptions and science education. <http://archiv.ipn.uni-kiel.de/stcse/>: Datum des letzten Aufrufs 13.04.2020.
- Engelhardt, P. V. & Beichner, R. J. (2004). Students' understanding of direct current resistive electrical circuits. *American Journal of Physics*, 72(1), 98-115.
- EFPA (2013). EFPA Review Model for the description and evaluation of psychological and educational tests. Brussels: European Federation of Psychology Associations.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2011). *Statistik und Forschungsmethoden*. Lehrbuch; mit Onlinematerialien (2. Aufl.). Weinheim [u.a.]: Beltz.
- Epskamp, S., Cramer, A.O.J., Waldorp, L.J., Schmittmann, V.D. & Borsboom, D. (2012) qgraph: Network Visualizations of Relationships in Psychometric Data. *Journal of Statistical Software*, 48 (4), 1-18. <http://www.jstatsoft.org/v48/i04/>: Datum des letzten Aufrufs 15.04.2020.
- Evers, A. (2001). Improving Test Quality in the Netherlands: Results of 18 years of Test Ratings. *International Journal of Testing*, 1, 137-153.
- Fetherstonhaugh, T. & Treagust, D. F. (1992). Students' understanding of light and its properties: Teaching to engender conceptual change. *Science Education*, 76(6), 653-672.
- Fox, J. (2010). Package 'polycor': Polychoric and Polyserial Correlations. Version 0.7-8. Freie Statistiksoftware R. <http://cran.r-project.org/web/packages/polycor/>: Datum des letzten Aufrufs 02.02.2012. Aktualisierte Version (2019) unter: <https://cran.r-project.org/web/packages/polycor/polycor.pdf>. Datum des letzten Aufrufs 02.04.2020.
- Galili, I. (1996). Students' conceptual change in geometrical optics. *International Journal of Science Education*. 18(7), S. 847-868.
- Gerdes, J. & Schecker, H. (1999). Der Force Concept Inventory - Ein diagnostischer Test zu Schülervorstellungen in der Mechanik. *Der Mathematische und Naturwissenschaftliche Unterricht*, 52(5), S. 283-288.
- Goldberg, F. M. & McDermott, L. C. (1987). An investigation of student understanding of the real image formed by a converging lens or concave mirror. *American Journal of Physics*, 55(2), 108-119.
- Guesne, E. (1985). Light. R. Driver, E. Guesne und A. Tiberghien (Hrsg.), *Children's ideas in science* (S. 10-32). Milton Keynes [Buckinghamshire], Philadelphia: Open University Press.
- Guzzetti, B. J., Snyder, T. E., Glass, G. V. & Gamas, W. S. (1993). Promoting Conceptual Change in Science: A Comparative Meta-Analysis of Instructional Interventions from Reading Education and Science Education. *Reading Research Quarterly*, 28(2), 117-155.
- Haagen-Schützenhöfer, C. & Hopf, M. (2012). Entwicklung eines Testinstruments zur geo-metrischen Optik. In Bernholt, S. (Hrsg.), *GDCP Jahrestagung in Oldenburg, 2011: Konzepte fachdidaktischer Strukturierung für den Unterricht*. Münster/New York: LIT Verlag.

- Haagen-Schützenhöfer, C. & Hopf, M. (2014a). Development of a two-tier test-instrument for geometrical optics. In: C. P. Constantinou, N. Papadouris und A. Hadjigeorgiou (Eds.), *E-Book Proceedings of the ES-ERA, 2013 Conference: Science Education Research For Evidence-based Teaching and Coherence in Learning*. (pp. 24-30) (Nicosia/Cyprus: European Science Education Research Association., 2014).
- Haagen-Schützenhöfer, C. & Hopf, M. (2014b). Testing students' conceptual understanding in geometrical optics with a two-tier instrument. In Taşar, M. (Hrsg.), *World Conference of Physics Education, 2012 in Istanbul: Book of Proceedings* (Ankara: Pegem Akademi, 2014); (pp. 1327-1336).
- Haagen-Schützenhöfer, C. (2014c). Students' conceptions on the nature of white light. In Dvorak, L., Koudelková, V. (Hrsg.), *ICPE-EPEC Conference, 2013: Active learning – in a changing world of new technologies* (MATFYZPRESS publisher, Prague, 2014); http://iupap-icpe.org/publications/proceedings/ICPE-EPEC_2013_proceedings.pdf: Datum des letzten Aufrufs 29.09.2014.
- Haagen-Schützenhöfer, C. & Hopf, M. (2018) Schülervorstellungen zur geometrischen Opti. (S. 89 – 114). In: H. Schecker., T. Wilhelm, M. Hopf, M.& R. Duit. *Schülervorstellungen und Physikunterricht: Ein Lehrbuch für Studium, Referendariat und Unterrichtspraxis*. Springer Spektrum: Berlin, Heidelberg.
- Haladyna, T. M., & Downing, S. M. (1989). The validity of a taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, 1, 51-78.
- Härtig, H. (2014). Der Force Concept Inventory Vergleich einer offenen und einer geschlossenen Version. *PhyDid A-Physik und Didaktik in Schule und Hochschule*, 1(13), 53-61.
- Herdt, D. (1990). Einführung in die elementare Optik. Vergleichende Untersuchung eines neuen Lehrgangs. Essen: Westarp-Wissenschaftsverlag.
- Hestenes, D., Wells, M. & Swackhamer, G. (1992). Force concept inventory. *Phys. Teach*, 30(3), 141-158.
- Hettmannsperger, R. (2015). Lernen mit multiplen Repräsentationen aus Experimenten: Ein Beitrag zum Verstehen physikalischer Konzepte. Springer Verlag Sozialwissenschaften: Wiesbaden.
- Heywood, D. S. (2005). Primary trainee teachers' learning and teaching about light: Some pedagogic implications for initial teacher training. *International Journal of Science Teaching*, 27(12). 1447-75.
- Horn, J. (1965). A rationale and test fort the number of factors in factor analysis. *Psychometrika* (30), 179-185.
- James, L. R., Demaree, R. G. & Wolf, G. (1984). Estimating withingroup interrater reliability with and without response bias. *Journal of Applied Psychology*, 69(1), 85-98.
- Jorion, N., Gane, B. D., James, K., Schroeder, L., DiBello, L. V., & Pellegrino, J. W. (2015). An analytic framework for evaluating the validity of concept inventory claims. *Journal of Engineering Education*, 104(4), 454-496.
- Jung, W. (1981). Ergebnisse einer Optik-Erhebung. *Physica Didactica* 9. 19-34.
- Kircher, E.; Girwitz, R.; Häußler, P. (JAHR) *Physikdidaktik - Theorie und Praxis*, Springer Verlag, Berlin, Heidelberg, New York.
- Kline, Th. J. B. (2005). *Psychological Testing. A Practical Approach to Design and Evaluation*. Thousand Oaks, London, New Delhi: Sage.
- Kline, P. (2015). *A Handbook of Test Construction: Introduction to psychometric design* London: Routledge.
- Langley, D., Ronen, M. & Eylon, B.-S. (1997). Light propagation and visual patterns: Preinstruction learners' conceptions. *Journal of Research in Science Teaching*, 34(4), 399-424.
- Lasry, N., Rosenfield, S., Dedic, H., Dahan, A. & Reshef, O. (2011). The puzzling reliability of the Force Concept Inventory. *American Journal of Physics*, 79(9), 909-912.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse*. 6. Aufl. Weinheim: Beltz. Psychologie Verl.-Union.
- Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: a critical appraisal. *Learning and Instruction*, 11, (4-5). 357-380.
- Lindell, M. K., Brandt, Ch. J. & Whitney, D. J. (1999). A Revised Index of Interrater Agreement for a Single Target. *Applied Psychological Measurement*, 23(2), 127-135.
- Lindell, R. S., Peak, E. & Foster, Th.M. (2007). Are They All Created Equal? A Comparison of Different Concept Inventory Development Methodologies., 2006 *Physics education research conference. AIP Conference Proceedings*, Volume 883, 14-17.

- Liu, X. (2012). Developing Measurement Instruments for Science Education Research. Second International Handbook of Science Education. (Springer International Handbooks of Education). Barry J. Fraser, Kenneth Tobin, Campbell J. McRobbie (Eds.). Springer. Dordrecht. 651-665.
- Madsen, A., McKagan, S. B., & Sayre, E. C. (2017). Resource Letter RBAI-1: Research-Based Assessment Instruments in Physics and Astronomy. *American Journal of Physics*, 85(4), 245-264.
- Maloney, D. P., O’Kuma, T. L., Hieggelke, C. J. & van Heuvelen, A. (2001). Surveying students’ conceptual knowledge of electricity and magnetism. *Am. J. Phys. Phys. Educ. Res., Suppl.* 69, 12-23.
- Mandl, H. & Spada, H. (Hrsg.). (1988): Wissenspsychologie. Psychologie Verlags-Union, München.
- Martinez-Borreguero, G., Pérez-Rodríguez, Á. L., Suero-López, M. I., & Pardo-Fernández, P. J. (2013). Detection of misconceptions about colour and an experimentally tested proposal to combat them. *International Journal of Science Education*, 35(8), 1299-1324.
- Meschede, D. (2006) (Hrsg.) Gerthsen Physik. (23. Aufl.) Berlin: Springer.
- Moosbrugger, H., & Kelava, A. (2012). Testtheorie und Fragebogenkonstruktion. Berlin: Springer.
- Morris, G. A., Branummartin, L., Harshman, N. et al. (2006). Testing the test: item reponse curves and test quality. *American Journal of Physics*, 74, 449-453.
- Mulford, D. R. & Robinson, W. R. (2002). An Inventory for Alternate Conceptions among First-Semester General Chemistry Students. *J. Chem. Educ.*, 79(6), 739-744.
- Naranjo-Correa, F. L., Martinez-Borreguero, G., Perez-Rodriguez, A. L., Lopez, S., Isabel, M., & Pardo Fernandez, P. J. (2015). A new online tool to detect color misconceptions. *Color Research & Application*.
- Nersessian, N. J. (1992). How Do Scientists Think? Capturing the Dynamics of Conceptual Change in Science. In: R. N. Giere (Hrsg.), *Cognitive models of science: Vol. 15. Minnesota studies in the philosophy of science* (pp. 129-186) Minneapolis, MN: University of Minnesota Press.
- Nieminen, P., Savinainen, A., & Viiri, J. (2010). Force concept inventory-based multiple-choice test for investigating students’ representational consistency. *Physical Review Special Topics Physics Education Research*, 6(2), 1-12.
- Özdemir, G. & Clark, D. B. (2007). An Overview of Conceptual Change Theories. *Eurasia Journal of Mathematics, Science and Technology Education*, 3(4), 351-361.
- Ramlo, S. (2008). Validity and reliability of the force and motion conceptual evaluation. *American Journal of Physics*, 76(9), 882-886.
- Rebello, N. S. & Zollman, D. A. (2004). The effect of distracters on student performance on the force concept inventory. *American Journal of Physics*, 72(1), 116-125.
- Reiner, M., Slotta, J. D., Chi, M. T. H. & Resnick, L. B. (2000). Naive Physics Reasoning: A Commitment to Substance-Based Conceptions. *Cognition and Instruction*, 18(1), 1-34.
- Revelle, W. (2013) psych [Computer Software]: freie Statistiksoftware R. <http://cran.r-project.org/web/packages/psych/index.html>: Datum des letzten Aufrufs 22.09.2013. Neueste Version: Revelle W (2019). psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois. R package version 1.9.12, <https://CRAN.R-project.org/package=psych>. Datum des letzten Aufrufs 17.04.2020.
- Rodriguez, M. C. (2005). Three Options Are Optimal for Multiple-Choice Items: A Meta-Analysis of 80 Years of Research. *Educational Measurement: Issues and Practice*, 24(2) 3-13.
- Rosseeel, Y. (2012). Package lavaan [Computer software]: freie Statistiksoftware R. <https://users.ugent.be/~yrosseel/lavaan/lavaanIntroduction.pdf>. Datum des letzten Aufrufs 18.04.2020.
- Rost, J. (2004). Lehrbuch Testtheorie - Testkonstruktion. 2. Auflage Bern, Göttingen [u.a]: Huber.
- Savinainen, A., & Viiri, J. (2008). The Force Concept Inventory as a measure of students conceptual coherence. *International Journal of Science and Mathematics Education*, 6(4), 719-740.
- Schecker, H. (o.J.). Geometrische Optik, Testaufgaben. <http://www.idn.uni-bremen.de/schuelervorstellungen/>. Datum des letzten Aufrufs 02.04.2020.
- Scheid, J. (2013). Multiple Repräsentationen, Verständnis physikalischer Experimente und kognitive Aktivierung: Ein Beitrag zur Entwicklung der Aufgabenkultur. Dissertationsschrift zur Erlangung des Doktors der Philosophie. In Niedderer, H., Fischler, H., Sumfleht E. (Hrg.). *Studien zum Physik- und Chemielernen Band*, 151, Berlin: Logos Verlag.

- Scheid, J., Müller, A., Hettmannsperger, R. & Schnotz, W. (2019). Improving learners' representational coherence ability with experiment-related representational activity tasks. *Physical review of physics education research* 15, 010142. DOI: 10.1103/PhysRevPhysEdu-Res.15.010142
- Scott, P. H., Adams, H., & Leach, J. (2007). Student Conceptions and Conceptual Learning in Science. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of Research on Science Education* (pp. 31-54). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Scott, T. F., Schumayer, D., & Gray, A. R. (2012). Exploratory factor analysis of a Force Concept Inventory data set. *Physical Review Special Topics-Physics Education Research*, 8(2), 020105.
- Selley, N. J. (1996). Children's ideas on light and vision. *International Journal of Science Education*, 18(6), 713-723.
- Shtulman, A., & Lombrozo, T. (2016). Bundles of contradiction: A coexistence view of conceptual change. In Barner, D., & Barron, A. S. (Eds.), *Core knowledge and conceptual change* (pp. 53-72). New York: Oxford University Press.
- Sokoloff, D. R. (2006). Action Research and the Light and Optics Conceptual Evaluation. (227 – 243). In: *Active learning in optics and photonics: training manual*. D.R. Sokoloff (Ed.). ALOP Manuel Training English version, UNESCO 2006. <https://unesdoc.unesco.org/ark:/48223/pf0000217100>. Datum des letzten Aufrufs 26.04.2020.
- Stewart, J., Griffin, H. & Stewart, G. (2007). Context sensitivity in the force concept inventory. *Physics Review Special Topics – Physics Education Research*, 3, 010102.
- Strobl, C. (2012). *Das Raschmodell. Eine verständliche Einführung für Studium und Praxis*. München und Mering: Rainer Hampp.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Boston: Allyn and Bacon.
- Tyson, L. M., Venville, G. J., Harrison, A. G. & Treagust, D. F. (1997). A Multidimensional Framework for Interpreting Conceptual Change Events in the Classroom. *Science Education*, 81(4), 387-404.
- Wiesner, H. (1986). Schülervorstellungen und Lernschwierigkeiten im Bereich der Optik. *Naturwissenschaften im Unterricht - Physik*, 34(13), 25-29.
- Wiesner, H. (1992a). Schülervorstellungen und Lernschwierigkeiten mit dem Spiegelbild. *Naturwissenschaften im Unterricht - Physik*, 3(14), 16-18.
- Wiesner, H. (1992b). Verbesserung des Lernerfolgs im Unterricht über Optik (1). *Physik in der Schule*, 30(9), 286-290.
- Wiesner, H. (1994). Ein neuer Optikkurs für die Sekundarstufe I, der sich an Lernschwierigkeiten und Schülervorstellungen orientiert. *Naturwissenschaften im Unterricht - Physik*, 5(22), 7-15.
- White, R. T. & Gunstone, R. F. (1989). Metalearning and conceptual change. *International Journal of Science Education*, 11(5), 577-586.
- Wilcox, R. R. (1981). Analyzing the Distractors of Multiple-Choice Test Items or Partitioning Multinomial Cell Probabilities with Respect to a Standard. *Educational and Psychological Measurement*, 41(4), 1051-1068.
- Wirtz, M. & Caspar, F. (2002). Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen. München: Hogrefe.
- Yeo, S., & Zadnik, M. (2001). Introductory Thermal Concept Evaluation: Assessing Students' Understanding. *The Physics Teacher*, 39, 495-504. Ins Deutsche übersetzt von Engelke, T. abrufbar unter http://www.didaktik.physik.uni-muenchen.de/forschung/testdatenbank/inhalt_testdatenbank/verst_waermelehre.pdf: Datum des letzten Aufrufs 02.05.2015. (passwortgeschützt).
- Zeilik, M., Schau, C., Mattern, N., Hall, S., Teague, K. W. & Bisard, W. (1997). Conceptual astronomy: A novel model for teaching postsecondary science courses. *American Journal of Physics*, 65(10), 987-996.

KONTEXT UND PROBLEMLÖSEN – EINE PROZESSANALYSE

Patrick Löffler¹, Alexander Kauertz¹

¹Universität Koblenz-Landau

*Please address all correspondence to Patrick Löffler, loeffler@uni-koblenz.de

STRUCTURED ABSTRACT

Hintergrund: Lernen mit Kontexten im engeren Sinne beinhaltet typischerweise problemorientierte Aufgaben, die durch die Anwendung von fachwissenschaftlichen Modellen gelöst werden müssen. Es bestehen wenig Zweifel an den positiven Effekten solcher kontextbasierten Ansätze auf affektive Variablen (z.B. Interesse). Gleichzeitig ist die beobachtete Varianz in der Leistung bislang nur unzureichend erklärt. In verschiedenen Ansätzen wurde seither versucht, diese Forschungslücke zu schließen; beispielsweise durch Untersuchung des Zusammenspiels von Aufgabenmerkmalen wie Authentizität oder Relevanz und deren Wechselwirkung mit Personenmerkmalen. Nach wie vor fehlt jedoch eine Erklärung für die heterogenen Befunde zum Einfluss von Kontext auf die Leistung im Problemlöseprozess.

Ziel: Leistungsmessung beim Problemlösen findet meist auf Ebene des gesamten Prozesses statt. Der Einfluss von Kontext kann daher auch nur in Bezug zu dieser übergreifenden Messung untersucht werden. Im Gegensatz dazu wird in der vorliegenden Studie erforscht, wie sich Kontext auf die Leistung in den einzelnen Phasen des Problemlöseprozesses auswirkt (Phase 1: Das Problem verstehen, Phase 2: Die Bedingungen/Variablen identifizieren, Phase 3: Die Auswahl begründen, Phase 4: Lösungsvorschlag formulieren, Phase 5: Eine Antwort auf das Problem geben, Phase 6: Weitere Argumente für die Richtigkeit der Lösung nennen). Dazu wird unter diesem Gesichtspunkt eine Reanalyse einer aktuellen Studie vorgenommen um die Frage zu beantworten: Welchen Einfluss hat Kontext auf die Leistung in diesen Phasen?

Stichprobe: Die Teilnehmer*innen besuchen die 10. Klassenstufe dreier Gymnasien aus Rheinland-Pfalz in Deutschland. Das Alter liegt zwischen 14 und 17 Jahren mit einem Schnitt von 15.4. Die Stichprobengröße beträgt $n = 123$ (63m, 60w).

Design und Methode: In dem Experiment wird der Einfluss der unabhängigen Variable *Kontextualisiertheit* auf die abhängige Variable *Leistung im Problemlöseprozess* für jede der sechs Phasen im Problemlöseprozess untersucht. Dabei werden folgende Personenmerkmale kontrolliert: Kognitive Fähigkeiten und konzeptuelles Vorwissen. Die Aufgabenmerkmale *Transparenz* sowie *Komplexität* werden ebenfalls als Kontrollvariablen in die Analyse aufgenommen. Um eine Alpha-Fehler-Inflation bei mehreren Regressionen auszuschließen, wird zunächst eine multivariate Kovarianzanalyse (MANCOVA) durchgeführt. Daran anschließend werden sechs hierarchische Regressionen vorgenommen und die Regressionsparameter angegeben. Als Maß für die Effektstärke wird Cohens f^2 bestimmt. Reliabilität und Validität der verwendeten Instrumente wurden in vorherigen Studien überprüft.

Ergebnis: *Kontextualisiertheit* hat einen signifikanten Einfluss auf die Leistung im Problemlöseprozess (Pillai's Trace: $V = 0.227$, $F(6, 112) = 5.483$, $p < .001$). Die Regressionsanalysen zeigen dabei einen positiven Einfluss auf die erste und dritte Phase des Problemlöseprozesses (1. Phase: $p < .023$, $f^2 = .035$, $\beta = .208$, 3. Phase: $p < .030$, $f^2 = .078$, $\beta = .194$). Auf die zweite und vierte Phase hat *Kontextualisiertheit* einen negativen Einfluss (2. Phase: $p < .017$, $f^2 = .042$, $\beta = -.198$, 4. Phase: $p < .018$, $f^2 = .040$, $\beta = -.200$). Auf die fünfte und sechste Phase kann keine signifikante Wirkung beobachtet werden.

Schlussfolgerung: Die Ergebnisse zeigen, dass die Wirkung von Kontext auf die Leistung im Problemlöseprozess stark variiert. Die positiven Effekte auf die Phasen 1 und 3 sind in etwa gleich stark wie die negativen Effekte auf die Phasen 2 und 4 und heben sich daher gegenseitig auf der Ebene des gesamten Prozesses auf. Dieses Ergebnis kann dabei helfen, die bisherigen, teils widersprüchlichen Befunde in der Kontextforschung zu erklären. Darüber hinaus gibt das Ergebnis Anlass zur Forderung, in zukünftiger Kontextforschung detailliert den Prozess zu untersuchen. Für den Unterricht ergibt sich die Folgerung, Lernende insbesondere in den Phasen 2 und 4 zu unterstützen.

Schlüsselwörter: Kontext, Modellanwendung, Problemlösen, Aufgaben.

Received: May 2020. **Accepted:** November 2020.

1 EINLEITUNG

Kontexte beim Problemlösen sind seit einigen Jahrzehnten ein wichtiger Bestandteil des naturwissenschaftlichen Lernens geworden (Pilot & Bulte, 2006): Sie ermöglichen es Lehrkräften, negativen Einstellungen gegenüber Naturwissenschaft zu begegnen (Bennett et al., 2007). Vor allem gibt es aber viele Argumente dafür, wie solche Aufgaben die Entwicklung des konzeptionellen Verständnisses fördern sollten: Lubben, Campbell, & Dlamini (1996) nennen beispielsweise die fokussierte Aufmerksamkeit von Lernenden bei der Verwendung von Kontexten mit einer empfundenen persönlichen Relevanz sowie deren erhöhte Mitarbeit bei der Problemlösung. Yager (1999) berichtet von besseren Fähigkeiten im Anwenden von Gelerntem. Gleichzeitig gibt es auch gute Gründe, das Gegenteil anzunehmen: Harp & Mayer (1998) sehen ansprechende Details in Kontexten als hinderlich: durch diese „verführerischen Details“ wird irrelevantes Vorwissen aktiviert, welches dann in die Überlegungen zur Problemlösung einfließt. Taasobshirazi & Carr (2008) kommen zu ähnlichen Überlegungen, wonach Kontexte zwar interessant sein sollten, jedoch auf eine Art und Weise, dass der fachliche Inhalt unterstützt und nicht davon abgelenkt wird. Zudem befürchten sie, Kontext könne Transfer unterdrücken, was im Gegensatz zu den Erkenntnissen von Yager (1999) steht. Folgerichtig werden in Studien zu diesem Thema heterogene Ergebnisse beobachtet (Bennett et al., 2007). In einer Metastudie kommen Park und Lee daher zu dem Schluss: “[...] it is difficult to observe a consistent effect of contexts on physics problem-solving.” (Park & Lee, 2004, p. 1586–1586). Seitdem haben mehrere Studien versucht, diese Forschungslücke zu schließen: Mögliche Themeneffekte (Gomez, Pozo, & Sanz, 1995; Kölbach, 2011), Aufgabenmerkmale (Löffler, 2016) sowie deren Zusammenspiel mit den Personenmerkmalen (van Vorst et al., 2014) und unterschiedliche Kontextdefinitionen wurden untersucht (Finkelstein, 2005; Gilbert, 2006). Zusammenfassend kann jedoch festgestellt werden, dass der Einfluss von Kontext auf die Leistung noch immer nicht ausreichend verstanden ist. Ein möglicher Grund für dieses Problem liegt in der Leistungsmessung, die in der Regel auf dem Ergebnis des Problemlösungsprozesses und nicht auf den einzelnen Phasen des Lösungsprozesses basiert. Dafür spricht auch, dass sich die zuvor beispielhaft genannten Einflussfaktoren bestimmten Phasen im Problemlöseprozess zuordnen lassen: Beispielsweise findet Aktivierung von Vorwissen vornehmlich zu Beginn statt.

Daher versuchen wir, mit einer Reanalyse die Frage "Welchen Einfluss hat Kontext in diesen Phasen?" zu beantworten. Die Leistung der Lernenden beim kontextualisierten Problemlösen wurde in der untersuchten Studie (Löffler, 2016) als Summenwert durch die Bewertung typischer Schritte im Problemlösungsprozess gemessen, was eine detailliertere Betrachtung im Nachhinein ermöglicht.

2 HINTERGRUND

Arbeiten mit Modellen ist ein zentrales Element naturwissenschaftlichen Unterrichts und somit auch der fachdidaktischen Forschung. Merkmale der Modellierungsanlässe, wie z.B. deren Bedeutsamkeit, Authentizität oder allgemein Sinnhaftigkeit (z.B. van Vorst et al., 2014), haben dabei zunehmend an Bedeutung gewonnen und stellen unter dem Überbegriff Kontext bereits seit vielen Jahren ein eigenes Forschungsfeld dar. Je nach Forschungsschwerpunkt werden allerdings unterschiedliche Kontextdefinitionen verwendet. Dennoch finden sich in diesen Definitionen wiederkehrende Elemente, die sich vor allem auf zwei miteinander verbundene Ebenen konzentrieren (z.B. Mestre, 2002 oder van Vorst, Dorschu, Fechner, Kauertz, Krabbe & Sumfleth, 2014): Kontext ist demnach ein realistisches Problem mit Relevanz für die Lernenden (Level 1), welches mit wissenschaftlichen Modellen gelöst werden kann (Level 2). Dieser Ansatz ermöglicht die Unterscheidung der Ebene der Problemsituation bzw. *surface structure* von der Ebene der angewandten fachwissenschaftlichen Modelle bzw. *deep structure* (Löffler, 2016; Löffler & Kauertz, 2014). In der untersuchten Studie von Löffler (2016) wurden daher Aufgabenmerkmale auf der Ebene *surface structure* sowie *deep structure* untersucht. Das zugrundeliegende Kontextmodell wird im Folgenden dargestellt.

2.1 Kontextmodell

Um den Einfluss von Kontext auf die Leistung im Problemlösen zu untersuchen, müssen wir die Vorstellung von Kontext weiter konkretisieren. Dazu betrachten wir die beiden zuvor genannten Ebenen *surface structure* und *deep structure* und analysieren deren Komponenten sowie den damit verbundenen Schwierigkeitseffekt.

Ausgehend von dem Ziel, strukturelle Beziehungen zwischen konzeptuellen und konkreten Objekten herzustellen (Bunge, 1973), besteht die grundlegende Intention des Kontextmodells darin, Informationssätze auf den Ebenen der Oberflächen- und Tiefenstruktur zu identifizieren und Teilmengen beider Ebenen miteinander zu verbinden. Die Informationen können nach zwei Dichotomien klassifiziert werden (Löffler, Pozas & Kauertz, 2018): erstens entweder als Elemente fachwissenschaftlicher Modelle (physikalischer Inhalt) oder als Elemente, die nicht Teil fachwissenschaftlicher Modelle (Kontext) sind, und zweitens als "Teil der Lösung" oder "nicht Teil der Lösung", von denen letztere beispielsweise dekorative oder ablenkende Elemente beinhaltet (vgl. z.B. „seductive details“ Harp & Mayer, 1998). Aus diesen Dichotomien ergeben sich daher vier Informationssätze (Tab. 1).

Die daraus abgeleiteten Informationen lassen sich durch drei Aufgabenmerkmale beschreiben: *Kontextualisiertheit* (Löffler & Kauertz, 2015), *Komplexität* (Kauertz, 2008) und *Transparenz* (Löffler & Kauertz, 2014).

Kontextualisiertheit bezieht sich auf das Vorhandensein und die Menge von Informationssätzen <nML> und <nMnL>.

Komplexität wird bestimmt durch die Anzahl und die Verbindungen der Elemente in <ML> und hat einen Einfluss auf die Aufgabenschwierigkeit (Kauertz & Fischer, 2006 oder Walpuski, Ropohl & Sumfleth, 2011).

Tab. 1. Informationssätze in Kontextaufgaben: ML = Elemente fachwissenschaftlicher Modelle & Teil der Lösung; MnL = Elemente fachwissenschaftlicher Modelle & nicht Teil der Lösung; nML = nicht Elemente fachwissenschaftlicher Modelle & Teil der Lösung; nMnL = nicht Elemente fachwissenschaftlicher Modelle & nicht Teil der Lösung (aus dem Englischen übersetzt, Löffler, Pozas & Kauertz 2018).

	Element fw. Modelle	Nicht Element fw. Modelle
Teil der Lösung	ML	nML
Nicht Teil der Lösung	MnL	nMnL

Die *Transparenz* einer Aufgabe beschreibt, ob:

- Elemente aus <ML> explizit im Text genannt werden,
- Informationssätze <MnL> existieren und so die Unterscheidung zwischen den Sätzen nötig machen,
- es zur Lösung ausreichend ist Elemente aus <nML> mit Elementen aus <ML> zu verknüpfen (Modellanwendung), oder ob
- fehlende Informationen auf Basis der Elemente aus <nML> gefunden werden müssen (Modellentwicklung).

Kontextualisiertheit und *Komplexität* sind offensichtlich hierarchisch messbar. Zur *Transparenz* sind weitere Ausführungen nötig: Wenn die *surface structure* einer Aufgabe bereits Elemente aus <ML> konkret anführt und gleichzeitig keine Elemente aus <MnL> enthalten sind, müssen Lernende nicht zwischen beiden Informationssätzen unterscheiden und die Aufgabe hat eine hohe *Transparenz*. Im umgekehrten Fall bedeutet eine niedrige *Transparenz*, dass Lernende den Distractionen von Elementen aus <MnL> widerstehen müssen und gleichzeitig die fehlenden Elemente von <ML> auf Basis von <nML> finden müssen. Diesem Beispiel folgend, wurde in der untersuchten Studie (Löffler, 2016) *Transparenz* dichotom kodiert um in den genannten Kombinationen den größtmöglichen Unterschied zu erzeugen.

Aus dem Ansatz der in fast allen Kontextdefinitionen implizit oder explizit aufgeführten Ebenen *surface structure* und *deep structure*, sowie den darauf definierten Aufgabenmerkmalen, ergibt sich so das der Studie (Löffler, 2016) zugrundeliegende Kontextmodell (Abb. 1). Zur besseren Verständlichkeit der Anwendung dieses Modells werden einige Beispiele aufgeführt. In der folgenden Tabelle (Tab. 2) sind unterschiedliche Ausprägungen des Aufgabenmerkmals *Kontextualisiertheit* dargestellt.

In der Aufgabenversion mit niedriger *Kontextualisiertheit* gibt es keine Elemente aus <nML> oder <nMnL>. Im Gegensatz dazu tauchen Elemente aus beiden Informationssätzen in der hoch kontextualisierten Variante auf.

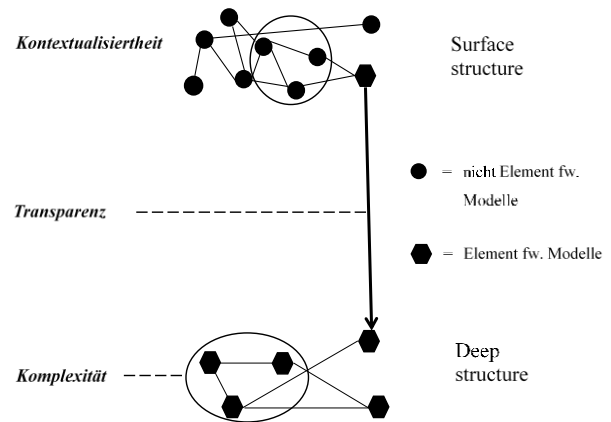


Abb. 1. Beispiel Kontextmodell (Löffler, Pozas & Kauertz, 2018) mit hoher *Kontextualisiertheit*, niedriger *Transparenz* und hoher *Komplexität*. Umkreiste Elemente sind zur Problemlösung nötig.

Das Aufgabenmerkmal *Komplexität* wird anhand der Lösung einer Aufgabe auf Ebene der *deep structure* bestimmt (Satz <ML>) und erlaubt Rückschlüsse auf die Aufgabenschwierigkeit. In der folgenden Tabelle (Tab. 3) werden der Verständlichkeit halber auch mögliche Probleme aufgeführt (obwohl das eigentliche Problem zur Bestimmung der *Komplexität* nicht bekannt sein muss).

Der Informationssatz <ML> in der Lösung der niedrig-komplexen Variante enthält lediglich ein Fakt (Kauertz & Fischer, 2006), was auf eine geringere Aufgabenschwierigkeit hinweist. Die Elemente aus <ML> der hochkomplexen Lösung lassen dagegen Rückschlüsse auf eine erhöhte Aufgabenschwierigkeit zu.

Tab. 2. Beispiele für *Kontextualisiertheit* (aus dem Englischen übersetzt, Löffler, Pozas & Kauertz, 2018)

Problem	Kontextualisiertheit	Informationssätze
Zwei Geräte haben eine Leistung von $P_1 = 2,5 \text{ kW}$ und $P_2 = 1,5 \text{ kW}$. Berechne die Stromstärke, wenn die Spannung $U = 230 \text{ V}$ beträgt.	Niedrig	Nur <ML>: Leistung, Stromstärke, $U = 230 \text{ V}$, ...
Herr Decker kauft einen Heizstrahler und einen Heizlüfter für den Wickeltisch im Kinderzimmer. Beide Geräte werben mit ihrer hohen elektrischen Leistung ($2,5 \text{ kW}$ und $1,5 \text{ kW}$) und er fragt sich, ob die Sicherung (16 A) hält, wenn er beide an die gleiche Steckdose anschließt.	Hoch	Satz <ML>: Elektrische Leistung, $2,5 \text{ kW}$, $1,5 \text{ kW}$, 16 A Satz <nML>: Heizstrahler, Heizlüfter, Steckdose, Sicherung Satz <nMnL>: Hr. Decker, Wickeltisch, Kinderzimmer

Tab. 3. Beispiele für *Komplexität* von Lösungen (aus dem Englischen übersetzt, Löffler, Pozas & Kauertz, 2018)

Problem	Lösung	Komplexität	Elemente aus <ML>
Um Wärmeverlust in kalten Umgebungen zu verhindern, plustern Vögel ihre Federn auf, um eine Luftschicht um den Körper zu erzeugen. Erkläre, warum das hilfreich ist.	Luft ist ein schlechter Wärmeleiter.	Niedrig	Ein Fakt (Kauertz & Fischer, 2006): Luft ist ein schlechter Wärmeleiter
Eine Thermosflasche hält Kaffee lange warm. Sie besteht aus zwei Flaschen: Eine davon ist in der anderen und am Flaschenhals sind beide verbunden. Aus der Lücke dazwischen wird die Luft abgepumpt. Die innere Flasche ist mit einer reflektierenden Schicht überzogen. Erkläre, warum diese Konstruktion Kaffee warmhält.	Um Wärmeübertragung zu minimieren, müssen Wärmeleitung, Wärme-strömung und Wärme-strahlung verringert werden	Hoch	Mehrere verbundene Fakten: Wärme-strahlung, Wärme-strömung und Wärmelei-tung ergeben zusammen Wärmeüber-tragung

Tab. 4. Beispiele für *Transparenz* in Problemlöseaufgaben und deren Lösungen (aus dem Englischen übersetzt, Löffler, Pozas & Kauertz, 2018)

Problem	Lösung	Transparenz
Julia springt vom 10m-Brett. Wie schnell wird sie auf das Wasser treffen, wenn ihre Masse 52 kg beträgt?	Für den freien Fall ohne Anfangsgeschwindigkeit gilt: $v = \sqrt{2gh}$. Mit $h = 10\text{ m}$ und $g \approx 9.81\text{ m/s}^2$ beträgt die	Niedrig
Julia springt vom Sprungbrett (Höhe $h = 10\text{ m}$). Mit welcher Geschwindigkeit v wird sie auf das Wasser treffen unter der Voraussetzung, dass die Erdbeschleunigung $g \approx 9.81\text{ m/s}^2$ beträgt?	Endgeschwindigkeit: $v \approx 14\text{ m/s}$ oder 50 km/h .	Hoch

In Tab. 4 sind Beispiele zur Veranschaulichung des Aufgabenmerkmals *Transparenz* aufgeführt. In der hochtransparenten Version des Problems werden nur Elemente fachwissenschaftlicher Modelle genannt, die Teil der Lösung sind (Satz <ML>): Höhe, Geschwindigkeit und Beschleunigung. Im Gegensatz dazu enthält die Version mit niedriger *Transparenz* Elemente fachwissenschaftlicher Modelle, die nicht Teil der Lösung sind (Satz <MnL>) und von der eigentlichen Lösung ablenken (Masse ist irrelevant für die Endgeschwindigkeit unter Vernachlässigung des Luftwiderstandes beim freien Fall aus niedriger Höhe). Gleichzeitig müssen Lernende in dieser Aufgabenvariante fehlende Elemente des Satz <ML> auf Grundlage der Elemente aus <nML> finden: „schnell“ → Geschwindigkeit, „10m-Brett“ → Höhe.

2.2 Problemlöseprozess

Diese verschiedenen Informationssätze und Elemente werden im Problemlöseprozess durch die Lernenden genutzt, um Lösungen zu generieren. Um den Zusammenhang zwischen den Informationssätzen und Elementen sowie dem Problemlöseprozess herzustellen, wird dieser zunächst näher charakterisiert:

In der untersuchten Studie von Löffler (2016) wurden typische Schritte im analytischen Problemlösungsprozess untersucht (vgl. Pólya, 1985 oder Abrams, 2001). Ziel war, die Leistung im Problemlösen über eine Bewertung der Lösung hinaus zu operationalisieren und so die Ursachen für erfolglose Problemlösungsprozesse untersuchbar zu machen. Dazu passte Löffler (2016) ein Instrument von Charles, Lester und O'Daffer (1987) an, um die Leistung der Schülerinnen und Schüler durch die Bewertung schriftlicher Antworten in sechs typischen Phasen eines Problemlösungsprozesses zu erfassen (siehe Tab. 5). Der daraus ermittelte Summenscore wurde von Löffler (2016) als abhängige Variable in eine Regressionsanalyse aufgenommen. Das Aufgabenmerkmal *Kontextualisiertheit* zeigte dabei entgegen der Hypothesen keinen signifikanten Einfluss. Durch Reanalyse der detaillierteren Prozessdaten soll nun mit dieser Studie die Frage, „Welchen Einfluss hat Kontext in diesen Phasen?“, beantwortet werden.

Tab. 5. Phasen des analytischen Problemlöseprozesses und daraus abgeleitete Items. “Bewertung” dient zur Veranschaulichung des detaillierten Beurteilermanuals. Beurteiler-Reliabilität Kendall’s Tau-b je nach Phase: $.48 \leq \tau_B \leq .97$, (Löffler, 2016).

Phase	Beschreibung	Item	Bewertung & τ_B
1	Das Problem verstehen	Formuliere das Problem zunächst mit eigenen Worten in Form einer Frage	Beurteiler entscheiden, wie gut das formulierte Problem zum intendierten Problem der Aufgabe passt. $.72 \leq \tau_B \leq .74$
2	Die Bedingungen/ Variablen identifizieren	Nenne drei Begriffe, die du zum Lösen des Problems am wichtigsten findest (das können Begriffe aus dem Text sein oder andere, z.B. aus der Physik).	Beurteiler schätzen die Relevanz der genannten Begriffe durch Vergleich mit der Lösung und anhand von Bewertungsbeispielen. $.89 \leq \tau_B \leq .92$
3	Die Auswahl begründen	Erkläre mit ganzen Sätzen für jeden dieser Begriffe, wieso er zur Lösung der Aufgabe wichtig ist.	Beurteiler entscheiden wie nützlich die Erklärungen im Hinblick auf die Lösung sind. $.48 \leq \tau_B \leq .55$
4	Lösungsvorschlag formulieren	Formulieren mit ganzen Sätzen deinen Lösungsvorschlag für das Problem. Du darfst die drei von dir	Beurteiler bewerten wie gut der Lösungsvorschlag zur Lösung passt. $.92 \leq \tau_B \leq .97$

		genannten Begriffe dabei verwenden.	
5	Eine Antwort auf das Problem geben	Schreibe nun einen ganzen Satz als Antwort auf die von dir notierte Frage auf.	Beurteiler entscheiden, in welchem Ausmaß die gegebene Antwort das Problem löst. $.71 \leq \tau_B \leq .85$
6	Weitere Argumente für die Richtigkeit der Lösung nennen	Findest du weitere Argumente, die deinen Lösungsvorschlag bekräftigen (z.B. aus der Physik oder deiner Erfahrung)? Antworte in ganzen Sätzen.	Beurteiler bewerten, wie gut neu aufgebrachte Ideen den durchlaufenen Problemlöseprozess stützen. $.73 \leq \tau_B \leq .82$

3 METHODE

Wie zuvor erläutert, stammen die Daten aus einer früheren Studie über die Auswirkungen von Kontexten auf die Leistung (Löffler, 2016). Die ursprüngliche Studie untersucht drei Aufgabenmerkmale: *Kontextualisiertheit* (Löffler & Kauertz, 2015 oder Löffler 2016) der in der Aufgabe beschriebenen Situation; die *Komplexität* (Kauertz, 2008), des der Aufgabe zugrundeliegenden wissenschaftlichen Modells; und *Transparenz* (Löffler & Kauertz, 2015 bzw. Löffler, 2016), die beeinflusst, ob und wie der Lernende dieses Modell identifizieren kann. Alle sind dichotom kodiert (hoch und niedrig). Das sich daraus ergebende 2x2x2-Design wurde in den 10. Klassen dreier Gymnasien in Deutschland (N = 123) wie folgt implementiert: Jeder Lernende arbeitet an einer von acht Aufgabenvarianten (t = 13 min). Die Leistung der Lernenden im Problemlösungsprozess wird dabei in den sechs Phasen des analytischen Problemlösungsprozesses in einem standardisierten Beurteilungsverfahren mit jeweils 0 bis 3 Punkten bewertet (Löffler, 2016): (1) Das Problem verstehen, (2) Die Variablen identifizieren, (3) die Auswahl der Variablen begründen, (4) einen Lösungsvorschlag formulieren, (5) eine Antwort auf das Problem geben, (6) weitere Argumente für die Gültigkeit der Lösung nennen. Jede Phase wird durch ein Item im Papier- und Bleistifttest (Löffler, 2016) im „offene Antwort“-Format abgebildet. Gründe für die eher geringe Reliabilität des Instruments ($\alpha_{\text{Cronbach}} = 0,59$) sind das komplexe Zusammenspiel der am Problemlösungsprozess beteiligten kognitiven und motivationalen Fähigkeiten, die geringe Anzahl der Items aufgrund von Testzeitbeschränkungen sowie der Ratingprozess, der die Ergebnisse zusätzlich mit Unsicherheit behaftet (Beurteiler-Reliabilität Kendalls Tau-b: $.48 \leq \tau_B \leq .97$, je nach Phase). Da die Operationalisierung der Prozessleistung jedoch eine große Herausforderung darstellt, argumentierte Schmitt (1996) an einem Beispiel mit $\alpha_c = 0,49$, dass, wenn "eine Messung andere wünschenswerte Eigenschaften aufweist, wie z.B. eine sinnvolle Inhaltsabdeckung einer Domäne und eine begründete Eindimensionalität, diese geringe Reliabilität möglicherweise kein großes Hindernis für ihre Verwendung darstellt" (aus dem Englischen übersetzt,

Schmitt, 1996, S. 351-352). Die hochkontextualisierte Version der in diesem Test verwendeten Aufgabe stammt aus dem Bereich der Thermodynamik (Kauertz, Löffler & Fischer, 2015) mit einer Aufgabenschwierigkeit von $p = 0,46$. Die niedrigkontextualisierte Version enthält das gleiche Physikproblem und hat eine Aufgabenschwierigkeit von $p = 0,47$. Insgesamt 90 Minuten lang arbeiten die Schülerinnen und Schüler an den Testheften, die zusätzlich Instrumente zur Erfassung von Kontrollvariablen enthalten: Zur Messung der kognitiven Fähigkeiten kommen Subskalen des IST-2000R (Liepmann, Beauducel, Brocke & Amthauer, 2007) zum Einsatz, konzeptuelles Vorwissen in der Thermodynamik wird mit dem Instrument Thermal Concept Evaluation (Yeo & Zadnik, 2001) erhoben. Im Unterschied zur ursprünglichen Studie wird in der vorliegenden Arbeit nicht die Leistung auf Ebene des Gesamtprozesses untersucht, sondern der Effekt der *Kontextualisiertheit* auf die Leistung in den sechs einzelnen Phasen des Problemlösungsprozesses erforscht. Zunächst wird dazu eine multivariate Kovarianzanalyse (MANCOVA) durchgeführt, da mehrere Regressionen oder Kovarianzanalysen (ANCOVA) allein zu einer Alphafehler-Inflation führen würden und jede Beziehung zwischen den abhängigen Variablen vernachlässigen würden (Tab. 6).

Tab. 6. Spearman-Korrelation der Items/abhängigen Variablen des Problemlösetests. Signifikante Ergebnisse sind hervorgehoben.

Item	1 <i>Problem verstehen</i>	2 <i>Variablen identifiz.</i>	3 <i>Auswahl begründ.</i>	4 <i>Lösungs-vorschlag</i>	5 <i>Antwort</i>	6 <i>Weitere Argumente</i>
1	r_s	1.000	-0,760	.100	.022	.038
	p	.	.405	.270	.810	.674
2	r_s		1.000	.300	.243	.183
	p		.	.001	.007	.043
3	r_s			1.000	.251	.170
	p			.	.005	.060
4	r_s				1.000	.348
	p				.	.000
5	r_s					1.000
	p					.
6	r_s					
	p					

Gleichzeitig werden die anderen, dichotomen Aufgabenmerkmale (*Komplexität* und *Transparenz*) statistisch als Kontrollvariablen behandelt, da wir in der Reanalyse nur den Effekt der *Kontextualisiertheit* untersuchen. Die Gesamtzahl der Kovariablen in einer MANCOVA wird durch die Faustregel "Maximale Anzahl der Kovariablen = (.10 x Stichprobengröße) - (Anzahl der Gruppen - 1)" begrenzt (Hair, Black, Babin, Anderson, & Tatham, 2006, S. 683). Mit den vier genannten Kontrollvariablen erfüllen wir diese Voraussetzung. Weitere Voraussetzungen für MANCOVA sind multivariate Normalität und Homogenität von Kovarianzmatrizen (Feld, 2009, S. 603). Letzteres kann vernachlässigt werden, "wenn die Gruppen ungefähr gleich groß sind (d.h. größte Gruppengröße ÷ kleinste Gruppengröße < 1,5)" (Hair et al., 2006, S. 685). Wir erfüllen diese Voraussetzung mit

57 Studenten in der Gruppe mit der niedrig kontextualisierten Aufgabe und 66 Studenten, die an der hoch kontextualisierten Aufgabe arbeiten ($66 / 57 < 1,5$). Hinsichtlich der multivariaten Normalität kommen Hair et al. (aus dem Englischen übersetzt, 2006, S. 686) zu dem Schluss, dass "Verstöße gegen diese Voraussetzung bei größeren Stichproben wenig Einfluss haben". Angesichts dessen fordern Hair et al. (2006, S. 679) als absolutes Minimum eine Stichprobengröße in Höhe der Anzahl der abhängigen Variablen und empfehlen eine Anzahl von 20 Beobachtungen pro Gruppe. Mit rund 60 Beobachtungen pro Gruppe in unserer Studie ist unsere Stichprobengröße daher ausreichend groß, um Verletzungen der univariaten Normalität zu ignorieren; insbesondere, da F-Tests im Allgemeinen robust sind (Hair et al., 2006, S. 687). Unter den verschiedenen Teststatistiken für MANCOVA ist Pillai's Trace besonders robust gegenüber Annahmeverletzungen bei gleichem Stichprobenumfang (Field, 2009, S. 605). Folglich haben wir die Pillai's Trace verwendet, um einen signifikanten Gesamteffekt von Kontextualisiertheit auf den Problemlöseprozess nachzuweisen. Basierend darauf führen wir anschließend mehrere hierarchische Regressionen mit dem gleichen Satz von Kontrollvariablen durch, um eine Vorstellung von der jeweiligen Wirkung der *Kontextualisiertheit* auf die einzelnen Phasen des Problemlöseprozesses zu bekommen (vgl. Abb. 2).

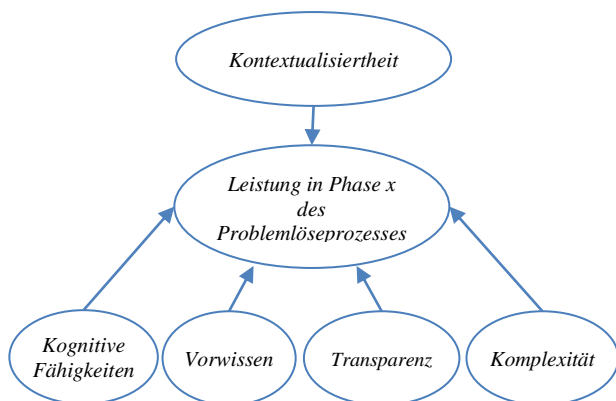


Abb. 2. Aufbau der hierarchischen Regressionen: unten Schritt 1 mit den Kontrollvariablen, oben Schritt 2 mit der unabhängigen Variable *Kontextualisiertheit*.

4 ERGEBNISSE

Mit Pillai's Trace kann ein signifikanter Gesamteffekt von *Kontextualisiertheit* auf die Leistung in den Phasen des Problemlösungsprozesses nachgewiesen werden: $V = 0.227$, $F(6, 112) = 5.483$, $p < .001$.

Tab. 7. Deskriptive Statistik der Items/abhängigen Variablen des Problemlösetests.

	Min	Max	Mittelwert	Standardabweichung	Standardfehler
Item 1	0	3	1.02	1.116	.101
Item 2	0	3	2.54	.727	.066
Item 3	0	3	1.36	.968	.087
Item 4	0	3	.68	.862	.078
Item 5	0	3	.146	.952	.086
Item 6	0	3	.122	1.156	.104

Tab. 7 zeigt die von den Lernenden erzielten Ergebnisse in den einzelnen Items des Problemlösetests. Die nach der MANOVA durchgeführten hierarchischen Regressionen zeigen signifikante Effekte von *Kontextualisiertheit* auf die Leistung für:

- Item 1/Phase 1: Das Problem verstehen: Die Kovariate haben keinen empirisch bestätigten Einfluss ($F(4, 118) = 0,947$, $p = .439$). Die Aufnahme von *Kontextualisiertheit* in das Modell (standardisiertes Beta: .208) erhöht das Bestimmtheitsmaß R^2 signifikant ($p = .023$) auf $R^2 = .034$ ($f^2 = .035$).
- Item 2/Phase 2: Die Variablen identifizieren: Die Kovariate klären 17,6 % der Varianz auf ($F(4, 118) = 7,505$, $p < .001$). Die Aufnahme von *Kontextualisiertheit* in das Modell (standardisiertes Beta: -.198) ändert R^2 signifikant ($p = .017$) auf $R^2 = .209$ ($f^2 = .042$).
- Item 3/Phase 3: Die Auswahl begründen: Die Kovariate haben keinen empirisch bestätigten Einfluss ($F(4, 118) = 2,325$, $p = .060$). Die Aufnahme von *Kontextualisiertheit* in das Modell (standardisiertes Beta: .194) erhöht R^2 signifikant ($p = .030$) auf $R^2 = .072$ ($f^2 = .078$).
- Item 4/Phase 4: Lösungsvorschlag formulieren: Die Kovariate erklären 13,4 % der Varianz ($F(4, 118) = 5,700$, $p < .001$). Die Aufnahme von *Kontextualisiertheit* in das Modell (standardisiertes Beta: -.200) ändert R^2 signifikant ($p = .018$) auf $R^2 = .167$ ($f^2 = .040$).

und nicht-signifikante Effekte für:

- Item 5/Phase 5: Eine Antwort auf das Problem geben: Weder die Kovariate noch das zweistufige Modell mit der unabhängigen Variable *Kontextualisiertheit* kann einen Einfluss auf die abhängige Variable Leistung empirisch absichern ($p = .056$ und $p = .443$).
- Item 6/Phase 6: Weitere Argumente für die Richtigkeit der Lösung nennen: Die Kovariate klären 23,2 % der Varianz auf ($F(4, 118) = 10,947$, $p < .001$). Die Aufnahme von *Kontextualisiertheit* in das Modell führt zu keiner signifikanten Änderung von R^2 ($p = .564$).

Die Zusammenfassung der Ergebnisse in Tab. 8 ermöglicht einen schnellen Vergleich der berichteten Effekte von *Kontextualisiertheit* auf die Leistung in den Phasen des Problemlöseprozesses. *Kontextualisiertheit* hat demnach einen entgegengesetzten Einfluss in den ersten vier Phasen: In Phase 1 und 3 steht eine hohe Kontextualisiertheit in Zusammenhang mit hohen Werten der entsprechenden Items des Problemlösetests. In den Phasen 2 und 4 steht eine hohe Kontextualisiertheit dagegen in Verbindung mit niedrigen Ergebnissen bei den entsprechenden Items.

Tab. 8. Zusammenfassung von Effekten von *Kontextualisiertheit* auf die Leistung in den Phasen des Problemlöseprozesses. Signifikante Werte sind hervorgehoben.

	R^2_I	R^2_{II}	ΔR^2	p	f^2	β
Phase 1						
<i>Problem verstehen</i>	-.002	.034	.034	.023	.035	.208
Phase 2						
<i>Variablen identifizieren</i>	.176	.209	.033	.017	.042	-.198
Phase 3						
<i>Auswahl begründen</i>	.042	.072	.072	.030	.078	.194
Phase 4						
<i>Lösungsvorschlag formulieren</i>	.134	.167	.033	.018	.040	-.200
Phase 5						
<i>Antwort auf das Problem geben</i>	.043	.040	/	.443	/	.069
Phase 6						
<i>Weitere Argumente nennen</i>	.232	.282	/	.564	/	-.047

5 DISKUSSION

Unsere Ergebnisse zeigen, dass die Wirkung von Kontext im Problemlösungsprozess stark variiert: So fördert er die Bemühungen der Lernenden, das Problem zu verstehen, während er sie gleichzeitig daran hindert, relevante Informationen zur Lösung des Problems auszuwählen. Geht es allerdings darum, einen Grund für die Bedeutung der ausgewählten Informationen zu liefern, scheinen die Lernenden von Kontext zu profitieren. Gleichzeitig erhöht Kontext die Anforderungen bei der Formulierung von Lösungsvorschlägen. Alle Effekte sind etwa gleich stark und heben sich daher auf der Prozessebene auf, was die Ergebnisse der ursprünglichen Studie (Löffler, 2016) erklären kann. Die Interpretation der Ergebnisse unserer Untersuchung unterliegen jedoch Einschränkungen hinsichtlich des Gültigkeitsumfangs:

Die Lernenden haben jeweils nur ein Problem aus dem Themenbereich Wärmelehre bearbeitet, Themeneffekte können daher nicht ausgeschlossen werden. Dennoch ist das Problem ein typischer Anwendungsfall für physikalische Modelle und eng sowohl an den Lehrplan als auch an Alltagserfahrungen angelehnt. Weitere Einschränkungen ergeben sich aus dem Test: Der vorstrukturierte Papier- und Bleistifttest

benachteiligt möglicherweise Novizen, da diese typischerweise von einer vermeintlichen Lösung ausgehend zurückarbeiten (Mayer, 1992). Obwohl es nicht explizit verboten ist, zwischen den Items zu wechseln, ist durch die Nummerierung eine Richtung implizit vorgegeben. Auch die durch die Implementation vorgegebene Testzeit kann Novizen benachteiligen, da diese üblicherweise mehr Zeit beim Erarbeiten einer Lösung benötigen als Experten (Larkin, McDermott, Simon & Simon, 1980a, 1980b). Daraus ergibt sich eine erhöhte Schwierigkeit für Novizen, die allerdings nicht allzu sehr ins Gewicht fällt, da die Korrelation zwischen den Items und dem Vorwissen .314 ($p < .001$) nicht übersteigt. Weiterhin wäre es wünschenswert, die Leistung in den einzelnen Phasen des Problemlöseprozesses nicht nur mit jeweils einem Item zu messen. Da die Daten aus der Reanalyse einer Studie stammen, konnte hierauf kein Einfluss genommen werden. Kovariablen sollten getrennt erhoben werden um längere Testzeiten und damit mehr Items zu ermöglichen. Demgegenüber steht eine gute ökologische Validität, da die Items die typischen Schritte im Problemlöseprozess abbilden und ein Papier- und Bleistifttest der üblichen Arbeit mit Textaufgaben entspricht.

Unter Beachtung der genannten Einschränkungen können die Ergebnisse unserer Studie dazu beitragen, die heterogenen Befunde der bisherigen Kontextforschung zu erklären: Je nachdem, wie die Leistung gemessen wird, dominieren entweder die positiven Auswirkungen von Kontext oder dessen Nachteile das Ergebnis, was zu widersprüchlichen Schlussfolgerungen führt. Zhou (2016) berichtet zum Beispiel: "Schülerinnen und Schüler schneiden im Allgemeinen mit physikalischen Kontexten besser ab als mit realen Kontexten" (aus dem Englischen übersetzt, Zhou, 2016, S. 12). Diese Schlussfolgerung ergibt sich aus ihrer Beobachtung über die Menge der zusätzlichen Variablen, die die Lernenden bei der Arbeit an einer Aufgabe berücksichtigen: Während die Schülerinnen und Schüler insgesamt fünf "unnötige" Variablen in einer Physik-Aufgabe nannten, wurden 17 zusätzliche Variablen in der kontextualisierten Version aufgeführt. Für die Kontextaufgabe schließt Zhou (2016) daraus, dass Lernende andere als die im Problem angegebenen Variablen berücksichtigen, was wiederum für die Messung der Leistung in ihrer Studie entscheidend ist. Dieser Befund passt zu unserer Beobachtung für Phase 2, in der der Kontext die Schülerinnen und Schüler bei der Identifizierung relevanter Variablen zu behindern scheint. In einer Studie zur Untersuchung der Leistung von Lernenden in kontextualisierten sowie abstrakten Problemlöseprozessen kommen Rennie und Parker (1996) zu dem Ergebnis, dass Kontext es leichter macht, das Problem zu verstehen. Sie stellen zudem fest, dass irrelevante Informationen in Kontextaufgaben (vgl. Satz <nMnL>) zwar interessant, jedoch auch verwirrend sind - in diesem Zusammenhang sprechen Harp und Mayer von „verführerischen Details“. Diese Ergebnisse passen zu unseren Beobachtungen zur Wirkung von *Kontextualisiertheit* auf Phase 1 und 2. Gleichzeitig stärkt es unsere Hypothese zu den heterogenen Ergebnissen in der Kontextforschung. Leider werden in den meisten Studien zum Einfluss von Kontext auf die Leistung im

Problemlösen keine detaillierten Angaben zur Leistungsmessung gemacht, was weitere Vergleiche erschwert: In der vielzitierten Studie von Park und Lee (2004) zu heterogenen, Kontexteffekten werden beispielsweise die Begründungen korrekter Antworten zu lebensweltlichen Problemaufgaben herangezogen um ein Leistungsmaß zu bilden. Zu den Bewertungskriterien werden allerdings keine Informationen angegeben. Taasobshirazi und Carr (S. 160, 2008) führen in einer Metastudie noch weitere Beispiele auf: „No information was provided about how the students' discourse, questionnaire responses, and written problem solutions were analyzed“; bezogen auf eine Studie von Heller und Hollabaugh (1992). Daher empfehlen wir, die Auswirkungen von Kontext auf einer detaillierteren Ebene zu untersuchen, d.h. in den einzelnen Phasen des Problemlösungsprozesses.

6 AUSBLICK

Die Ergebnisse dieser Studie führen unter anderem auch zu Fragen hinsichtlich der unmittelbaren Implikationen für die unterrichtliche Praxis. In diesem Kapitel sollen daher vielversprechende Ansatzpunkte für die weitere Forschung auf Grundlage der erhobenen Daten genannt werden:

Es ist wünschenswert, dass die negativen Auswirkungen von Kontext im Sinne einer ökonomischen Lösung für den Unterricht bereits durch Aufgabenmerkmale kompensiert werden. Konkret sollten Lernende also bei der Identifizierung relevanter Variablen (Phase 2) sowie bei der Formulierung eines Lösungsvorschlags (Phase 4) unterstützt werden. Die Ergebnisse der Regressionsanalyse legen nahe, dass das Aufgabenmerkmal *Transparenz* das erste Problem adressieren kann, indem der Zugang der Lernenden zum zugrundeliegenden wissenschaftlichen Modell erleichtert wird (Abb. 3). In Bezug auf Phase 4 scheint keines der untersuchten Aufgabenmerkmale den negativen Einfluss von Kontext zu kompensieren. Stattdessen stellt sich ein entsprechendes Vorwissen als besonders hilfreich bei der Formulierung eines Lösungsvorschlags heraus (Abb. 4).

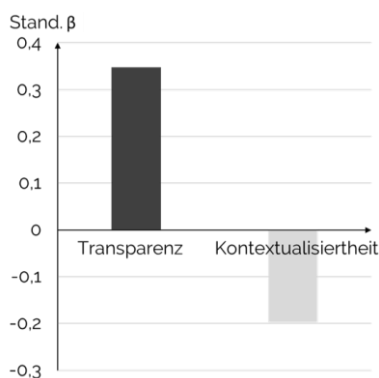


Abb. 3. Vergleich der standardisierten Betas über die Auswirkungen von *Transparenz* und *Kontextualisiertheit* auf die Leistung in Phase 2 des Problemlösungsprozesses (Identifizierung relevanter Variablen).

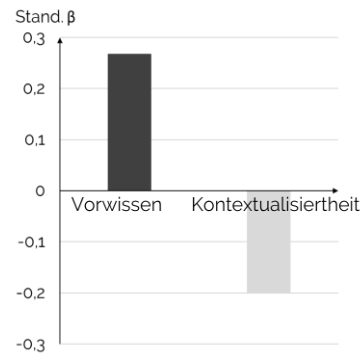


Abb. 4. Vergleich der standardisierten Betas über die Auswirkungen von *Vorwissen* und *Kontextualisiertheit* auf die Leistung in Phase 4 des Problemlösungsprozesses (Formulierung eines Lösungsvorschlags).

Daher sind weitere Untersuchungen nötig, um darüber hinaus wirksame Aufgabenmerkmale und deren Interaktion mit Persönlichkeitsmerkmalen zu identifizieren. In diesem Zusammenhang weisen aktuelle Studien auf die besondere Bedeutung affektiver Variablen sowie Metakognition hin (Löffler, Pozas & Kauertz, 2018). Der Einfluss von Kontext auf die Leistung im Problemlöseprozess ist demnach hochkomplex und künftige Untersuchungen müssen dieser Komplexität gerecht werden, um die Wirkzusammenhänge aufzuklären.

DANKSAGUNG

Die Ergebnisse dieser Arbeit sind Teil eines Projekts, das von der Deutschen Forschungsgemeinschaft gefördert wurde (GRK 1561).

LITERATURVERZEICHNIS

Bennett, J., Lubben, F., & Hogarth, S. (2007). Bringing science to life: A synthesis of the research evidence on the effects of context-based and STS approaches to science teaching. *Science Education*, 91(3), 347-370. doi:10.1002/sce.20186

Bennett, J., & Lubben, F. (2006). Context-based Chemistry: The Salters approach. *International Journal of Science Education*, 28(9), 999-1015. doi:10.1080/09500690600702496

Field, A. (2009). *Discovering statistics using spss: (and sex and drugs and rock 'n' roll)* (3rd ed.). Thousand Oaks, CA: SAGE Publications.

Finkelstein, N. (2005). Learning Physics in Context: A study of student learning about electricity and magnetism. *International Journal of Science Education*, 27(10), 1187-1209. doi:10.1080/09500690500069491

Gilbert, J. K. (2006). On the Nature of "Context" in Chemical Education. *International Journal of Science Education*, 28(9), 957-976. doi:10.1080/09500690600702470

- Gomez, M.-A., Pozo, J.-I., & Sanz, A. (1995). Students' ideas on conservation of matter: Effects of expertise and context variables. *Science Education*, 79(1), 77-93. doi:10.1002/sce.3730790106
- Hair, J. F., Black, Babin, Anderson, & Tatham. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, N.J.: Pearson Prentice Hall.
- Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, 90(3), 414-434. doi:10.1037//0022-0663.90.3.414
- Heller, P. (1992). Teaching problem solving through cooperative grouping.: Part 2: Designing problems and structuring groups. *American Journal of Physics*, 60(7), 637. doi:10.1119/1.17118
- Kauertz, A., Löffler, P., & Fischer, H. E. (2015). Physikaufgaben. In E. Kircher, R. Girwidz, & P. Häußler (Eds.), *Springer-Lehrbuch. Physikdidaktik. Theorie und Praxis* (3rd ed., pp. 451-475). Berlin: Springer.
- Kauertz, A. (2008). Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben. Berlin: Logos.
- Kauertz, Alexander; Fischer, Hans E. (2006): Assessing Students' Level of Knowledge and Analysing the Reasons for Learning Difficulties in Physics by Rasch Analysis. In: William John Boone und Xiufeng Liu (Eds.): *Applications of Rasch measurement in science education*. Maple Grove, Minnesota: JAM Press, S. 212-246.
- Kölbach, E. (2011). *Kontexteinflüsse beim Lernen mit Lösungsbeispielen*. Berlin: Logos.
- Larkin, J., McDermott, J., Simon, D., & Simon, H. (1980a). Expert and novice performance in solving physics problems. *Science*, 208(4450), 1335-1342.
- Larkin, J., McDermott, J., Simon, D. & Simon, H. A. (1980b). Models of Competence in Solving Physics Problems. *Cognitive Science*, 4(4), 317-345.
- Löffler, P., Pozas, M. & Kauertz, A. (2018). How do students coordinate context-based information and elements of their own knowledge? An analysis of students' context-based problem-solving in thermodynamics. In: *International Journal of Science Education (IJSE)*. DOI: 10.1080/09500693.2018.1514673
- Löffler, P. (2016). *Modellanwendung in Problemlöseaufgaben -- Wie wirkt Kontext? Studien zum Physik- und Chemielernen*, Vol. 205. Berlin: Logos Berlin.
- Löffler, P., & Kauertz, A. (2015). *Modellanwendung in kontextualisierten Problemlöseaufgaben*. In S. Bernholt (Ed.), *Heterogenität und Diversität - Vielfalt der Voraussetzungen im naturwissenschaftlichen Unterricht* (Vol. 35, pp. 648-650). Kiel: IPN.
- Löffler, P., & Kauertz, A. (2014). Applying physics models in context-based tasks in physics education. In: C. Constantinou, N. Papadouris, & A. Hadjigeorgiou (Hrsg.), *E-Book Proceedings of the ESERA 2013 Conference: Science Education Research For Evidence-based Teaching and Coherence in Learning - Nikosia, Zypern 2013*, (Strand 10, pp. 171-179).
- Lubben, F., Campbell, B., & Dlamini, B. (1996). Contextualizing science teaching in Swaziland: some student reactions. *International Journal of Science Education*, 18(3), 311-320. doi:10.1080/0950069960180304
- Mayer, R. (1992). *Thinking, problem solving, cognition* (2nd ed.). A series of books in psychology. New York: Freeman.
- Park, J., & Lee, L. (2004). Analysing cognitive or non-cognitive factors involved in the process of physics problem-solving in an everyday context. *International Journal of Science Education*, 26(13), 1577-1595. doi:10.1080/0950069042000230767
- Pilot, A., & Bulte, A. M. (2006). The Use of "Contexts" as a Challenge for the Chemistry Curriculum: Its successes and the need for further development and understanding. *International Journal of Science Education*, 28(9), 1087-1112. doi:10.1080/09500690600730737
- Pólya, G. (1985). *How to solve it: A new aspect of mathematical method* (2nd ed.). Princeton, N.J.: Princeton University Press.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4), 350-353. doi:10.1037/1040-3590.8.4.350
- Taasoobshirazi, G., & Carr, M. (2008). A Review and Critique of Context-Based Physics Instruction and Assessment. *Educational Research Review*, 3(2), 155-167.
- van Vorst, Helena; Dorschu, Alexandra; Fechner, Sabine; Kauertz, Alexander; Krabbe, Heiko; Sumfleth, Elke (2014): *Charakterisierung und Strukturierung von Kontexten im naturwissenschaftlichen Unterricht – Vorschlag einer theoretischen Modellierung*. In: *Zeitschrift für Didaktik der Naturwissenschaften* 21 (1), S. 29-39. DOI: 10.1007/s40573-014-0021-5.
- Whitelegg, E., & Parry, M. (1999). Real-life contexts for learning physics: Meanings, issues and practice. *Physics Education*, 34(2), 68-72. doi:10.1088/0031-9120/34/2/014

Yager, R. E. (1999). Scope, sequence and coordination: The Iowa Project, a national reform effort in the USA. *International Journal of Science Education*, 21(2), 169-194. doi:10.1080/095006999290778

Zhou, S., Han, J., Koenig, K., Raplinger, A., Pi, Y., Li, D., ... (2016). Assessment of Scientific Reasoning: The Effects of Task Context, Data, and Design on Student Reasoning in Control of Variables. *Thinking skills and creativity*, 19, 175-187. doi:10.1016/j.tsc.2015.11.004

DÉMARCHE D'INVESTIGATION DANS L'ENSEIGNEMENT DES SCIENCES DE LA NATURE À L'ÉCOLE PRIMAIRE: Effets sur la motivation et les représentations des élèves et conséquences sur la formation des enseignant(e)s

Laurent Dubois¹, Andreas Müller², Marine Deleval³

¹Section des sciences de l'éducation, Faculté de psychologie et des sciences de l'éducation, Université de Genève

²Section Physique, Faculté des Sciences, Université de Genève

³Inspé Lille Hauts de France & Univ. Lille, ULR 4072 - PSITEC - Psychologie : Interactions Temps Émotions Cognition, F-59000 Lille, France

*Please address all correspondence to Laurent Dubois, laurent.dubois@unige.ch

STRUCTURED ABSTRACT

Purpose: In the current study, we investigate the effects of new science & technology teaching materials and sequences for primary school with a focus on inquiry on aspects related to motivation (pleasure, academic self-concept/perceived competence) as well as student ideas about of science (scientific approach, usefulness of science).

Sample/Setting: Participants ($N = 180$) were from ten 5th and 6th year classes of primary schools in the canton of Geneva (international classification: ISCED 2, UNESCO 2011). They followed one among five newly designed learning sequences with 4 lessons (90 min. each) over 2 months.

Design and Methods: In a quasi-experimental pre / post design, the three dimensions “Usefulness of Science”, “Scientific approach” and “Perceived Competence in Science” were assessed by a questionnaire. Internal consistency was acceptable to good ($\alpha_c = 0.72, 0.75, 0.66$ at pre-test, respectively; similar at post-test). An ANCOVA with sequence type and measurement time as independent variables and gender as covariate was carried out. Effect sizes are reported as Cohen d .

Results: The sequences as a whole show a positive effect on the understanding of the scientific approach ($d = 0.28$), no effect on perceived competence, and a negative effect on perceived utility ($d = -0.27$). Some interesting findings result from more detailed analysis:

Girls benefit more from the learning sequences than boys: for understanding of the scientific approach and perceived science competence a difference to the disadvantage of girls was found before the learning sequences ($d = -0.38, -0.36$, respectively), but not after; understanding of the scientific approach improves for girls ($d = 0.5$), but not for boys, and perceived science competence stays stable for girls, but decreases for boys ($d = -0.37$).

Note that while most of these effects are small, they occurred after a rather short teaching time (360 min.), indicating that they are worth of interest.

Conclusions/Implications for classroom practice and future research: While there are some positive effects by the introduction of new teaching sequences like the ones studied here, there are also negative findings (e.g. for perceived utility), and positive ones are inconsistent for boys and girls, and across types of sequence. To see practices evolve and really have a lasting impact on students' motivation and ideas about science developing a few new sequences is probably not enough. A systemic change including a coherent, continued development of several sequences, an intensification of initial and continuous training, and an institutional upgrading of science education seem essential.

Keywords: Inquiry based learning, motivation, self-concept, nature of science, science beliefs
Démarche d'investigation, motivation, concept de soi, nature de la science, représentations

Received: July 2020. **Accepted:** January 2021.



1 INTRODUCTION

Cette recherche est ancrée dans un contexte de changement, tant au niveau de l'introduction de nouveaux moyens d'enseignement et de nouvelles séquences de sciences de la nature mettant en exergue la démarche d'investigation, qu'au niveau des dispositifs mis en place par les institutions scolaires pour revaloriser l'enseignement des sciences et des MITIC (Médias, images et technologies de l'information et de la communication) à l'école obligatoire (Plan MSN¹, année des sciences et des technologies 2018-2019²). Ces mesures cherchent également à amoindrir les stéréotypes de genre, étant donné que les filières scientifiques restent l'apanage des garçons.

Depuis plusieurs décennies, la didactique des sciences a développé des séquences didactiques que l'ingénierie didactique a pu mettre à l'épreuve dans le cadre de recherches spécifiques. Leur efficacité a donc été éprouvée dans des contextes bien définis auprès d'un groupe souvent restreint d'élèves.

Notre intention ici consiste à nous interroger sur la diffusion des séquences d'enseignement-apprentissage auprès des enseignants du primaire et sur l'effet de ces séquences sur les élèves. En effet, si la mise à disposition de moyens d'enseignement et de séquences d'enseignement-apprentissage de sciences de la nature est une bonne chose, il est nécessaire de s'interroger sur leurs conditions d'implémentation et les effets produits sur les élèves.

Par ailleurs, une grande partie des recherches en didactique des sciences ont mis l'accent sur l'acquisition de notions et de concepts scientifiques. Nous nous intéresserons dans le cadre de ce travail aux aspects liés à la motivation (plaisir, estime de soi) ainsi qu'aux représentations des élèves sur la science (démarche scientifique, utilité des sciences, épistémologie).

En effet, une des grandes intentions de l'enseignement des sciences à l'heure actuelle consiste non seulement à faire acquérir des connaissances scientifiques, des notions, des concepts et des modèles, mais aussi de rendre les élèves conscients des modes de production des connaissances scientifiques, des caractéristiques, des démarches et du statut des sciences (notion de « culture scientifique », OECD, 2006). Ainsi, comme le mentionne le Plan d'Études Romand – PER - (CIIP, 2010) : l'enseignement des sciences vise à permettre aux élèves

- d'acquérir un certain nombre de **notions, de concepts et de modèles scientifiques** développés progressivement par l'humanité et de **réaliser la manière dont les savoirs scientifiques se sont construits** ;

- d'identifier des questions, de développer progressivement la capacité de problématiser des situations, de mobiliser **des outils et des démarches**, de

tirer des conclusions fondées sur des faits, notamment en vue de comprendre le monde naturel et **de prendre des décisions à son propos**, ainsi que de comprendre les changements qui sont apportés par **l'activité humaine** ;

- de se montrer capable d'évaluer des faits, **de faire la distinction entre théories et observations**, et d'estimer le degré de confiance que l'on peut avoir dans les explications proposées.

On le voit, si l'acquisition de notions, de concepts et de modèles scientifiques reste bien présent, il s'agit en plus de traiter de questions d'épistémologie des sciences (réaliser la manière dont les savoirs scientifiques se sont construits), d'être capable de mener une démarche scientifique dites d'investigation (d'identifier des questions, de développer progressivement la capacité de problématiser des situations, de mobiliser des outils et des démarches, de tirer des conclusions fondées sur des faits, notamment en vue de comprendre le monde naturel et de prendre des décisions à son propos, ainsi que de comprendre les changements qui sont apportés par l'activité humaine) et de s'interroger sur les caractéristiques et le statut des sciences (se montrer capable d'évaluer des faits, de faire la distinction entre théories et observations, et d'estimer le degré de confiance que l'on peut avoir dans les explications proposées).

Cependant, si bon nombre de moyens d'enseignement et de séquences pédagogiques intègrent plus ou moins ces nouveaux objectifs d'apprentissage, peu mettent véritablement l'accent sur le développement de ces nouvelles compétences. La démarche d'investigation et la démarche scientifique sont souvent perçues plutôt comme des dispositifs pédagogiques permettant l'acquisition de connaissances que des objectifs d'apprentissage à part entière.

C'est pourquoi nous nous sommes penchés sur les effets sur l'intérêt et la motivation des élèves de 7P et 8P HARMOS (classification Suisse : HARMOS 7P/8P, EDK/CDIP 2015 ; 5^e et 6^e classification internationale : ISCED 2, UNESCO 2011) lors du suivi de différentes séquences d'enseignement-apprentissage de sciences et technologies. Dans un premier temps, le cadre théorique permettra d'explicitier les concepts de motivation, d'intérêt, de représentations et de rappeler l'importance des stéréotypes de genre en sciences de la nature. Une deuxième partie décrira le dispositif de recherche. La troisième partie présentera les résultats principaux. Nous tenterons dans la quatrième partie d'interpréter les résultats. Notre conclusion identifiera les limites de notre recherche, les pistes futures et cherchera à tirer des enseignements en matière d'implémentation de moyens d'enseignement et de séquences d'enseignement-apprentissage de sciences de la nature.

¹ Le « Plan MSN » est un plan organisé et coordonné par la Direction de l'Enseignement Obligatoire du canton de Genève, dès l'année 2016, visant à mettre en place des actions de revalorisation de l'enseignement des sciences et des mathématiques.

² L'année des « Sciences et des technologies » a été initié et coordonnée par la Direction Générale de l'Enseignement Obligatoire du canton de Genève et avait comme objectifs de stimuler ces disciplines par l'organisation d'événements ponctuels pour les classes.

2 CADRE THÉORIQUE

2.1 La motivation ou l'intérêt pour les sciences et les technologies

La motivation est considérée à juste titre comme un moteur de l'apprentissage (Viau, 2003). Or, le désir d'apprendre semble s'étioler au fil des années (Lepper et al, 2005). En cause, divers facteurs psychosociaux, familiaux, la concurrence des médias, les changements culturels, l'esprit de compétition (Giordan, 2010). Les sciences n'échappent probablement pas à cette involution.

Le concept de motivation, générique, comprend à la fois des aspects liés au métier de l'élève (envie de bien faire, de réussir) et des aspects plus intériorisés, en relation avec les intérêts de chaque élève (Viau, 2003).

C'est ce concept d'intérêt que nous avons retenu dans ce travail. Le concept d'intérêt a été largement étudié dans le cadre de différents champs de recherche en sciences humaines et sociales (psychologie, psychologie de l'éducation, sociologie, éducation scientifique et technologique, etc.). Hasni et Potvin (2015) ont documenté ce concept en se référant aux travaux de Krapp et Prenzel (2011), qui ont identifié trois éléments interdépendants : les caractéristiques du concept d'intérêt, les dimensions de l'intérêt et les niveaux d'étude de l'intérêt.

Ainsi, l'intérêt permet d'étudier les relations qu'entretiennent les élèves avec les contenus disciplinaires que sont les sciences. Il s'agit d'une relation personnelle entre un individu, l'élève, et une discipline, les sciences, construite selon plusieurs dimensions : l'intérêt pour un sujet en particulier (les animaux, les planètes), l'intérêt pour un contexte (un enjeu de société, une question d'actualité), l'intérêt pour le type d'activité (expérience, manipulation,...)³.

L'intérêt présente également trois dimensions interconnectées, la dimension cognitive (comme le sentiment de maîtrise), la dimension émotionnelle (comme le plaisir), et la dimension « valeur » (comme l'importance de la discipline pour l'individu). D'autres auteurs (Genoud & Guillod, 2014), de manière similaire, parlent de trois registres d'attitude, le registre cognitif, le registre affectif et le registre comportemental. Au niveau cognitif, Genoud et Guillod dégagent plusieurs facteurs, notamment l'utilité perçue de la discipline et le sentiment de compétence. Sur le plan affectif, registre souvent négligé dans les prédicteurs de l'engagement scolaire, le plaisir ou l'anxiété à suivre un cours disciplinaire constituent des éléments déterminants. Il s'agit donc ici de se référer au rapport que les élèves entretiennent aux disciplines ou aux activités proposées. Le registre comportemental consiste en « l'évaluation que fait l'élève de sa propre implication actuelle pour ses apprentissages » (Genoud & Guillod, 2014).

En outre, l'intérêt peut se référer à trois niveaux distincts, l'intérêt situationnel (en relation avec le type de tâches), l'intérêt général (en fonction du vécu de chacun) et l'intention d'agir (intention de s'orienter vers des études scientifiques) (Hasni & Potvin, 2015).

Notre étude se réfère à ces différents éléments, plus particulièrement :

- à l'utilité des sciences
- à l'importance des sciences
- au plaisir de faire des sciences
- au sentiment de compétence vis-à-vis des sciences
- à l'envie de s'engager dans une orientation (étude, métier) scientifique

Nous pourrions par là même comparer ainsi nos données avec celles de la recherche menée par Dutrévis et al. (2017), qui a pu déterminer que, concernant le plaisir d'apprendre en sciences chez les élèves suisses, celui-ci « ne diffère pas de manière importante du score moyen des élèves de l'OCDE et est relativement élevé, puisque globalement plus de 50% des élèves interrogés affirment aimer les sciences ». Cette recherche mentionne en outre que la motivation instrumentale (utilité perçue des sciences) obtient un score moyen plutôt inférieur à 50%.

2.2 Les représentations des sciences de la nature

2.2.1 Les représentations sociales

À l'instar de Moscovici (1961) et Jodelet (1989), et selon Minier et Gauthier (2006), « nous considérons que les représentations sociales sont constituées d'images, de croyances, d'attitudes, de valeurs, de symboles, de réseaux de concepts élaborés au fil des interactions sociales et dans un univers culturel ». Alors que Moscovici parlait de système, nous rejoignons Minier et Gauthier (2006) qui recourent à la notion de réseau, reflétant mieux l'organisation de ces divers éléments ainsi que l'aspect dynamique de la relation entre ces éléments hétéroclites, mais qui s'autoalimentent.

Les sciences, à la fois discipline, ensemble de connaissances (sur un fait, un domaine ou un objet) et méthode (mode de production de ces connaissances), revêtent diverses représentations parfois antagonistes.

2.2.2 Les représentations de l'enseignement des sciences

Pour comprendre l'état actuel des représentations de l'enseignement des sciences, il est nécessaire de faire un détour par l'histoire de l'enseignement de cette discipline. L'enseignement des sciences est le fruit d'une évolution des principes et des fondements épistémologiques, pédagogiques et didactiques, influencés par des contextes historiques (de Boer, 1991 ; Coquidé & al., 2009). En effet, les enjeux de société ont eu une incidence considérable sur les grandes orientations en matière

³ Haussler et Hoffmann, cités in Hasni et Potvin (2015)

d'enseignement des sciences. Des modes de vie, essentiellement ruraux au début du siècle dernier, aux questions environnementales qui ont jalonné la fin du XX^e siècle et le début du XXI^e siècle, en passant par les deux guerres mondiales, le développement des technologies avec la conquête spatiale, le développement des médias (télévision, ordinateurs, internet) et l'explosion des connaissances scientifiques ; ces différents contextes sont en train de transformer radicalement l'enseignement des sciences (Hulin, 2007).

Parallèlement, le début du XX^e siècle a été marqué par l'apparition des sciences de l'éducation avec la création de différents centres de recherche et d'instituts ayant pour objet les sciences de l'apprentissage. Depuis maintenant plus d'un siècle, différentes figures marquantes, comme Binet, Claparède, Ferrière, Dewey, Freinet ou Piaget, ont permis de mieux connaître les processus d'apprentissage et de développer des théories de l'apprentissage qui ont servi de bases à la didactique des sciences et de cadre théorique aux dispositions institutionnelles touchant l'enseignement des sciences (Hulin, 2001 ; Giordan, 1999).

Les quatre grands mouvements distincts

L'enseignement des sciences se décline en quatre grands mouvements distincts (Kahn, 2000) qui, dans une logique chronologique, se sont succédé dans les directives et les programmes, mais qui, dans les faits, ont coexisté et coexistent encore : la leçon de choses, la pédagogie de l'éveil, la démarche expérimentale et la démarche d'investigation (Rocard & al. 2007).

La leçon de choses

L'enseignement des sciences a grandement évolué depuis le siècle dernier. Ainsi, la leçon de choses s'est imposée comme une méthode d'enseignement des sciences jusqu'à la refonte des programmes en 1957 (Kahn, 2000). Cette méthode consistait à mettre l'élève en présence d'objets concrets (textes, schémas, maquettes), censés présenter le savoir scientifique, afin de lui faire acquérir une idée abstraite.

La pédagogie de l'éveil

La pédagogie de l'éveil, mise en place dès les années 1960, s'est appuyée sur les nouvelles théories du développement de l'enfant diffusées à cette époque (Kahn, 2000). Ainsi, la psychologie cognitive a progressivement révélé certains mécanismes de l'apprentissage, notamment l'importance de tenir compte des conceptions de l'élève et montré les limites d'une méthode basée sur l'observation, une approche positiviste, empiriste et inductive, comme l'était la leçon de choses. L'apprentissage consiste dès lors plus en une interprétation, une reconstruction ou une transformation de concepts qu'en une simple mémorisation. La pédagogie de l'éveil part donc des questionnements de l'élève, de ses conceptions et du primat de la

problématisation sur l'observation. L'accent est mis sur l'activité de l'élève et sur l'importance du tâtonnement expérimental.

La démarche expérimentale

La pédagogie de l'éveil, jugée trop générale, a été écartée au début des années 1980 avec l'apparition dans les plans d'études d'un plus grand cloisonnement disciplinaire, parallèlement à l'engagement, dans le secondaire, d'enseignant(e)s possédant des formations universitaires spécifiques (physicien(ne)s, chimistes, biologistes etc). La démarche expérimentale s'est imposée sous l'impulsion des enseignant(e)s qui voyaient une méthode permettant d'améliorer l'acquisition des connaissances scientifiques tout en rendant les élèves « actifs ». Cependant, elle a souvent été dénaturée par ceux et celles qui l'ont introduite dans les classes. Une vision figée de la démarche expérimentale se met alors en place dans les pratiques et devient un modèle à suivre pas à pas, donnant naissance à la fameuse formule OHERIC (Observation, Hypothèse, Expérience, Résultats, Interprétation, Conclusion), dénoncée par Giordan (1999). Aussi, sous prétexte de les rendre actifs, les élèves sont amenés à réaliser des expériences pour voir ou pour comprendre des concepts scientifiques, plutôt que pour véritablement mettre à l'épreuve des hypothèses.

La démarche d'investigation

Durant la première décennie du 21^{ème} siècle, de nouveaux plans d'études (CIIP, 2010 – PER⁴ et MEN, 2008⁵) ont vu le jour, en Suisse comme en France, souhaitant marquer, une fois encore, un tournant vis-à-vis d'un enseignement scientifique s'attachant presque exclusivement à faire acquérir des connaissances théoriques. En réaction aux lacunes des élèves dans le traitement des problèmes scientifiques et à une désaffection des filières scientifiques, un groupe d'experts de la Commission européenne, présidé par Michel Rocard (Rocard et al., 2007), recommande d'instaurer une approche basée sur la démarche d'investigation. Cette méthode, initiée dès la fin des années 1990 dans les pays anglo-saxons (National Research Council, 1996), met alors tout autant l'accent sur le développement de compétences que sur la construction de concepts scientifiques, ainsi que sur la motivation des élèves. Ainsi un repositionnement apparaît clairement dans les finalités de l'enseignement des sciences : l'acquisition des connaissances est relativisée en faveur d'un enseignement cherchant à développer chez les élèves des démarches, des attitudes et une culture scientifique. La démarche expérimentale se voit ainsi progressivement déstituée au profit de la démarche d'investigation, qui affiche de manière plus explicite des ambitions de développement de savoir-faire et de savoir-être.

⁴ CIIP, 2010. Plan d'Études Romand, CIIP, Conférence intercantonale de l'instruction publique de la Suisse romande et du Tessin, Neuchâtel.

⁵ Ministère de l'Éducation Nationale, 2008. Programmes de l'école primaire, Ministère de l'Éducation Nationale, Paris.

2.3 La réalité sur le terrain

Cette évolution de la didactique des sciences n'a malheureusement pas été suivie sur le terrain.

Différents travaux de recherche montrent en effet que les pratiques des enseignant(e)s en sciences de la nature sont extrêmement hétérogènes, non seulement au niveau de l'approche, mais également en termes de temps d'enseignement (Dubois, 2008). Plusieurs raisons sont évoquées par les enseignant(e)s dans cette étude : une pression politique et sociale insistant sur d'autres disciplines, le manque de maîtrise avoué par certain(e)s enseignant(e)s, les moyens d'enseignement jugés insuffisants, la grande liberté d'action concernant cette discipline, pour n'en mentionner que quatre. Nous nous retrouvons donc avec des pratiques hétérogènes, tant sur le plan de l'importance de l'enseignement des sciences attribuée par les enseignants, que par les démarches pédagogiques qu'ils mettent en place avec leurs élèves.

Nous pensons, tout comme Boivin-Delpieu et Bécu-Robinault (2015), que cette hétérogénéité témoigne d'une grande diversité des représentations épistémologiques des enseignant(e)s sur l'enseignement des sciences et sur l'implémentation de séquences de sciences. Ces auteurs affirment d'ailleurs que les gestes d'enseignement trouvent leurs fondements dans les soubassements épistémologiques de l'action professorale. En s'appuyant sur Brousseau (1998), qui définit l'épistémologie du professeur comme étant « l'ensemble de ses connaissances, de ses savoirs, de ses méthodes et de ses croyances explicites ou implicites sur la façon de trouver, d'apprendre ou d'organiser un savoir », Boivin-Delpieu et Bécu-Robinault attribuent donc une importance plus grande à la posture de l'enseignant(e) qu'à la séquence elle-même.

Selon Blanquet (2014), un(e) enseignant(e) doit avoir une notion claire de la nature de son sujet pour pouvoir enseigner efficacement les sciences. Harlen (citée dans Blanquet, 2014), l'une des pionnières de la réflexion sur l'enseignement des sciences au primaire, insiste d'ailleurs dans un rapport pour l'InterAcademy Panel – Groupe inter-académique pour les questions internationales qui réunit les académies des sciences du monde entier - (Harlen & Allende, 2009) sur « le caractère crucial d'une vision claire de la nature de la science pour enseigner les sciences ».

La nature de la science comprend à la fois les questions épistémologiques, le statut des sciences dans notre société et le rapport au savoir scientifique. Bien entendu, ces réflexions s'appliquent également aux élèves, puisqu'eux aussi se construisent, dès leur plus jeune âge, des représentations des sciences et de l'enseignement de cette discipline. Leurs représentations des sciences sont la conséquence d'une multitude d'interactions et d'expériences, que ce soit dans le cadre familial, en consultant les médias, dans leur vie sociale ou encore à l'école.

Si les représentations - appelées conceptions par certains auteurs, dont De Vecchi et Giordan (1987) - sur

différents concepts et notions scientifiques ont été largement documentés, celles touchant le rapport au savoir scientifique, les connaissances épistémologiques, les perceptions de la discipline et son statut n'ont pas été étudiées chez les élèves de l'école primaire.

Notre étude se réfère à ces différents éléments, plus particulièrement :

- les types de savoirs
- l'importance du questionnement
- le rôle des manipulations et des expériences
- la démarche scientifique

2.4 Définition et mesures des attitudes

Avant de décrire précisément l'instrument mis en œuvre pour répondre à nos questions de recherche, il nous paraît opportun de définir la notion d'attitude. L'attitude est généralement définie par « un état interne à l'individu, résultant de la combinaison de perceptions, de représentations, d'émotions, d'expériences et de l'analyse de leurs résultats. Cet état interne rend plus ou moins probable un comportement déterminé dans une situation donnée » (Raynal & Rieunier, 2001). En d'autres termes, les attitudes révèlent les représentations.

Selon Thurstone (dans Debaty, 1967, p. 11), l'attitude est la somme des sensations, idées, convictions, sentiments, relatifs à un objet déterminé. Il s'agit donc d'une variable latente, fonction de nombreuses autres variables. Pour Debaty (1967, p. 14), « l'attitude est un syndrome issu de la convergence des différents symptômes manifestes que sont les opinions verbales ou non-verbales d'un sujet à propos d'un objet défini ». Variable complexe et ambiguë, l'attitude d'un sujet se manifeste au travers de ses opinions dévoilées tant par ses expressions verbales que ses expressions non-verbales. Seule la manifestation extérieure d'une attitude qui « apparaît comme une organisation mentale dynamique des expériences sociales de l'individu » (Bayer & Tuyns, 1996, p. 48) peut être appréhendée et mesurée, à l'aide d'instruments spécifiques, selon différents paramètres :

- La direction, « positive ou négative en fonction des réactions d'attraction ou de répulsion que l'attitude met en jeu » ;
- L'intensité, « selon que ces réactions sont plus ou moins marquées » ;
- Le champ, « qui est déterminé par l'étude et la complexité de l'univers des objets que l'attitude implique ».

Parmi les outils de mesure des attitudes à disposition des sciences humaines (Khine, 2015), l'échelle de Likert, de par sa relative simplicité, figure en bonne place sur la liste des instruments les plus fréquemment employés. Par conséquent, il nous a paru évident de nous en servir dans le cadre de ce travail.

2.5 Sciences de la nature et genre

L'étude de Dutrévis et al. (2017) se réfère à différentes recherches menées ces dernières années à Genève et qui montrent que les orientations dans les filières scientifiques sont largement genrées. Ainsi, l'étude de Rastoldo et Mouad (cités dans Dutrévis et al., 2017) mentionne que déjà au cycle d'orientation, les filles sont plus nombreuses à rejoindre l'option langue, alors que les garçons optent davantage pour la filière science.

Cette différence se révèle également en ce qui concerne le secondaire II et le tertiaire, puisque là aussi les filles sont plus nombreuses à s'engager dans des formations professionnelles Santé et social et moins nombreuses que les garçons dans les formations techniques ou en option complémentaires Physique et application des maths.

Cette étude (Dutrévis et al., 2017) montre en outre que les élèves peinent à se projeter dans une carrière scientifique et surtout que le choix du premier métier chez les filles et les garçons de 8P est très différent, les garçons privilégiant nettement certains métiers scientifiques, le domaine de la santé mis à part. Tout comme nous le ferons dans notre recherche, Dutrévis et al. ont cherché à mettre en évidence des différences genrées concernant l'utilité des sciences, le sentiment de compétence et le plaisir à apprendre les sciences (affects positifs/négatifs). Des différences significatives ont ainsi été observées, notamment en ce qui concerne le plaisir à faire des sciences et le sentiment de compétence, plus élevés chez les garçons de 8P que chez les filles de 8P.

Notre étude cherchera à voir si les stéréotypes de genre apparaissent dans les discours de notre échantillon et tentera de déterminer si le genre constitue une variable significative jouant un rôle sur l'évolution de la motivation et des représentations des élèves participants à notre recherche. À partir de ces constats, nous avons identifié une question de recherche et quelques sous-questions.

2.6 Questions de recherche et hypothèses

La question de recherche faitière qui nous intéresse dans cette contribution est la suivante : Quels types de séquences de sciences ont une influence sur l'intérêt et les représentations des élèves de 5^e et 6^e année de l'école primaire (classification internationale : ISCED 2, UNESCO 2011 ; classification Suisse: HARMOS 7P/8P, EDK/CDIP 2015) ? Il semble en effet pertinent de déterminer l'influence de quelques séquences de sciences et technologies de types différents. De manière intuitive, nous pourrions penser que toutes les séquences permettent d'éveiller la curiosité scientifique des élèves, et donc de susciter un intérêt accru, de développer la motivation. Mais est-ce vraiment le cas ?

Trois sous-questions ont été également identifiées :

- Quelles sont les représentations des élèves de 10 à 12 ans sur les sciences et l'enseignement des sciences ?

- Quels registres d'attitude (intérêt, représentations des sciences) subissent une évolution significative et en fonction de quelles séquences ?
- Dans quelle mesure le genre constitue une variable significative dans l'évolution de la motivation et des représentations de notre échantillon ?

Si la motivation et les représentations des élèves en sciences de la nature sont induites par l'ensemble des activités de la classe, ainsi que par le contexte socio-culturel et institutionnel, il ne devrait pas y avoir – ou peu – d'évolution en ce qui concerne la motivation et les représentations des élèves. Les effets des quelques périodes d'enseignement cherchant à développer les compétences mentionnées ci-dessus - l'attitude scientifique, la démarche d'investigation, la motivation - se voient être atténués par l'ensemble des autres activités et circonstances, qui ne suivent pas la même logique.

Nos postulats relativisent ainsi l'impact que peuvent avoir les séquences d'enseignement-apprentissage en faveur:

1. de la posture de l'enseignante
2. du contexte socio-culturel
3. du contexte institutionnel

C'est d'ailleurs ce que sous-entendent Boivin-Delpieu et Bécu-Robinault (2015) lorsqu'elles identifient les déterminants de l'action professorale dans le contexte des démarches d'investigation. Ces chercheuses démontrent l'influence des postures épistémologiques ainsi que des instructions officielles. Nous pensons cependant que des séquences d'enseignement-apprentissage orientées et construites spécifiquement pour modifier la motivation et les représentations des élèves auront une incidence sur ces facteurs.

3 MÉTHODOLOGIE

3.1 Préambule

Cette recherche se veut être quantitative. Comme toutes les recherches en sciences humaines, nous avons conscience que le nombre de variables interagissant tout au long de notre étude est élevé. Par conséquent, les résultats de cette recherche doivent être interprétés avec prudence. Toutefois, notre hypothèse mentionnée ci-dessus, estimant justement que la motivation et les représentations des élèves en sciences de la nature sont induites par une multitude de facteurs et que donc l'évolution observée devrait être très faible, donne du sens à notre démarche.

3.2 Échantillon

Dix classes de 5^e et 6^e année de l'école primaire (classification internationale : ISCED 2, UNESCO 2011 ; classification Suisse : HARMOS 7P, 8P, EDK/CDIP 2015) du canton de Genève ont participé à cette recherche.

Le choix des 10 classes a été effectué par la DGEO⁶, selon quelques critères que nous lui avons communiqués : classes d'une même région, population de niveau socio-économique relativement homogène, ni trop favorisé, ni trop défavorisé.

Notre échantillon se compose de 180 élèves, 97 filles et 83 garçons (voir tableau 1 pour une description plus détaillée). Seuls 170 élèves ont répondu au pré-test et au post-test (suite à des valeurs manquantes le n est légèrement différent en fonction de la variable concernée). Ces données sont présentées dans le tableau 1.

Tab. 1. Échantillon

Classe	1 - Classification	2 - Classification	3 - Animascience	4 - Animascience	5 – Dans la peau de scientifiques	6 – Dans la peau de scientifiques	7 – Robotique	8 – Robotique	9 – Moyens d'enseignement habituels	10 – Moyens d'enseignement habituels	Total
Filles	3	13	12	7	8	7	12	13	9	13	97
Garçons	14	7	3	9	7	7	8	10	8	10	83
Total	17	20	15	16	15	14	20	23	17	23	180

3.3 Dispositif général

Les enseignant(e)s des 10 classes ont été contacté(e)s en octobre afin de leur communiquer les conditions de participation. Un pré-test a été effectué en novembre, les séquences ont été conduites entre décembre et janvier, un post-test a eu lieu fin février-début mars. L'ensemble du dispositif a donc duré trois mois et demi.

Chaque séquence est constituée de 4 leçons de 90 minutes, une par semaine. En fonction de la disponibilité des 10 classes, les séquences ont été menées entre le début décembre et la fin janvier. Sur les 10 classes, 2 ont suivi une séquence sur la classification animale, 2 sur la robotique et la programmation, 2 sur les activités proposées par Animascience, 2 ont suivi une séquence proposée dans différents moyens d'enseignement, par exemple Odysséo (CIIP, 2013) et 2 ont suivi des activités menées dans le cadre du projet « Dans la peau de scientifiques »⁷.

3.4 Protocole de passation

Afin de garantir que les questionnaires ont été passés dans les mêmes conditions, un protocole de passation a été rédigé et communiqué aux deux opérateurs ayant fait passer le pré-test et le post-test dans les 10 classes.

Les 10 classes de 7P/8P ont suivi le même dispositif :

1. Passation d'un pré-test soumis par un opérateur de recherche⁸.
2. Quatre leçons de sciences et technologies sur une période de 2 mois, données par le titulaire ou un médiateur d'Animascience, organisme agréé par la DGEO.
3. Passation d'un post-test soumis par un opérateur de recherche.

Les questionnaires ont été testés auprès d'une classe de 7P/8P quelques mois avant la phase d'interventions, de manière à ajuster les items.

3.5 Description des 5 séquences de sciences et technologie

Cinq types de séquences différentes ont été sélectionnés, séquences qui se distinguent notamment par le thème abordé, les objectifs poursuivis, les concepteurs, le mode d'implémentation et les personnes ayant donné la séquence. Une vue d'ensemble des 5 séquences est fournie dans le tableau 2.

⁶ Direction Générale de la Scolarité Obligatoire

⁷ Projet pédagogique « Dans la peau de scientifiques » : <https://www.unige.ch/fapse/lides/projets/dans-la-peau-de-scientifiques/>

⁸ Questionnaires et passation des tests réalisés en collaboration avec le Bchangelab (<http://bchangelab.com/>)

Tab. 2. Caractéristiques des séquences

Séquences	Thème	Objectifs	Concepteurs	Mode d'implémentation	Tenue de la classe
Classification animale	La classification animale	Découvrir la classification phylogénétique du monde animal	Groupe de travail de la DGEO (coordinatrices de discipline, enseignantes)	Distribution du matériel dans les écoles. Dossiers sur le site Internet de la DGEO.	Titulaire de la classe
Ateliers Animascience	Voir l'invisible, locomotion des animaux, articulations, corps humain	Découvrir différents thèmes, outils et démarches scientifiques	Animascience ¹ (consultants pédagogique et scientifique)	Ateliers donnés par Animascience sur inscription des enseignants.	Médiateurs scientifique d'Animascience
Dans la peau de scientifiques	Les caractéristiques de la matière et des objets	Découvrir comment mener une démarche scientifique	Initiateurs du projet (CERN ² , Université de Genève ³ , DIP ⁴ , MEN ⁵)	Projet en lien avec une formation continue de 2 demi-journées, sur inscription.	Titulaire de la classe
Robotique	La robotique et la programmation	Initier les élèves à la robotique et à la programmation	Didacticien des sciences et technologies	Projet en lien avec une formation continue, sur inscription.	Animateurs de l'Université de Genève et d'Animascience
Moyens d'enseignement habituels	Différents thèmes	Selon le sujet choisi	Auteurs des moyens d'enseignement	Manuels distribués à toutes les classes.	Titulaire de la classe

3.5.1 Séquence – Classification Animale (CA)

Cette séquence (Groupe de rédaction du DIP, 2018) fournit aux enseignants du matériel et un ensemble d'activités (guide didactique et fiches pour les élèves), permettant de travailler sur la classification phylogénétique du vivant. Le guide didactique affiche clairement deux intentions :

1. rompre avec la classification traditionnelle des vertébrés (mammifères, oiseaux, amphibiens, reptiles et poissons), qui se réfère à des critères inconstants pour construire une classification scientifique fondée sur une logique unique rendant compte des liens de parenté entre les taxons ;
2. élaborer une méthodologie de classification basée sur des critères phylogénétiques et notamment sur des critères visibles représentatifs de la phylogénie scientifique.

Distribuée au printemps 2018 à l'ensemble des enseignant(e)s, l'introduction de cette séquence n'a fait l'objet d'aucun recyclage, déléguant l'appropriation des nouveaux contenus au seul guide didactique.

3.5.2 Séquence - Ateliers Animascience (AE)

Dirigé par le Mouvement Jeunesse Suisse Romande (MJSR), en collaboration avec l'Université de Genève et le DIP, Animascience vise à mettre en place des activités scientifiques pour les enfants et les jeunes de 3 à 18 ans. Depuis 2017, différents ateliers, validés par la DGEO, sont proposés aux classes de l'école primaire du canton de Genève. Ces ateliers, animés par des médiateurs d'Animascience, cherchent avant tout à inviter les élèves

à se poser des questions, à partager des émotions, des idées et à découvrir des instruments et des méthodes scientifiques originales. Ces ateliers traitent de différents thèmes : « Voir l'invisible », « la locomotion des animaux », « les articulations », « le fonctionnement du corps humain ».

3.5.3 Séquence - Dans la peau de scientifiques (DS)

Le projet « Dans la peau de scientifiques » cherche à promouvoir la démarche d'investigation. Le dispositif allie journées de formation, activités pédagogiques à mettre en œuvre en classe, relations avec des scientifiques et avec des institutions actives dans la recherche fondamentale, accompagnement à distance par des formateurs, formatrices et des enseignant(e)s universitaires et espace numérique de travail prenant la forme d'un réseau social conçu spécialement pour le projet. Sur une durée de 3 mois, 30 classes, représentant plus de 600 élèves de la région franco-genevoise, ont comme défi d'imaginer des dispositifs leur permettant de découvrir – le plus précisément possible – ce qu'il y a à l'intérieur d'une boîte en carton, ceci sans l'endommager. Élèves et enseignant(e)s doivent, pour cela, mettre en place une véritable démarche d'investigation. Pour la recherche, 2 classes participant au projet ont été sélectionnées.

3.5.4 Séquence – Robotique (RE)

Cette séquence, développée dans le cadre d'une formation continue intitulée « Sciences et technologie par la construction et la programmation du robot LEGO WeDo 2.0 », initie les élèves à la robotique. Cette séquence a été conduite dans les classes par des animateurs de l'Université de Genève et d'Animascience.

3.5.5 Séquence - Moyens d'enseignement habituels (CE)

Cette séquence représente en fait le groupe témoin, puisqu'ici aucune consigne particulière n'a été transmise aux enseignants, leur laissant le choix du thème et de la conduite des activités. Ainsi, comme la pratique le veut, les enseignants se sont référés aux moyens d'enseignement habituels Odysséo (Rocard et al., 2007), souvent complétés par d'autres ressources et d'autres documents.

3.6 L'échelle d'évaluation

L'instrument utilisé pour mesurer les attitudes des élèves dans cette recherche est une échelle de Likert, qui se caractérise notamment par la présentation d'une liste d'items homogènes et représentatifs des opinions relatives à un objet déterminé ainsi que par une échelle d'évaluation, identique pour chaque item, à cinq catégories.

Échelle d'évaluation utilisée dans nos questionnaires :

- 1 = Pas du tout d'accord
- 2 = Plutôt pas d'accord
- 3 = Ni d'accord, ni pas d'accord
- 4 = Plutôt d'accord
- 5 = Tout à fait d'accord

Afin de varier quelque peu le mode de réponse et de maintenir l'attention des élèves, quelques questions utilisent une échelle à positionnement, d'adhésion, de fréquence, toujours sur 5 catégories.

3.7 Description du questionnaire

3.7.1 Items et dimensions

Le questionnaire utilisé comprend, selon les séquences, entre 50 et 60 items, dont l'énoncé est présenté en annexe. Trois dimensions sont en particulier identifiées (Dutrévis et al., 2017) :

- Utilité des sciences (5 items ; ex : « Ce que j'apprends en science est utile pour la vie de tous les jours. »)
- Démarche scientifique (5 items précédés de « Selon toi, à quel point te sens-tu capable d'effectuer les tâches suivantes ? ex : « Identifier ce qui pourrait influencer les résultats d'une expérience ». »)
- Compétences perçues en sciences (5 items ; ex : « Je suis doué•e en sciences. »)

Ces trois dimensions seront traitées statistiquement pour évaluer l'évolution entre les deux temps de mesure dans les cinq groupes.

Les autres items concernent la perception des sciences, les émotions liées aux sciences, le plaisir pour le cours de sciences, la volonté de devenir scientifique, les stéréotypes de genre en sciences, la définition des sciences et de ses domaines. Ces items ne seront pas utilisés dans

l'analyse statistique mais pourront être utilisés à titre d'exemples pour une analyse descriptive des représentations des élèves.

3.7.2 Test de fiabilité (« cohérence interne »)

L'alpha de Cronbach est utilisé comme mesure de fiabilité (plus précisément, cohérence interne ; Cortina, 1993). Les valeurs obtenues sont satisfaisantes au pré-test et au post-test pour les dimensions « Utilité perçue » (.72 et .75) et « Démarche scientifique » (.75 et .74). Elles restent en revanche assez basses pour la dimension « Compétence perçue » (.54 et .48) mais s'améliorent lorsque l'item « J'ai peur d'échouer en sciences » est retiré (.65 et .60). Cet item est donc définitivement retiré pour le reste de l'analyse.

3.8 Tailles d'effet

Pour évaluer l'importance des différences de scores, c'est devenu un standard méthodologique de rapporter des tailles d'effet (Wilkinson, 1999), parmi lesquelles Cohen *d* est une des plus répandue. Cette grandeur statistique est calculée comme $d = (M_1 - M_2) / SD_p$, avec

$$SD_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2}}$$

comme écart type pondéré ; Cooper et al. 2009, ch. 15. En plus, on utilise pour la discussion les seuils conventionnels (effet petit : $.2 \leq d < .5$; effet moyen : $.5 \leq d < .8$, grand effet : $.8 \leq d$; Cohen, 1988).

4 ANALYSES

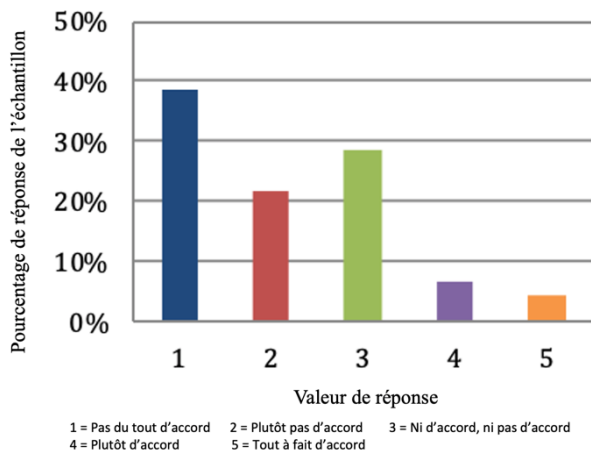
Des analyses statistiques ont été réalisées grâce au logiciel RStudio (RStudio Team, 2020).

L'objectif de ces analyses est de tester l'évolution des scores des trois sous-dimensions identifiées du questionnaire (Utilité Perçue, Démarche Scientifique, Compétence Perçue) entre le pré-test et le post-test et une différence entre les cinq groupes dans cette évolution. Pour ce faire, des analyses de covariance (ANCOVA) ont été réalisées pour tester les différences de scores entre les groupes (AE, CA, CE, DS ou RE) et entre les deux temps (pré-test ou post-test). Une interaction est également testée pour évaluer si l'évolution entre les deux temps de mesure est différente selon les groupes. Enfin, le genre est entré comme covariable dans le modèle pour contrôler son impact potentiel.

4.1 Analyse descriptive du pré-test

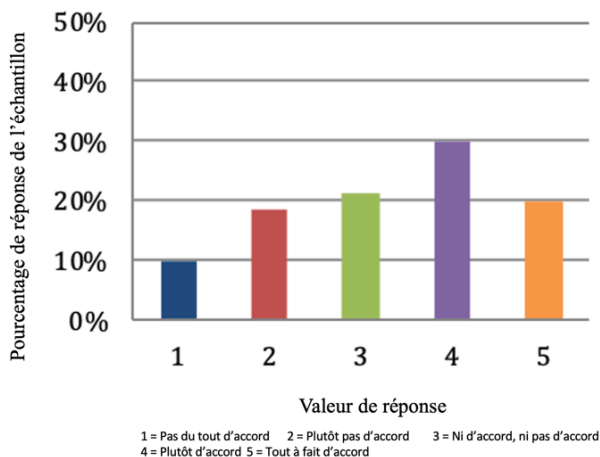
Dans un premier temps, une exploration descriptive des données issues du pré-test est réalisée afin d'avoir un aperçu des attitudes et représentations des élèves avant le début des séquences de sciences. Nous montrons ici, dans les graphiques 1, 2, 3 et 4, quelques exemples qui nous paraissent pertinents.

4.1.1 Quelques exemples représentatifs



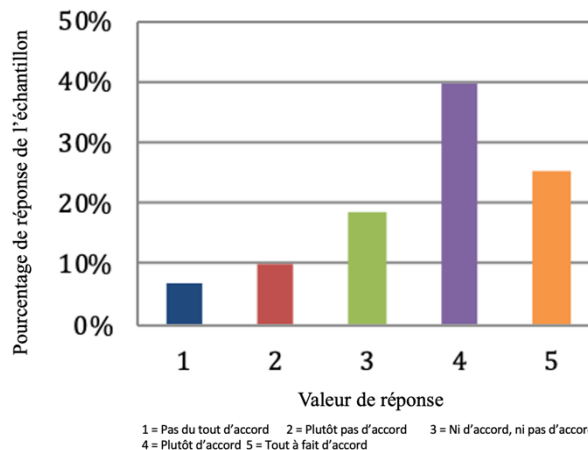
Graphique 1. Aimerais-tu devenir scientifique plus tard ?

Ici, plus de 60% des élèves interrogés sont défavorables au fait de devenir scientifique plus tard.



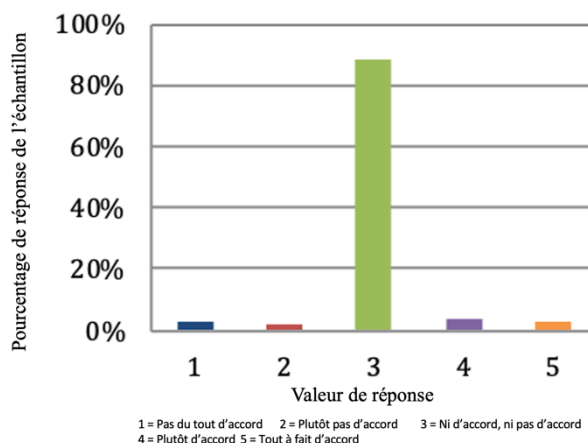
Graphique 2. Faire des sciences c'est concevoir une expérience pour tester une hypothèse

Les réponses sont assez contrastées, avec 28% en désaccord avec cette affirmation et 50% en accord.



Graphique 3. As-tu du plaisir à suivre les cours de sciences ?

Ici, près de 65% des élèves interrogés affirment avoir du plaisir à suivre les cours de sciences. Notons cependant que 16% des élèves disent le contraire.



Graphique 4. Selon toi, les sciences c'est plus facile pour les garçons ou pour les filles ?

Réponses assez unanimes pour cet item, avec 88% des élèves interrogés ayant un avis neutre, 5% affirmant que les sciences sont plus faciles pour les garçons et 7% pour les filles.

4.1.2 Les questions du genre

Concernant l'item « Selon toi, les sciences c'est plus facile pour les garçons ou pour les filles ? » (voir graphique 4), nous pouvons noter qu'il y a peu de préjugés par rapport au genre dans le discours des élèves.

Parmi les autres items, huit se démarquent quant aux différences de réponses entre les filles et les garçons, comme on peut le voir dans le tableau 3. Les tailles d'effet indiquent dans quelle mesure la différence entre les deux genres est importante. On voit que les tailles d'effet sont entre -0.4 et -0.6 (effets petits à moyens) en défaveur des filles, avec les effets les plus prononcés pour « À quel point te sens-tu capable de proposer une explication ou une solution à un problème scientifique » ($d = -0.62$) et pour la préférence des sciences par rapport au français

($d = -.6$). À noter que pour les autres items, aucune différence n'a été trouvée.

Tab. 3. Différences filles-garçons sur une sélection de huit items

Items	Pour faire des sciences, il faut être créatif	On peut apprendre toutes les connaissances scientifiques dans les livres et sur internet	À l'école, je préfère les sciences plutôt que le français	Concevoir une expérience pour tester une hypothèse*	Identifier ce qui pourrait influencer les résultats d'une expérience*	Faire les observations ou les mesures appropriées pendant une expérience*	Proposer une explication ou une solution à un problème scientifique*	Utiliser correctement un langage scientifique*
Moyenne filles	2.56	2.74	3.06	2.98	2.72	3.01	2.95	2.27
Écart-type	1.20	1.16	1.43	1.26	1.11	1.18	1.25	1.08
Moyenne Garçons	3.07	3.36	3.91	3.65	3.21	3.51	3.67	2.83
Écart-type	1.26	1.25	1.42	1.20	1.13	1.16	1.07	1.34
Différence Filles-Garçons	-.51	-.62	-.85	-.68	-.50	-.50	-.71	-.55
Écart-type pondéré	1.23	1.21	1.43	1.23	1.12	1.17	1.16	1.22
Cohen d	-.41	-.51	-.6	-.54	-.44	-.43	-.62	-.46

4.2 Analyse statistique de l'évolution des scores entre pré-test et post-test

4.2.1 Effet du groupe

Utilité perçue

Tab. 4. Moyennes et écarts-types des scores d'utilité perçue au pré-test et post-test

Groupe	Pré-test		Post-test		Diff (post-pré)
	M	SD	M	SD	
Animascience	3.13	.59	3.05	.47	-.06
Classification	3.38	.51	3.17	.59	-.21
Moyens d'enseignement habituels	3.66	.54	3.46	.70	-.20
Dans la peau de scientifiques	3.44	.56	3.36	.48	-.08
Robotique	3.42	.63	3.22	.67	-.20
TOTAL	3.42	.59	3.26	.61	-.16

Note. Score min = 1 ; score max = 5

L'analyse montre, comme on le voit dans le tableau 4, une diminution significative des scores d'utilité perçue entre le pré-test et le post-test, $F(1,160) = 16.78$, $p < .001$. La diminution correspond à une taille d'effet de $d = -.27$. Cette évolution ne diffère pas entre les groupes, $F(4,160) = .57$, *ns*. Par ailleurs, on observe

un effet principal du groupe (différence significative des scores sans distinction pré/post), $F(4,160) = 3.1$, $p = .017$. Des analyses complémentaires indiquent qu'en réalité, seuls les groupes CE et AE (soit les extrêmes si on classe les groupes selon leur score moyen) diffèrent significativement ($t = -3.36$, $p = .009$, $d = .73$).

Pour illustrer ce phénomène, nous donnons des exemples pour trois items dans le tableau 5 ci-après :

Tab. 5. Évolution des scores pour trois items d'utilité perçue

	Items	Apprendre les sciences c'est utile.	Les sciences me seront précieuses dans mon futur (formation et métier).	Ce que j'apprends en science est utile pour la vie de tous les jours.
CLASSIFICATION	PRE-TEST	4.20	3.33	3.13
	POST-TEST	3.65	3.03	2.81
	DIFFÉRENCE	-.55	-.30	-.32
	Items	Apprendre les sciences c'est utile.	Être bon.ne en sciences donne un avantage considérable pour trouver un métier.	Les sciences ne servent pas dans la vie de tous les jours.
ANIMASCIENCE	PRE-TEST	3.71	3.08	3.00
	POST-TEST	3.52	2.70	3.08
	DIFFÉRENCE	-.19	-.38	.08

Démarche scientifique

Tab. 6. Moyennes et écarts-types des scores de démarche scientifique au pré-test et post-test

Groupe	Pré-test		Post-test		Diff (post-pré)
	M	SD	M	SD	
Animascience	3.23	.69	3.67	.64	+.44
Classification	3.6	.81	3.64	.69	+.04
Moyens d'enseignement habituels	3.33	.80	3.31	.74	-.02
Dans la peau de scientifiques	2.99	.82	3.46	.73	+.47
Robotique	3.23	.79	3.46	.68	+.23
TOTAL	3.29	.80	3.50	.70	+.21

Note. Score min = 1 ; score max = 5

L'analyse montre, comme le présente le tableau 6, une augmentation significative des scores de démarche scientifique entre le pré-test et le post-test ($F(1,157) = 16.00, p < .001$), correspondant à une taille d'effet de $d = .28$. Elle montre également que cette évolution est différente en fonction du groupe ($F(4,157) = 3.33, p = .012$). Des analyses complémentaires indiquent que

l'augmentation des scores est significative pour le groupe DS ($t = 3.64, p < .001, d = .61$) et pour le groupe AE ($t = 3.08, p = .002, d = .67$). Pour les trois autres groupes, la différence entre pré et post n'est pas significative. Pour illustrer ce résultat, nous donnons des exemples pour trois items de cette sous-échelle dans le tableau 7 ci-après :

Tab. 7. Évolution des scores pour trois items de démarche scientifique

		Formuler des hypothèses.	Concevoir une expérience pour tester une hypothèse.	Identifier ce qui pourrait influencer les résultats d'une expérience.
DANS LA PEAU DE SCIENTIFIQUES	PRE-TEST	3.25	2.99	2.63
	POST-TEST	4.01	3.89	3.11
	DIFFÉRENCE	.77	.90	.48
		Formuler des hypothèses.	Concevoir une expérience pour tester une hypothèse.	Identifier ce qui pourrait influencer les résultats d'une expérience.
ANIMASCIENCE	PRE-TEST	3,27	3.04	2.70
	POST-TEST	3.81	3.71	3.36
	DIFFÉRENCE	.54	.67	.66

Compétence perçue

Tab. 8. Moyennes et écarts-types des scores de compétence perçue au pré-test et post-test

Groupe	Pré-test		Post-test		Diff (post-pré)
	M	SD	M	SD	
Animascience	3.19	.73	2.98	.56	-.21
Classification	3.15	.51	3.23	.54	+.07
Moyens d'enseignement habituels	3.17	.41	3.03	.46	-.14
Dans la peau de scientifiques	2.96	.35	2.86	.38	-.1
Robotique	3.03	.63	2.99	.6	-.04
TOTAL	3.10	.54	3.02	.53	-.08

Note. Score min = 1 ; score max = 5

Pour cette dimension, et comme le montre le tableau 8, il y a une légère diminution des scores qui est seulement tendancielle, $F(1,159) = 3.54, p = .062$. Il n'y a pas d'effet principal du groupe, ni de différence d'évolution pré-post entre les groupes.

4.2.2 Effets de genre

Utilité perçue

Pour cette dimension, il n'y a pas d'effet de genre.

Démarche scientifique

Pour cette dimension on n'obtient pas d'effet principal du genre ($F(1,157) = 2.37, ns$), mais surtout un effet différent au pré-test et au post-test ($F(1,157) = 5.92, p = .016$). Des analyses complémentaires indiquent :

- Que le score moyen du genre « fille » augmente significativement entre le pré-test et le post-test ($t = 4.75, p < .001, r = .35$), alors que celui du genre

"garçon" reste stable ($t = -1.07, ns$). Cette augmentation auprès des filles correspond à une taille d'effet $d = .50$.

- Qu'il y a une différence entre les deux genres au pré-test en défaveur des filles ($t = -2.54, p = .012, d = -.42$) mais pas au post-test ($t = -.15, ns$).

Le tableau 9 ci-dessous détaille cette différence entre les filles et les garçons.

Tab. 9. Moyennes et écarts-types des scores pour la démarche scientifique en fonction du genre

Genre	Pré-test		Post-test	
	M	SD	M	SD
Fille	3.14	.81	3.50	.62
Garçon	3.47	.77	3.51	.79

Note. Score min = 1 ; score max = 5

Compétence perçue

Il n'y a également pas d'effet principal du genre, mais par contre un effet différent du genre au pré et au post-test ($F(1,159) = 12.63, p < .001$). Des analyses complémentaires montrent :

- Que le score moyen du genre "garçon" diminue significativement entre le pré-test et le post-test ($t = 3.70, p < .001$), contrairement au genre "fille" ($t = -1.22, ns$). Cette diminution auprès des garçons correspond à une taille d'effet $d = -.37$.
- Qu'il y a une différence entre les deux genres au pré-test ($t = -2.41, p = .017, d = -.37$) mais pas au post-test ($t = .96, ns$)

Le tableau 10 ci-dessous détaille cette différence entre les filles et les garçons.

Tab. 10. Moyennes et écarts-types des scores pour les compétences perçues en fonction du genre

Genre	Pré-test		Post-test	
	M	SD	M	SD
Fille	3.01	.51	3.06	.49
Garçon	3.21	.57	3.00	.57

Note. Score min = 1 ; score max = 5

Tab. 11. Scores d'une sélection d'items concernant l'effet "classe"

Classe	1	2	3	4	5	6	7	8	9	10	Variance	Écart-type	Corrélation $r; r^2$
Devenir scientifique	2.29	1.85	1.87	1.81	1.65	2.83	3.47	2.43	2.85	1.96	.39	.59	-.45
Aime pas les science	1.35	2.25	2.20	3.38	2.06	1.61	1.80	2.29	2.35	3.04	.37	.61	.21

4.3 En résumé

Les données obtenues montrent que les séquences d'enseignement-apprentissage proposées ont un impact limité sur la motivation et les représentations des élèves en sciences de la nature.

Cependant, quelques effets ont été révélés par les données :

- une diminution globale de l'utilité perçue avec une différence entre les groupes (diminution représentative pour CE et AS), mais pas de différence d'évolution ;

4.2.3 Effet « groupe classe »

Selon nos postulats de départ, si les séquences d'enseignement-apprentissage ont un impact limité sur la motivation et les représentations des élèves en sciences de la nature, d'autres facteurs pourraient être plus déterminants, notamment les postures de l'enseignant(e) (Boivin-Delpieu et Bécu-Robinault, 2015).

L'analyse des données nous a donné l'opportunité de la vérifier en identifiant les scores moyens avec chacune des 10 classes, comme l'illustre le tableau 11 ci-dessous, grâce à deux items.

Les résultats témoignent d'un effet « classe » assez important. L'anti-corrélation est prononcée ($r = -.45$) et la taille d'effet est grande. L'effet « groupe classe » concernant l'intérêt va de pair avec l'effet négatif sur la perception d'une carrière scientifique.

D'autres items, touchant la motivation et le souhait d'orientation des élèves, mais aussi touchant la démarche scientifique, suivent cette tendance marquée.

Cette analyse révèle donc que l'un des facteurs déterminants sur la motivation et les représentations des élèves est le facteur « groupe classe ». Bien que ne faisant pas partie de nos questions de recherche, cette confirmation nous permettra d'interpréter nos résultats.

- une augmentation des capacités en démarche scientifique uniquement chez les groupes DS et AE ;
- une augmentation de la démarche scientifique pour le genre "fille" (et pas pour le genre "garçon") ;
- une diminution de la compétence perçue pour le genre "garçon" (mais pas pour le genre "fille").

Nous allons discuter ces constats dans la section suivante.

5 DISCUSSION DES RÉSULTATS

5.1 Discussion des résultats du pré-test

Le pré-test montre des représentations attendues : Pour les élèves, la science permet de découvrir de nouvelles connaissances, c'est plutôt une discipline intéressante, globalement utile et pas trop difficile.

5.1.1 Une vision stéréotypée de la demande scientifique ?

Si les élèves interrogés affirment que faire des sciences c'est acquérir de nouvelles connaissances, ils semblent avoir conscience que les sciences sont plus complexes que cela, et qu'il s'agit aussi d'apprendre à poser des questions et réfléchir, à effectuer des manipulations et des expériences, à formuler des hypothèses, à faire des observations et des mesures, et à proposer une explication ou une solution à un problème scientifique (tous ces items ont une adhésion supérieure à 70%).

Ainsi, les enjeux de l'enseignement des sciences semblent en apparence être bien maîtrisés par les élèves de notre échantillon. Sans doute que ces objectifs apparaissent de manière explicite lors des séquences d'enseignement-apprentissage qu'ils vivent en classe. Cependant, nous pouvons nous interroger sur le fait que ces représentations correspondent à des stéréotypes des activités scientifiques proposés dans le cadre scolaire. En effet, il semble ici que les élèves restituent une vision figée de la démarche scientifique, telle qu'elle apparaît dans la fameuse formule OHERIC (Observation, Hypothèse, Expérience, Résultats, Interprétation, Conclusion), décriée depuis longtemps par les didacticien(ne)s des sciences (Giordan, 1999).

Ainsi, les élèves – comme les enseignants ? – ont peut-être intégré cette transposition didactique que représente la démarche OHERIC. Ce n'est qu'une hypothèse pour laquelle des recherches complémentaires seraient nécessaires. Mais trois items, pourtant essentiels lors d'une démarche scientifique, ne recueillent respectivement que 50 %, 31 % et 54 % d'adhésion :

- Concevoir une expérience pour tester une hypothèse.
- Identifier ce qui pourrait influencer les résultats d'une expérience. Comprendre ce qui n'a pas fonctionné dans une expérience.

5.1.2 Un intérêt important pour les sciences... ...alors que les orientations scientifiques sont délaissées

De manière générale, les sciences ont la cote. Les 2/3 des élèves interrogés trouvent les sciences intéressantes et ont du plaisir à suivre les cours de sciences. Cela ne doit pas occulter le fait que près de 20% des élèves ne trouvent pas les sciences intéressantes et affirment ne pas avoir de plaisir en sciences.

Aux questions liées à l'orientation professionnelle, seulement 11% des élèves interrogés souhaitent devenir scientifique plus tard, ce qui correspond aux résultats des études mentionnées dans notre cadre théorique (Dutrévis et al. 2017).

Il est dès lors légitime de s'interroger sur le fait que 2/3 des élèves affirment être intéressés par les sciences alors que seulement à peine plus de 1/10 souhaitent s'orienter vers une carrière scientifique.

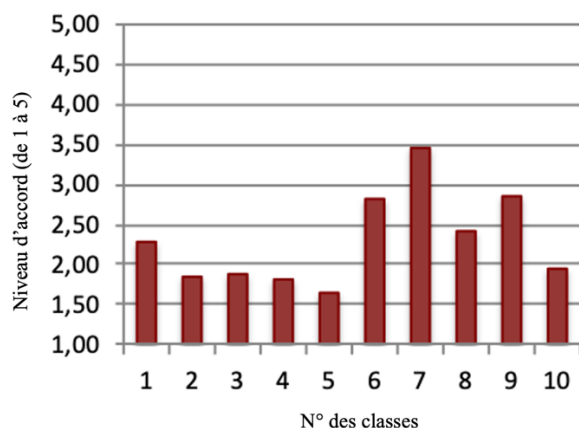
Différentes hypothèses peuvent être envisagées, notamment la méconnaissance de l'ensemble des filières scientifiques ou encore les représentations sociales concernant les métiers scientifiques. Cela dit, l'intérêt des élèves par rapport aux disciplines enseignées à l'école primaire n'est pas en rapport avec leurs projections concernant leur futur métier. Il en est certainement de même avec l'éducation physique, les arts plastiques ou les sciences humaines.

5.1.3 Les questions de genre

Comme attendu, les garçons affirment davantage aimer plutôt les sciences que le français. Par exemple, pour l'item « À l'école je préfère les sciences plutôt que le français », la différence des scores est de 0,85, le score moyen des garçons étant de 3,91 et celui des filles de 3,06. Une autre différence concerne la perception des élèves de la démarche scientifique. En effet, la sous-échelle « démarche scientifique » montre qu'initialement les filles identifient la démarche scientifique moins bien que les garçons ($d = -.38$). Finalement, la perception assez différente de la compétence perçue en défaveur des filles ($d = -.37$) est consistante avec la littérature (OCDE, 2007).

5.1.4 L'effet groupe classe

Un « effet groupe classe » assez prononcé a été révélé pour certaines sous-échelles, par exemple concernant l'envie de devenir scientifique ou encore par rapport au plaisir à suivre les cours de sciences. Voici une illustration concernant deux items, présentés grâce aux graphiques 5 et 6 ci-dessous



Graphique 5. Aimerais-tu devenir scientifique plus tard ?⁹

L'effet groupe classe assez prononcé découvert dans cette recherche nous permet de nous interroger sur les causes. Cet effet se révèle particulièrement important pour certains items, ceux concernant l'intérêt pour les sciences et le plaisir à suivre les cours de sciences et ceux concernant l'orientation professionnelle. Ainsi par exemple, pour les items ci-dessus, nous remarquons que la classe n°7, ainsi que dans une moindre mesure les classes n°6 et n°9, affichent une plus grande appétence pour devenir scientifique plus tard. Le second item témoigne du fait que la classe n°4 et la n°10 semblent moins aimer les sciences que les autres classes.

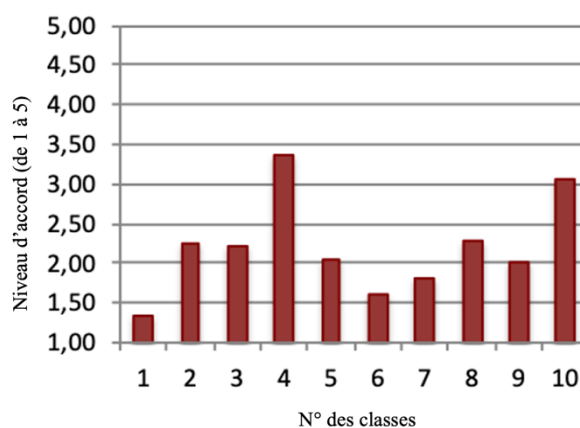
Il est difficile à ce stade de commenter ces résultats car trop de variables pourraient être susceptibles d'y jouer un rôle. Une recherche consacrée à ce phénomène pourrait permettre d'identifier les causes.

Est-ce que l'effet groupe classe est en fait un effet « enseignant », avec une affinité plus ou moins grande pour tel ou tel enseignant ? On pourrait en effet émettre l'hypothèse que certains enseignants sont moins à l'aise avec l'enseignement des sciences ou que leur style d'enseignement ne plaît pas aux élèves.

Est-ce que ce phénomène est lié aux pratiques ou au type d'enseignement des sciences de l'école. Des cultures d'école, avec des pratiques distinctes et/ou avec un intérêt (ou non) partagé au sein de la communauté scolaire pourraient expliquer cette hétérogénéité.

Est-ce qu'il n'y aurait pas un effet « enseignant complémentaire » ? Puisque dans certaines classes, l'enseignement des sciences est attribué à un enseignant complémentaire afin de délester l'enseignant titulaire, il se pourrait que celui-ci soit perçu négativement par les élèves, occasionnant problèmes de discipline ou désintérêt de la part des élèves pour la branche enseignée par ce

⁹ Ici, la moyenne de trois classes dépasse les 2,5 points. Une classe atteint même les 3,5 points.



Graphique 6. Je n'aime pas les sciences.¹⁰

complémentaire. De manière implicite en effet, les disciplines enseignées par l'enseignant titulaire, responsable de classe, sont probablement mieux considérées. Cela dit, cette hypothèse ne peut pas être vérifiée ici, d'une part, par le fait que dans l'une des deux classes ayant le plus haut score, les activités de sciences ont été données par l'enseignant titulaire et d'autre part, parce que les activités de sciences ont parfois été données par un intervenant externe (AE et RE).

Est-ce que le contexte socio-économique a pu jouer un rôle, les classes retenues pour la recherche n'étant pas situées dans des quartiers absolument identiques ? En effet, le contexte socio-économique pourrait jouer un rôle sur l'intérêt pour les disciplines scientifiques et l'orientation des élèves. Un milieu abritant un nombre important de parents d'élèves ayant une profession scientifique pourrait expliquer ces différences.

Une complémentarité entre ces facteurs n'est pas exclue, et seule une étude spécifique pourrait le déterminer.

5.2 Discussion des résultats du post-test

5.2.1 Une évolution pour les facteurs « démarche scientifique »

Au post-test, nos postulats de départ semblent se confirmer au moins en partie, puisqu'on observe peu de variation sur l'ensemble des items concernant la motivation et les représentations des élèves en sciences.

En effet, la majorité des items n'enregistrent que de faibles variations, non significatives.

Ces constats confirment que l'intérêt pour les sciences et les représentations en sciences se construisent sur du long terme et ne dépendent pas forcément directement des cours de science, mais de toute une série de facteurs liés à l'enseignement

¹⁰ Avec respectivement 3,4 points et 3,0 points, deux classes se démarquent nettement des autres. Sur le bas du graphique, une classe atteint 1,3 points.

(statut des savoirs, type de pédagogie, représentations épistémologiques des enseignants, ...) ou non (représentations socialement et culturellement partagée des sciences, environnement familial, représentations transmises par les médias,...).

Cependant, quelques séquences ont des effets sur certains paramètres, comme pour le projet « Dans la peau de scientifiques » et les activités proposées par Animascience, assurément influant sur les facteurs « démarches scientifiques », mais pas sur les autres facteurs.

Ainsi, pour déplacer quelque peu les représentations des élèves sur les facteurs « démarches scientifiques », il se pourrait qu'il y ait une nécessité impérative de travailler spécifiquement et de manière explicite ces compétences, ce que visaient les séquences « Dans la peau de scientifiques » et « Animascience ».

Cela pourrait indiquer, en tout cas pour l'enseignement primaire, dont les enseignants sont des généralistes, qu'il faille implémenter des séquences cherchant explicitement à développer la démarche scientifique pour avoir de l'effet.

Une hypothèse complémentaire consisterait à rompre avec cette idée que la démarche d'investigation vise à mieux intégrer, ou à donner du sens, aux apprentissages de notions et de concepts scientifiques. Celle-ci pourrait être utilisée alors plutôt et peut-être même exclusivement pour faire prendre conscience aux élèves la manière dont les savoirs scientifiques se sont construits, de leur faire s'approprier des démarches scientifiques, voire développer leur rapport aux savoirs. Mais il se pourrait aussi qu'il soit nécessaire que chaque enseignant(e) acquière, et ce pour toutes les disciplines, des compétences épistémologiques. L'étude de Boivin-Delpieu et Bécu-Robinault (2015) montre d'ailleurs « l'impact des représentations épistémologiques des enseignants sur l'implémentation de séquences de sciences ». Elles affirment notamment qu'il serait pertinent « de proposer aux enseignants une formation en épistémologie dans laquelle ils seraient amenés à prendre conscience non seulement de leurs propres postures épistémologiques, mais aussi du caractère déterminant de cette posture notamment sur le rôle attribué à l'expérimentation ».

5.2.2 Une diminution globale de l'utilité perçue

La diminution globale de l'utilité perçue est surprenante. La taille d'effet est petite ($d = -.27$), mais cet effet se manifeste après un enseignement assez court (360 min. au total, intégré dans une durée totale de 3.5 mois), et il est évident qu'il amène le contraire de ce qu'il est censé produire. Nous pensons que cette diminution et ce pour tous les groupes, tient au fait que les séquences sélectionnées (classification, ateliers Animascience, Dans la peau de scientifiques, robotique), ne cherchent pas vraiment à mettre l'accent sur la question de l'utilité des sciences.

5.2.3 Une variabilité selon le genre

Le fait que le score « démarche scientifique » ait progressé pour les filles d'une façon prononcée ($d = +.5$, une taille d'effet moyenne, mais obtenue par une séquence assez courte) et ait stagné pour les garçons pourrait signifier que les activités proposées dans le cadre de cette recherche permettent de combler des faiblesses pour cette sous-échelle.

La diminution de la compétence perçue pour le genre « garçon » est plus surprenant. En effet, si le score reste stable pour les filles, il diminue significativement pour les garçons, comme si les activités avaient influencé négativement la compétence perçue des garçons. Encore une fois, il s'agit d'un effet petit ($d = -.37$), mais apparaît après une durée d'enseignement assez courte. Il pourrait s'agir d'une contre-réaction au fait que ces activités mettent en avant les capacités de tous les élèves, les garçons comme les filles. Sur ce même sujet, l'activité robotique a joué un rôle positif sur les stéréotypes de genre, puisqu'elle a permis d'infléchir la croyance que les garçons ont plus de facilité que les filles dans ce domaine technologique.

5.2.4 Les objectifs de notions et de concepts sont atteints

Les items de connaissances spécifiques pour chaque séquence ont également enregistré des évolutions significatives. Ainsi, par exemple, les élèves ayant suivi les activités de robotique affirment être mieux capable de programmer un robot et comprendre un langage de programmation, ceux du projet « Dans la peau de scientifiques » se sentent plus à l'aise pour réaliser un graphique en fonction de l'évolution de la masse d'un objet, ceux ayant suivi la séquence sur la classification disent arriver mieux à différencier les arachnides des insectes et ceux ayant suivi les activités proposées par Animascience se sentent plus à même de nommer des méthodes qui permettent de voir l'invisible.

La séquence « Classification » par contre est décevante concernant les acquisitions visées. Certes, au post-test les élèves arrivent mieux à nommer un amphibien, mais lorsqu'on leur demande des catégories d'animaux qui possèdent une colonne vertébrale, beaucoup nomment les insectes, les reptiles (qui ne représente pas une catégorie dans la classification phylogénétique), les arachnides. Apparemment la séquence n'a pas réussi à modifier les fausses conceptions à ce sujet.

5.3 Limites de cette étude

Cette recherche comporte certaines limites qu'il s'agit d'évoquer ici.

Comme les 10 classes retenues dans cette recherche ont été choisies selon des critères d'homogénéité, celles-ci ne peuvent pas être totalement représentatives de la réalité du canton.

Par ailleurs, le fait d'avoir choisi un questionnaire comme outil de recueil de données peut entraîner un

biais de désirabilité sociale, les élèves cherchant peut-être à répondre en fonction de normes sociales en vigueur.

Enfin, seuls quelques items utilisés avaient déjà fait l'objet de validation. La qualité psychométrique n'est donc pas optimisée.

Il faut donc considérer les résultats de cette recherche avec la prudence qui s'impose. Néanmoins, nous pouvons tirer des conclusions qui fourniront certainement des informations utiles dans le cadre de la mise en place de mesures visant à revaloriser l'enseignement des sciences.

6 CONCLUSION

La réponse à notre question de recherche est globalement négative. Le type de séquence n'a en effet que peu d'incidence sur l'intérêt et les représentations des élèves.

L'intérêt (utilité des sciences, importance des sciences, plaisir de faire des sciences, sentiment de compétence, envie de s'engager dans une orientation scientifique) n'évolue pas vraiment. Cela témoigne peut-être d'une dévalorisation des sciences au niveau institutionnel, voire d'un manque de visibilité de la science et de son importance dans notre société.

Ce résultat témoigne en tout cas de l'utilité de renforcer les dispositifs de revalorisation des sciences à l'école primaire.

D'autres facteurs pourraient être probablement plus déterminants pour faire évoluer la motivation et les représentations des élèves. Nous pensons notamment à la posture de l'enseignant : ses représentations épistémologiques, ses connaissances de la nature de la science, son intérêt personnel. Cela suppose donc qu'il pourrait être nécessaire de former les enseignants sur ces différents aspects.

Estelle Blanquet (2014) le précise d'ailleurs dans sa thèse : « L'idée que *pour enseigner efficacement les sciences, un enseignant doit avoir une notion claire de la nature de son sujet* n'a rien d'original. Je l'emprunte à Wynne Harlen, l'une des pionnières de la réflexion sur l'enseignement des sciences au primaire, qui insistait dans un rapport pour l'InterAcademy Panel – Groupe inter-académique pour les questions internationales qui réunit les académies des sciences du monde entier - (Harlen & Allende, 2009) sur le caractère crucial d'une vision claire de la nature de la science pour enseigner les sciences. Plus généralement, la question de la nature de la science (*NoS*, dans le jargon de la didactique anglo-saxonne) et de son appropriation par les enseignants est devenue un enjeu considérable de la recherche en didactique des sciences. Son enseignement explicite est par ailleurs l'un des enjeux des récentes réformes des programmes américains et anglais, y compris pour l'école primaire. »

La question d'un enseignement explicite de la nature de la science et de l'épistémologie dans notre pays, tant dans les classes de l'école primaire que dans

la formation initiale et continue des enseignants semble être une piste plus qu'intéressante. Il faut mentionner également que les enseignants de l'école primaire sont des généralistes et que leurs connaissances de base en sciences ne sont pas aussi élevées que celles des enseignants du secondaire. La formation initiale et continue ne peut pas faire l'économie non plus, d'une mise à niveau dans ce domaine.

Par ailleurs, les changements attendus par l'introduction du PER apparaissent avec une distorsion, laissant entrevoir une vision stéréotypée de l'enseignement des sciences chez les élèves, conséquence probable d'un même malentendu chez les enseignants. Les notions et concepts scientifiques restent prédominants et la démarche scientifique persiste à être comprise selon la formule OHERIC.

L'implémentation de nouveaux moyens ou de nouvelles séquences ne peuvent pas à eux seuls engendrer les changements voulus par le PER.

Seules les séquences cherchant explicitement à développer la démarche scientifique obtient des résultats positifs. Ainsi, pour honorer tous les objectifs du PER il s'agirait sans doute de développer des séquences spécifiques pour ces compétences visées.

Les résultats obtenus dans le cadre de cette recherche concernant les questions de genre sont plutôt rassurants. En effet, les activités de sciences proposées semblent avoir un effet plus marqué chez les filles, leur permettant ainsi de stimuler leur motivation vis-à-vis des sciences. Un autre constat intéressant, la mise en place d'activités de technologies, notamment l'activité de robotique, permettent de lutter contre les stéréotypes de genre concernant les sciences. Il est donc essentiel de développer ces activités.

On ne pourra pas faire l'économie d'un changement important en matière de formation initiale et continue si l'on veut infléchir les tendances.

Il faut préciser que :

- les sciences sont plurielles et qu'il n'existe pas *une* démarche scientifique;
- les sciences n'ont pas encore réussi à rendre visibles, lisibles et compréhensibles leurs méthodes;
- l'enseignement des sciences a été influencé par une série de courants pédagogiques et didactiques qui ont jalonné le vingtième siècle et qui donnent une image chaotique de la discipline.

En outre, aucun nouveau moyen d'enseignement de sciences n'a été spécialement conçu à la suite de l'adoption du PER, contrairement aux moyens d'enseignement des autres disciplines (français, mathématiques, allemand, histoire, géographie). La

CIIP¹¹ ayant choisi de retenir un moyen d'enseignement français existant, 64 enquêtes pour comprendre le monde aux éditions Magnard, ouvrage largement décrié par les enseignants et par les autorités scolaires, comme le confirme la récente décision de la CIIP de l'abandonner au profit de nouveaux moyens d'enseignement qui seront élaborés ces prochaines années.

Se côtoient ainsi dans les classes des brochures des années 80, les moyens d'enseignement des années 2000 et les moyens « officiels » qui arborent un statut de provisoire ! Cela incite les enseignants à se procurer des ouvrages du commerce ou à télécharger des activités dénichées sur Internet, imprégnés eux aussi par 50 ans d'évolution des méthodes d'enseignement des sciences.

Ainsi, les manuels présents dans les classes, publiés à des époques différentes et rédigés par des auteurs provenant d'horizons divers, ne font qu'accroître la diversité des approches pédagogiques auxquelles les enseignants se réfèrent pour mettre en place leurs séquences d'enseignement-apprentissage en sciences. Cela pourrait constituer un atout, puisque déléguant à l'enseignant le soin d'adapter et de s'approprier ces documents, pour autant qu'il ait acquis les outils indispensables pour le faire.

Cette complexité, ce manque de clarté, ces errements et ces adjonctions se traduisent dans la pratique par une confusion générale tant au niveau des objectifs à poursuivre dans l'enseignement des sciences, qu'au niveau des approches pédagogiques recommandées.

Pour voir évoluer les pratiques et avoir une incidence sur la motivation et les représentations des élèves, le développement de séquences ne suffira pas. Un changement systémique comprenant le développement de séquences spécifiques, une intensification de la formation initiale et continue, et une revalorisation institutionnelle de l'enseignement des sciences semble indispensable.

ANNEXES

Les annexes sont disponibles sur le site de la revue.

BIBLIOGRAPHIE

Bayer, E. et Tuyns, G. 1996. *Elaboration de procédures et d'instruments de recherche*. Genève: Université de Genève.

Blanquet, E. (2014). *La Construction de critères de scientificité pour la démarche d'investigation : une approche pragmatique pour l'enseignement de la physique à l'école primaire*. [Thèse de doctorat, Université de Genève]. <https://doi.org/10.13097/archive-ouverte/unige:42783>

Boivin-Delpieu, G., & Bécu-Robinault, K. (2015). Influence des postures épistémologiques sur l'action professorale : Les phases de la Lune au cycle 3. RDST.

Recherches en didactique des sciences et des technologies, 12, 25-58. <https://doi.org/10.4000/rdst.1126>

Brousseau, G. (1998). *Théorie des situations didactiques. Didactiques des mathématiques 1970-1990*. La Pensée Sauvage.

CIIP (2010). PER – Plan d'Etudes Romand. Neuchâtel, CIIP.

CIIP (2013). *Adaptation en 2 ouvrages 5P-6P et 7P-8P de Rolando, J.-M. et al. (2013). 64 enquêtes pour comprendre le monde*. Collection Odysséo, 3^{ème} cycle – édition 2011. Magnard.

CIIP/EDK (2015). *Objectifs nationaux de formation*. Consulté le 16/6/2020 sur http://www.edudoc.ch/static/web/arbeiten/harmos/gru/ndkomp_faktenblatt_f.pdf

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (No. 300.72 C6).

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.) (2009). *The handbook of research synthesis and meta analysis*. New York: Russell Sage Foundation.

Coquidé, M., Fortin, C., & Rumelhard, G. (2009). L'investigation : Fondements et démarches, intérêts et limites. *ASTER*, 49, 51-78. <https://doi.org/10.4267/2042/31129>

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98-104. <https://doi.org/10.1037/0021-9010.78.1.98>

Debaty, P. (1967). *La Mesure des attitudes : par Pol Debaty...* Presses universitaires de France.

De Boer, G.E.A (1991). *History of Ideas in Science Education: Implications for Practice*. Teachers College Press.

De Vecchi, G., & Giordan, A. (1987). *Les origines du savoir: Des conceptions des apprenants aux concepts scientifiques*. Delachaux et Niestlé.

Dubois, L. (non publié). L'enseignement des sciences à l'école primaire et les traces écrites en sciences de la nature : attitudes d'enseignants. Étude exploratoire.

Dutrévis, M., Soussi, S. A., & Genoud, S. P. A. (2017). *Les attitudes et aspirations scientifiques des filles et des garçons à Genève*. Genève: Service de la recherche en éducation (SRED).

Genoud, P. A., & Guillod, M. (2014). Développement et validation d'un questionnaire évaluant les attitudes socio-affectives en maths. *Recherches en Éducation*, 20, 140-156.

Giordan, A. (1999). *Une didactique pour les sciences expérimentales*. Belin.

Giordan, A. (2010). *Le désir d'apprendre : un oubli de l'école...* École changer de cap. <http://www.ecolechangerdecap.net/spip.php?article61>

¹¹ Conférence intercantonale de l'instruction publique

Groupe de rédaction du DIP, (2018). *Classer le vivant 7P/8P, Guide didactique et fiches pour les élèves*. DIP Genève.

Harlen, W., & Allende, J. E. (2009). Teacher professional development in pre-secondary school inquiry-based science education (IBSE). *GraficAndes*, Santiago Chile.
<http://www.interacademies.net/12238/2951/12250/9348.aspx>

Hasni, A., & Potvin, P. (2015). *L'intérêt pour les sciences et la technologie à l'école: résultats d'une enquête auprès d'élèves du primaire et du secondaire au Québec*. Chaire de recherche sur l'intérêt des jeunes à l'égard des sciences et de la technologie.

Hulin, N. (2007). *L'Enseignement Secondaire Scientifique En France d'un Siècle à l'Autre : 1802-1980 : Évolution, Permanences et Décalages*. INRP.

Hulin, N. (2001). *Études Sur l'Histoire de l'Enseignement des Sciences Physiques et Naturelles*. ENS Editions.

Jodelet, D. (1989). Représentations sociales : un domaine en expansion. Dans Jodelet, D., (dir.). *Les représentations sociales* (pp. 47-78). PUF.

Kahn, P. (2000). L'enseignement des sciences, de Ferry à l'éveil. *ASTER*, 31, 9-35.
<https://doi.org/10.4267/2042/8751>

Khine, M.S. (2015). *Attitude Measurements in Science Education: Classic and Contemporary Approaches*. Information Age Publishing.

Krapp, A. et Prenzel, M. (2011). Research on Interest in Science: Theories, Methods, and Findings. *International Journal of Science Education*, 33(1), 27-50.

Lepper, M. R., Corpus, J. H., & Iyengar, S. S. (2005). Intrinsic and extrinsic motivational orientations in the classroom: Age differences and academic correlates. *Journal of educational psychology*, 97(2), 184.

Minier, P., & Gauthier, D. (2006). Représentations des activités d'enseignement-apprentissage en sciences et liens avec les stratégies pédagogiques déployées par des enseignants du primaire. *Journal International sur les Représentations sociales*, 3(1), 35-46.

Moscovici, S. (1961). *La psychanalyse, son image et son public. Étude sur la représentation sociale de la psychanalyse*. PUF.

National Research Council. 1996. *National Science Education Standards*. The National Academies Press. <https://doi.org/10.17226/4962>.

OCDE. (2006). *Compétences en sciences, lecture et mathématiques. Le Cadre d'évaluation de PISA 2006*.

OECD. (2007). *Pisa 2006 : Data* (Vol. 2). OECD Publishing.

Raynal, F., & Rieunier, A. (2001). *Pédagogie : Dictionnaire des concepts clés* (3ème éd.). ESF.

Rocard, M., Csermely, P., Jorde, D., Lenzen, D., Walberg-Henriksson, H., & Hemmo, V. (2007). *L'enseignement scientifique aujourd'hui : Une pédagogie renouvelée pour l'avenir de l'Europe*. Commission Européenne.
http://ec.europa.eu/research/science-society/document_library/pdf_06/report-rocard-on-science-education_fr.pdf

RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.

UNESCO. (2011). *International Standard Classification of Education: ISCED*. <http://www.uis.unesco.org/Education/Documents/isc-ed-2011-en.pdf>

Viau, R. (2003). La motivation en contexte scolaire. De Boeck Supérieur.

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8): 594-604.

AN INSTRUMENT TO MEASURE STUDENTS' PERCEPTION OF THE AUTHENTICITY OF AN OUT-OF-SCHOOL LEARNING PLACE

Daniela Schriebl¹, Andreas Müller², Nicolas Robin¹ & Beat Henrich³

¹University of Teacher Education St.Gallen, Switzerland

²Section Physique, Faculté des Sciences, Université de Genève

³Paul Scherrer Institute, Switzerland

*Please address all correspondence to Daniela Schriebl, daniela.schriebl@phsg.ch

STRUCTURED ABSTRACT

Background: One of the big opportunities offered by out-of-school learning places at research institutions is their authenticity, as they can provide insight into authentic research and work of scientists.

Purpose: To what extent the students perceive this “authenticity of place” may however be individually different. In order to measure whether students indeed perceive a given out-of-school learning offer as an authentic learning place from their individual perspective an instrument is needed.

Sample/setting: The Paul Scherrer Institute (PSI) is a genuine research environment for natural and engineering sciences and can therefore be considered as an authentic out-of-school learning place. Students from 3 different cantons in Switzerland participated in a field trip to the PSI, including a guided tour to one of its research facilities (on renewable energies) and a hands-on workshop in its science outreach lab (iLab) related to that topic. Data about test characteristics were collected in a pilot and in a main study (n = 80, March 2018 and n = 94, May to September 2018).

All the classes were taught by the same teacher to learn about the basics of the research being done in that particular field of research at PSI. The guided tour was done by the same scientist from PSI for all classes.

Design and methods: The questionnaire consists of a 6 point Likert scale with 9 items. An item analysis was carried out, as well as a factor analysis testing for the dimensionality of the questionnaire.

Results: In terms of content the items for authenticity of place can be divided into one group with a cognitive focus and another group with an emotional focus. The item analysis of the total instrument yields good to very good characteristics (Cronbach's Alpha as estimate of internal consistency $\alpha_C = .91$, average item-test-correlation $r_{it} = .71$), similarly for the sub-tests with cognitive and emotional focus ($\alpha_C = .80$, $r_{it} = .63$ and $\alpha_C = .89$, $r_{it} = .77$)

A performed confirmatory factor analysis proved compatible with a two-factor and a one-factor model (CFI = 0.98 and 0.97, respectively). The fact that the correlation between the two factors “cognitive” and “emotional” is very high (.94) argues in favour of the one factor model (McDonald's omega as estimate of internal consistency adapted to factor analysis: $\omega = 0.92$).

Conclusions/Implications for practice and future research: The instrument presented here can be used as a one factor scale with good to very good test characteristics, if an overall measure of perceived authenticity of place is needed. The two subscales with cognitive and emotional focus could also be used separately, as their test characteristics are also satisfactory to good. Due to its short format and administration time (around 2 minutes) the instrument can be well integrated in the evaluation of out-of-school learning places.

The scale was developed specifically for a research institute and has to be adapted for other out-of-school learning places such as museums, science centres or field trips. For future research it will be interesting to include other dimensions of perceived authenticity (such as authenticity of a person, e.g. the scientist at a research institute) and to study their combined effects on educational outcomes. Work along these lines within the framework of a larger research project on out-of-school science learning is in progress.

Keywords: *authenticity, out-of-school science learning, place-based learning, context-based science education*

Received: July 2020. **Accepted:** September 2020



1 INTRODUCTION

Teachers are faced with a fundamental transformation of the spaces and opportunities for experiencing, learning and teaching scientific knowledge and practices. The meaningful use of learning spaces outside the school, in museums, science centres, industry and nature is resulting in a change in their teaching practices (Braund & Reiss, 2006; National Research Council, 2009). Research in science education highlights the positive effects of such a contextualized teaching, whether in terms of an increase of students' self-efficacy (Martin, Durksen, Williamson, Kiss & Ginns, 2016; Muthersbaugh, Kern & Charvoz, 2014; Pawek, 2009), strengthening interest in science and technology (Henriksen, Jensen & Sjaastad, 2015; Hulleman & Harackiewicz, 2009), or a better understanding of science and technology and a positive impact on attitudes (National Research Council, 2009; Place-based Education Collaborative, 2010; Rennie & Johnston, 2007; Rennie & Norman G. Lederman, 2014; Tal, 2012).

A central aspect of out-of-school learning and teaching is the sought-after authenticity of the activities, of the settings and of the social context. For example, with regard to place-based outdoor learning, authentic science opportunities are valued by teachers who report positive experience with students on the motivational, emotional and social levels (Glackin, 2016; James & Williams, 2017; Lloyd, Truong & Gray, 2018). Similar results were found within the framework of the Place-based Education Evaluation Collaborative (2010). On the basis of more than 1250 interviews and more than 2600 surveys it shows clearly that place-based education improves students' achievements and their "environmental, social and economic vitality" (Place-based Education Collaborative, 2010, p. 2). According to Rickinson et al. (2004, p. 5) affective and cognitive impacts influence each other and their reinforcement can "provide a bridge to higher order learning". However, there are indications that impacts on attitudes are stronger than the cognitive and behavioural ones (Knapp & Poff, 2001).

In the case of an out-of-school learning experience at a research institute, a genuine research environment and other aspects of authenticity can be experienced which are not (or less) possible by other offers (Braund & Reiss, 2006; National Research Council et al., 2009; Stocklmayer, Rennie & Gilbert, 2010). In the present work, we investigate effects of lower secondary visits to the Paul Scherrer Institute (PSI). It is the largest research institute for natural and engineering sciences in Switzerland, covering a wide range of well-renowned research activities („The Paul Scherrer Institute in brief“, 2019) and can be considered a very authentic place to experience scientific research. On behalf of the Swiss federal 'energy strategy 2050' the PSI is currently doing research on possible solutions to store renewable energy. This is a highly relevant topic in the Swiss science curriculum (D-EDK, 2017). As several authors argue, authenticity has to emerge within the individual (Hutchison, 2008; Petraglia, 1998). In the case of PSI the potential for an authentic experience is given, but we

cannot expect every student to perceive this authenticity in the way we intended.

Some authors claim that only an expert (e.g. the scientist working at the institute or the curator of a museum) can evaluate the authenticity of a place (Hede, Garma, Josiassen & Thyne, 2014). We have a different approach by taking into consideration both perspectives, the one from the expert and the one from the student (De Bruyckere & Kirschner, 2016; Schriebl, Robin & Müller, o. J.; Weiss & Müller, 2015). How will students evaluate the authenticity of a learning place – in our case PSI? To evaluate the students' perspective and to find out to which degree students rate the research institute an authentic place, an instrument is needed, which is the focus of the present study.

Below, we present the setting of the study and the development and properties of the newly developed instrument for authenticity of place.

2 METHODS

2.1 Setting

Five school classes of secondary level one visited the PSI for one day with the goal to learn about the research done on the ESI (Energy System Integration) platform. In the morning they attended a program held at the science outreach lab iLab of the PSI. The iLab is located in the environment of the PSI but not within the fenced area, where scientists are working on their research.

To be able to understand what research is done on the ESI platform of the PSI the students discussed energy issues with a special focus on the storage of renewable energy ("power-to-gas"), performed experiments (e.g. electrolysis of water) and learned about methanization of hydrogen, a key step in the power-to-gas process (Götz et al., 2016; Rubner, Grofe & Oetken, 2019). These basics were taught in the iLab before visiting the actual research site. After the teaching unit held at the iLab all the students were given a batch with which they entered the fenced area for the guided tour. The scientist being in charge for the iLab – a physicist who used to work in research and is now responsible for the iLab – led them to the research area of the ESI – platform.

2.2 Sample and Data Collection

A pilot study took place in March 2018 (n = 80). On the basis of its outcomes, the instrument has been adapted (selection and wording of items). The main study was carried out from May to September 2018 with n = 94 secondary level one students (48 female, 45 male, 1NA). Quantitative data were collected with pre- and post-questionnaires for authenticity of place. In the larger framework of studies on out-of-school learning offers, this has to be compatible with testing for several other variables (interest, attitudes, knowledge, curiosity, self-concept and more). Thus, a short questionnaire for authenticity of place is needed, as only strongly limited testing time is available.

2.3 Item development

Items emphasize on various aspects of the personal perception of authenticity and are systematically based on conceptualisations of authenticity of place discussed in the literature. A few items could be taken verbatim from published instruments. However, to our knowledge, there is no specific instrument for authenticity of place available in literature yet. Other items are inspired by other instruments or specific aspects of authenticity of place discusses by various authors (e.g. understanding the benefit of science institutions for society; see Tab. 1 for an overview.)

The items were developed and tested in German (see Tab. 1 for an English translation). Scale format is a 6 level Likert scale (1 = don't agree at all ...6 = completely agree).

2.4 Statistical methods

To determine which items should be used for the instrument, we conducted an item analysis with the following statistics: Cronbach's Alpha of the scale, Cronbach's Alpha if the item is eliminated, the item-test correlation r_{it} (as a measure of the reliability of individual items) and the mean and standard deviation (Bortz & Schuster, 2010; Field, Miles & Field, 2012).

In a second step we ran a confirmatory factor analysis (CFA, Bollen & Curran, 2005; Ellis & Mayer, 2019) to establish whether the construct of authenticity of place consists of one or two factors. As the measurement model of CFA is different from that of classical test theory (congeneric instead of tau equivalent), a congeneric (or "composite") reliability measure such as McDonald's Omega has to be used¹ (McDonald, 1970; McDonald, 1999). It provides a better estimate for reliability than Cronbach's Alpha and has to be used instead (Raykov, 2001).

All analyses were carried out using R Studio (RStudio Team, 2019).

¹We thank one of the reviewers for this methodological advice.

Tab. 1. Items of the authenticity of place questionnaire

	Item number (one/ two factor solution)	Item wording	Item source
Cognitive aspects	AutPI02 AutPIc01	Der Besuch am PSI hat mir Einblick in Hightech-Forschung gegeben. (The visit at PSI has given me an insight into high-tech research.)	“big science” and high tech institutions (Braund & Reiss, 2006)**
	AutPI04 AutPIc02	Der Besuch am PSI hat mir einen Eindruck davon gegeben, was die Gesellschaft von einem Grossforschungsinstitut hat. (The visit at PSI has given me an impression of how society benefits from a major research institute.)	benefit for society (Rennie, 1994; PLACES, 2012)***
	AutPI06 AutPIc03	Die Führung zur ESI-Plattform hat mir geholfen zu verstehen, was Power to Gas ist. (The tour to the ESI platform helped me understand what Power to Gas is.)	as AutPI09, but for a specific research unit (Rennie, 1994)***
	AutPI09 AutPIc04	Ich habe einen Eindruck erhalten, was am PSI gemacht wird. (I got an impression of what is done at PSI.)	(Rennie, 1994)***
Emotional aspects	AutPI01 AutPIe01	Das PSI hat mir besser gefallen als alle anderen ausserschulischen Lernorte, die ich bisher besucht habe. (I liked PSI better than any other out-of-school learning place I have visited so far.)	(Schreiner & Sjøberg, 2007)***
	AutPI03 AutPIe02	Ich würde mich gerne weiter über das PSI informieren (im Internet, Büchern etc.) (I would like to get more information about PSI (from the internet, books etc.))	wish to know more, related to research institution (Kuhn, 2010; Kuhn & Müller, 2014)***
	AutPI05 AutPIe03	Ich fand die Atmosphäre am PSI faszinierend. (The atmosphere at PSI was fascinating.)	fascination related to research institution atmosphere (Litman & Spielberger, 2003)***
	AutPI07 AutPIe04	Ich fand es beeindruckend, den Arbeitsort von Wissenschaftlern sehen zu können. (It was impressive to see the workplace of scientists.)	authenticity of workplace: (Swinbank & Lunn, 2004)**
	AutPI08 AutPIe05	Das PSI ist ein spannender Arbeitsort. (PSI is an exciting workplace.)	

* *item identical or close to an item in the given reference*

** *item based on an aspect of authenticity discussed in the given reference*

*** *item based on an item in the given reference*

3 RESULTS

3.1 Item analysis

Item analysis showed satisfactory to good indices ($n = 85$, see Tab. 1): Cronbach's $\alpha = .91$, for the overall scale, 0.80 and 0.89 for the cognitive and emotional part, respectively. The item discrimination r_{it} of AutPI09 shows the lowest and only satisfactory value of .46 and could be dropped. Yet, the content of the item provides information about an important aspect about authenticity of place (Rennie, 1994) and therefore, it was decided to keep it in the scale.

Tab. 1. Item analysis AutPI

Item	Cronbach's α if item eliminated	r_{it}	Mean (SD)
AutPI01	.91	.60	3.37 (1.50)
AutPI02	.90	.73	4.23 (1.06)
AutPI03	.90	.70	3.08 (1.23)
AutPI04	.90	.77	4.20 (1.21)
AutPI05	.89	.85	3.88 (1.18)
AutPI06	.91	.66	4.22 (1.11)
AutPI07	.90	.78	4.28 (1.39)
AutPI08	.90	.82	4.19 (1.25)
AutPI09	.92	.46	4.70 (0.86)

3.2 One factor solution

First, we used authenticity of place as a single factor scale with all nine items (McDonald's Omega = .92). Confirmatory factor analysis for this model showed the following standardised values (Bollen & Curran, 2005; Ellis & Mayer, 2019; RStudio Team, 2019): ML Chi-Square = 42 (df = 27, $p = .031$), Comparative Fit Index (CFI) = 0.97, Tucker-Lewis Index (TLI) = 0.96, Root Mean Square Error of Approximation (RMSEA) = 0.08 (90% confidence interval), Sample-size adjusted Bayesian (BIC) = 1991.47, Akaike (AIC) = 2004.29.

The factor loadings for most items are satisfactory or good with exception of item AutPI09.

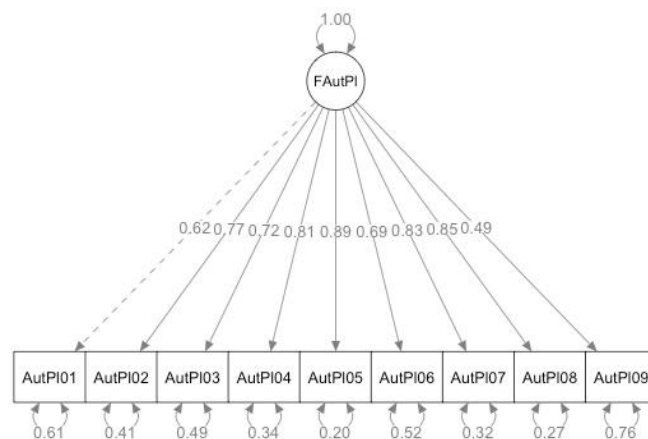


Fig. 1. Factor loadings for the one factor model

3.3 Two factor solution

Looking deeper into the content of the items, there are two different groups we can distinguish: The first focuses on whether students get insight into what is happening at the research institute and how it can help the student or society to understand the purpose of the institute and its research. These items only ask students to state whether they got an idea about these issues, but it is not asked whether or not this is a fascinating or interesting insight for the learner. So even though some students may not particularly like the place – maybe because they are not interested very much in science – they can still experience this aspect of authenticity of place. We summarize the focus of these items as “cognitive aspects” of the authenticity of place. This item group consists of four items. For the two factor solution, we renamed these items with the abbreviation “AutPlc” (with “c” for “cognitive”, see Tab. 1).

The second group of items inquires about the emotional impact which an authentic place may have on a student. They can like the place, be impressed, fascinated etc. The item group AutPle (with “e” for “emotional”) measures the emotional aspects of the authenticity of place. There are five items in this group.

Confirmatory factor analysis seems to confirm the considerations based of the content of the items. Fit indices (standardised values) are slightly better than for the one factor model: ML Chi-Square = 37 (df = 26, $p = .066$), Comparative Fit Index (CFI) = 0.98, Tucker-Lewis Index (TLI) = 0.97, Root Mean Square Error of Approximation (RMSEA) = 0.07 (90% confidence interval), Sample-size adjusted Bayesian (BIC) = 1988.12, Akaike (AIC) = 2001.65. The reliability shows a good but slightly lower value than the one factor model: McDonald's Omega = .83.

The correlation between the two factors was found to be very high ($r = 0.94$). The comparison of the two models with the Likelihood Ratio - Test (Werner, 2012) shows that the one factor model fits better with the data than the two factor model (Chi-Square-Difference = 4.64, $p = .031$).

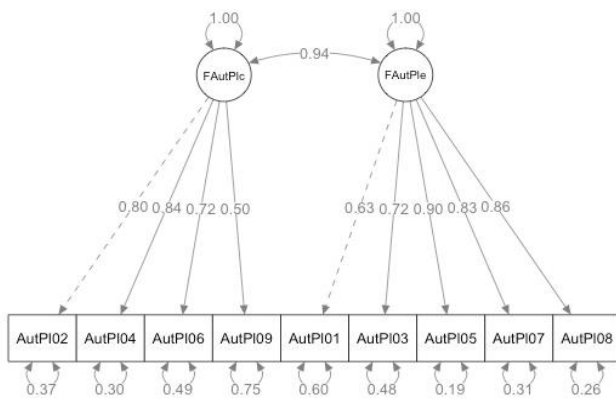


Fig. 2. Factor loadings for the two factor model

4 DISCUSSION

Allen et al. (2008) provide a list of impact categories for informal science education projects consistent with the ones found here. Their category “awareness, knowledge or understanding” can be related to the cognitive aspects of authenticity of place, whereas their category “engagement and interest” or “attitudes” can be related to the emotional aspects of authenticity of place.

Similarly, Rickinson et al. (2004) point out that there are cognitive and affective impacts of fieldwork or visits. They review several studies which looked into cognitive and affective impacts and their influence towards each other. Even though the cognitive impact often refers to knowledge or conceptual understanding in general, the same categories (“cognitive” and “affective”) can also be applied to the experience of authenticity. The cognitive aspect of authenticity at an out-of-school learning place can be understood as part of a conceptual understanding which includes that students get a deeper insight of what is done at a research institute.

Our scale with the two factors is consistent with these findings and concepts from previous research. As the correlation of the two factors is very high (.94), the instrument as presented in this article should rather be used in the form of the one factor solution. However, the two subscales with cognitive and emotional focus could also be used separately, as their test characteristics are also satisfactory to good; we thus think that this information might be useful for potential users of the test.

As for limitations, on the individual item level, AutPI09 has a smaller factor loading than the other items of this dimension. One reason for this is probably the fact that students could not get a full idea of what is done at PSI, they only saw one single research project and solely visited a small part of the whole area. So due to the setting they merely got a glimpse of what is done at PSI. However, the content of the item is important for the scale (Rennie, 1994) and it was kept as part of the factor for that reason. For AutPI01, the item discrimination r_{it} has a satisfactory value of .60, yet we would drop it in hindsight. In the case of students who have never visited an out-of-school learning place before, their answer might be distorted by the way the item is worded.

On a more general level, impact categories of authenticity discussed in the literature (Allen et al., 2008; Hede et al., 2014) are much broader than authenticity of

a place, which can be seen as only one individual aspect of a broader set of students’ attitudes or students’ understanding towards an out-of-school learning place. For future research it will be interesting to include other dimensions of perceived authenticity (such as authenticity of a person, e.g. the scientist at a research institute) and to study their combined effects on educational outcomes. Work along these lines is within the framework of a larger research project on out-of-school science learning.

Further analysis might also be needed regarding the distinction and relation between the concepts of “interest” in general (Hulleman & Harackiewicz, 2009; Krapp & Prenzel, 2011; Potvin & Hasni, 2014) and authenticity of place (emotional aspects).

Affective constructs and terms are often quite ambiguous and overlapping; an example are ‘interest’ and ‘curiosity’ (Pekrun, 2019). This holds in particular for the term ‘authenticity’, which became a kind of educational “buzzword [...], applied loosely and inconsistently to a wide range of theoretical and practical work (Shaffer & Resnick, 1999). However, a quite widespread and useful understanding of ‘authenticity’ in science education (closely related to the etymological origin: gr. authentikós “true”; lat. authenticus “reliable”) is that learning should be related to actual, real(istic), genuine contexts and experiences for learners. This point of view is also essential to and strongly advocated in the framework of scientific literacy (Fensham, 2009; OECD, 2006). We believe that an out-of-school learning opportunity is a special kind of such a context, where authenticity means a place-based experience of and insight into real research and science, its relevance, and positive emotions (interest, fascination) related to and triggered by this experience. While different motivational aspects come into play here, they are intrinsically connected to the ‘authenticity of place’, which is therefore used as a kind of overarching term.

5 CONCLUSION

The scale for authenticity of place can be a useful tool for researchers to measure authenticity of place as experienced by students with satisfactory to good psychometric properties.

The short format and administration time allow for integration in the evaluation of out-of-school learning places, where authenticity of place is only one among other variables of interest. Further items might be developed to improve the instrument if authenticity of place is the main target variable.

ACKNOWLEDGEMENTS

This research was financially supported by swissuniversities as part of the program P9 (development of disciplinary educational research) in the framework of the project CoBaLT.

REFERENCES

- Allen, S., Campbell, P. B., Dierking, L. D., Flagg, B. N., Garibay, C. & Ucko, D. A. (2008). Framework for evaluating impacts of informal science education projects. In A. J. Friedman (Hrsg.), *Report from a National Science Foundation Workshop. The National Science Foundation, Division of Research on Learning in Formal and Informal Settings*.
- Bollen & Curran. (2005). *Latent Curve Models: A Structural Equation Perspective*. Hoboken, N.J: John Wiley & Sons.
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler*. Springer-Lehrbuch (7., vollständig überarbeitete und erweiterte Auflage.). Berlin Heidelberg: Springer.
- Braund, M. & Reiss, M. (2006). Towards a More Authentic Science Curriculum: The Contribution of Out-of-School Learning. *International Journal of Science Education*, 28(12), 1373-1388.
- De Bruyckere, P. & Kirschner, P. A. (2016). Authentic teachers: Student criteria perceiving authenticity of teachers. (Y. Xian-han Huang, Hrsg.) *Cogent Education*, 3(1).
- D-EDK. (9. Mai 2017). Lehrplan 21. Abgerufen am 5. September 2017, unter www.lehrplan.ch
- Ellis, A. & Mayer, B. (30. August 2019). 10 CFA and SEM with lavaan | Introduction to R. Abgerufen am 3. Juni 2020, unter <http://methodenlehre.github.io/SGSCLM-R-course/>
- Fensham, P. J. (2009). Real world contexts in PISA science : implications for context-based science education. *Journal of Research in Science Teaching*, 46(8), 884–896.
- Field, A. P., Miles, J. & Field, Z. (2012). *Discovering statistics using R*. London ; Thousand Oaks, Calif: Sage.
- Glackin, M. (2016). ‘Risky fun’ or ‘Authentic science’? How teachers’ beliefs influence their practice during a professional development programme on outdoor learning. *International Journal of Science Education*, 38(3), 409-433.
- Götz, M., Lefebvre, J., Mörs, F., McDaniel Koch, A., Graf, F., Bajohr, S., Reimert, R., et al. (2016). Renewable Power-to-Gas: A technological and economic review. *Renewable Energy*, 85, 1371-1390.
- Hede, A.-M., Garma, R., Josiassen, A. & Thyne, M. (2014). Perceived authenticity of the visitor experience in museums: Conceptualization and initial empirical findings. *European Journal of Marketing*, 48(7/8), 1395-1412.
- Henriksen, E. K., Jensen, F. & Sjaastad, J. (2015). The Role of Out-of-School Experiences and Targeted Recruitment Efforts in Norwegian Science and Technology Students’ Educational Choice. *International Journal of Science Education, Part B*, 5(3), 203-222.
- Hulleman, C. S. & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *science*, 326(5958), 1410-1412. American Association for the Advancement of Science.
- Hutchison, P. S. (2008). *Epistemological authenticity in science classrooms*. College Park, Md. : University of Maryland. <http://hdl.handle.net/1903/8833>
- James, J. K. & Williams, T. (2017). School-Based Experiential Outdoor Education: A Neglected Necessity. *Journal of Experiential Education*, 40(1), 58-71.
- Knapp, D. & Poff, R. (2001). A Qualitative Analysis of the Immediate and Short-term Impact of an Environmental Interpretive Program. *Environmental Education Research*, 7(1), 55-65.
- Krapp, A. & Prenzel, M. (2011). Research on Interest in Science: Theories, methods, and findings. *International Journal of Science Education*, 33(1), 27-50.
- Kuhn, J. (2010). *Authentische Aufgaben im theoretischen Bereich von Instruktions-und Lehr-Lern-Forschung: Effektivität und Optimierung von Ankermedien für eine neue Aufgabenkultur im Physikunterricht*. Wiesbaden: Vieweg + Teubner - Verlag.
- Kuhn, J. & Müller, A. (2014). Context-based science education by newspaper story problems: A study on motivation and learning effects. *Perspectives in Science*, 2(1–4), 5-21. Elsevier.
- Litman, J. A. & Spielberger, C. D. (2003). Measuring epistemic curiosity and its diverse and specific components. *Journal of personality assessment*, 80(1), 75-86.
- Lloyd, A., Truong, S. & Gray, T. (2018). Place-based outdoor learning: more than a drag and drop approach. *Journal of Outdoor and Environmental Education*, 21(1), 45-60.
- Martin, A. J., Durksen, T. L., Williamson, D., Kiss, J. & Ginns, P. (2016). The role of a museum-based science education program in promoting content knowledge and science motivation: Museum-based Science Education, Learning and Motivation. *Journal of Research in Science Teaching*, 53(9), 1364-1384.
- Muthersbaugh, D., Kern, A. L. & Charvoz, R. (2014). Impact Through Images: Exploring Student Understanding of Environmental Science Through Integrated Place-Based Lessons in the Elementary Classroom. *Journal of Research in Childhood Education*, 28(3), 313-326.

- National Research Council, Bell, P., Shouse, A. W. & Feder, M. A. (2009). *Learning Science in Informal Environments: People, Places, and Pursuits*. Washington, DC: The National Academies Press.
- OECD. (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Publications de l'OCDE.
- Pawek, C. (2009). *Schülerlabore als interessesfördernde außerschulische Lernumgebungen für Schülerinnen und Schüler aus der Mittel- und Oberstufe*. Kiel: Universitätsbibliothek Kiel.
- Pekrun, R. (2019). The Murky Distinction Between Curiosity and Interest: State of the Art and Future Prospects. *Educational Psychology Review*, 31(4), 905-914.
- Petraglia, J. (1998). The real world on a short leash: The (mis) application of constructivism to the design of educational technology. *Educational Technology Research and Development*, 46(3), 53-65.
- Place-based Education Collaborative. (2010). *The Benefits of Place-based Education: A Report from the Place-based Education Evaluation Collaborative (Second Edition)*. Abgerufen am 30. April 2012, unter <http://tinyurl.com/PEECBrochure>
- Potvin, P. & Hasni, A. (2014). Interest, motivation and attitude towards science and technology at K-12 levels: a systematic review of 12 years of educational research. *Studies in Science Education*, 50(1), 85-129.
- Raykov, T. (2001). Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. *British Journal of Mathematical and Statistical Psychology*, 54(2), 315-323. Wiley Online Library.
- Rennie, L. J. D. (1994). Measuring affective outcomes from a visit to a Science Education Centre. *Research in Science Education*, 24(1), 261-269.
- Rennie, L. J. & Johnston, D. J. (2007). Visitors' Perceptions of Changes in Their Thinking about Science and Technology Following a Visit to Science Center. *Visitor Studies*, 10(2), 168-177.
- Rennie, L. & Norman G. Lederman, S. K. A. (2014). *Learning Science Outside of School*. <http://hdl.handle.net/20.500.11937/35673>
- Rickinson, Mark., Dillon, J., Teamey, K., Morris, M., Choi, M. Y., Sanders, D. & Benefield, P. (2004). *A review of research on outdoor learning*. Preston Montford: Field Studies Council.
- RStudio Team. (2019). *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, Inc. <http://www.rstudio.com/>
- Rubner, I., Grofe, T. & Oetken, M. (2019). Speicherung erneuerbarer Energien: Power-to-Gas: Energiewende für die Schulpraxis. *Chemie in unserer Zeit*, 53(2), 104-110.
- Schreiner, C. & Sjøberg, S. (2007). Science education and young people's identity construction—Two incompatible projects.
- Schriebl, D., Robin, N. & Müller, A. (o. J.). Authenticity in STEM education - a model. *in preparation*.
- de Semir, V., Revuelta, G., Dimopoulos, K., Maesele, P., & others. (2012). PLACES (Platform of Local Authorities and Communicators Engaged in Science): toolkit for the impact assessment of science communication initiatives and policies.
- Stocklmayer, S. M., Rennie, L. J. & Gilbert, J. K. (2010). The roles of the formal and informal sectors in the provision of effective science education. *Studies in Science Education*, 46(1), 1-44. Taylor & Francis.
- Swinbank, L. & Lunn, M. (2004). Learning physics and astronomy outside the classroom. *Learning science outside the classroom*, 169-183. London: Routledge Falmer.
- Tal, T. (2012). Out-of-School: Learning Experiences, Teaching and Students' Learning. In B. J. Fraser, K. Tobin, & C. J. McRobbie (Hrsg.), *Second International Handbook of Science Education* (1109-1122). Dordrecht: Springer Netherlands.
- The Paul Scherrer Institute in brief. (2019). *Paul Scherrer Institut (PSI)*. Abgerufen am 7. Februar 2020, unter <https://www.psi.ch/en/about/psi-in-brief>
- Weiss, L. & Müller, A. (2015). The notion of authenticity in the PISA units in physical science: an empirical analysis. *Zeitschrift für Didaktik der Naturwissenschaften*, 21(1), 87-97.
- Werner, C. (2012). *Parameter tests und Modellvergleiche in Strukturgleichungsmodellen*. Universität Zürich. Abgerufen am 10. Juni 2020, unter https://www.psychologie.uzh.ch/dam/jcr:fffff-f-b371-2797-0000-00007f3165a2/parameter_tests_modellvergleiche.pdf

