

Handeln und Lernen an Experimentierstationen

1 Einleitung

Nicht nur in Deutschland bleiben die Schüler*innen besonders mit Blick auf naturwissenschaftlich-technisches Lernen hinter ihren Lernmöglichkeiten zurück (TIMSS 2015), weshalb im Rahmen von schulischen sowie außerschulischen Fördermaßnahmen (z.B. MINT-Initiativen, siehe BMBF 2012) naturwissenschaftliches Lernen öffentlich und privat gefördert wird (Freericks 2011; Scharfenberg et al. 2019). Vor allem die Effekte außerschulischer Lernangebote sind jedoch eher als bescheiden oder wenigstens widersprüchlich einzuschätzen (Guderian 2007), u.a. weil die Einbindung in den Unterricht und eine inhaltlich abgestimmte Lerner*innenaktivierung nicht gelingen (Lucas 2000; Anderson et al. 2003; Waltner & Wiesner 2006; Falk 2006). Dennoch werden sie als notwendige Unterstützung angesehen, um die Qualität naturwissenschaftlichen Unterrichts zu verbessern. Aus diesem Grund ist es besonders geboten, entsprechende Projekte (im Sinne von Praxis-, Feld- bzw. Entwicklungsforschung) sorgfältig zu evaluieren.

Dies ist Gegenstand der vorliegenden Studie. Im Rahmen eines von den Stadtwerken Potsdam geförderten Projekts sollten praxistaugliche Experimentierstationen zu physikalischen Inhalten des Sachunterrichts (Wasser, Energie, Umwelt, Mobilität) für außerschulisches Lernen entwickelt und evaluiert werden.

Da die Praxisbedingungen beachtet werden sollten, ist es im Gegensatz zu Laborstudien oder auch empirisch-analytischen quasi-experimentellen Studien hier nicht möglich, über Variablenkontrolle alle wesentlichen Bedingungen der untersuchten Effekte zu kontrollieren oder die Randomisierung des Samples zu realisieren. Daher sind keine Aussagen darüber zu erwarten, ob und unter welchen sauber kontrollierten Bedingungen bestimmte Effekte (z.B. des Lernens) zu beobachten sind. Geleistet werden kann zu untersuchen, welche Effekte unter „normalen“ Feld- oder Praxisbedingungen hinsichtlich außerschulischer Lernangebote beobachtbar und in der (gegebenen) Praxis beeinflussbar sind.

2 Problem

2.1 Naturwissenschaftliches Lernen im Sachunterricht

Ein wichtiges Ziel von (unterrichtlichen und außerunterrichtlichen) Lernangeboten zur naturwissenschaftlichen Perspektive im Sachunterricht ist es, einen Beitrag zur Förderung des Lernens mit Blick auf die Entwicklung von Scientific Literacy zu leisten. Dabei bestehen besondere Lernprobleme hinsichtlich der begrifflichen Entwicklung dadurch, dass auf einer neuen gedanklichen Ebene die zunächst auf dem Hintergrund von Alltagsanforderungen konstruierten Begriffe und ihre Relationen neu gebildet, umgebildet oder parallel zu den Alltagsbegriffen neu kontextualisiert und repräsentiert werden müssen (Conceptual Change, siehe Vosniadou 2015). Psychologisch bedeutet dies, dass die begrifflich gefassten Sachverhalte mit neuen (naturwissenschaftlichen) Merkmalen (neuer Merkmalssatz) und Relationen verbunden werden und andererseits eine neue (wissenschaftliche) Bedeutung/ einen neuen Sinn erlangen müssen. Kognitionspsychologisch bedeutet das, neben der Repräsentation der Phänomene auf einer Oberflächenebene eine Repräsentation auf Tiefenebene aufzubauen. Beide

unterscheiden sich durch unterschiedliche Bezüge zum Kontext (Alltag vs. Wissenschaft/Fach), zur Anschaulichkeit (anschaulich-konkret, unanschaulich-abstrakt), zum praktisch-gegenständlichen Handeln (Hands-on vs. Minds-on-Aktivität, siehe Kaltman 2010; Marquardt-Mau & Hoffmann 2009) sowie zur Art der Begriffsbildung (empirische vs. theoretische Begriffsbildung, siehe Vygotskij 2002, Giest 2020). Hinsichtlich ihrer empirischen Erforschung wird in einschlägigen Untersuchungen (nicht nur) zum Sachunterricht von drei Niveaustufen ausgegangen, auf die Lerntests zu richten sind: 1. naive Vorstellungen, 2. Zwischenvorstellungen, 3. wissenschaftliche Vorstellungen (vgl. Vosniadou 2015; Möller et al. 2006; Hardy et al. 2010 bzw. Koerber et al. 2011, 2015; Edelsbrunner et al. 2018).

2.2 Experimentieren als Lernhandlung

Das Experimentieren ist eine wesentliche naturwissenschaftliche Lernhandlung und Erkenntnisweise bzw. -methode, von deren Beherrschung sowohl das Verstehen bzw. Erklären naturwissenschaftlicher Phänomene als auch die emotional-motivationale Haltung gegenüber den Naturwissenschaften wesentlich abhängt (GDSU 2013; Hartinger 2020; Haider 2015; Körber et al. a.a.O.).

Es gibt kaum eine naturwissenschaftliche Methode, die mit Blick auf den Unterricht so gut erforscht ist, wie das Experimentieren (vgl. Alberts & Giest 2011, 2012; Bohrmann, Todorowa & Möller 2016; Dunker 2016; Elschenbroich 2005; Giest 2016; Giest & Alberts 2010; Hartinger 2020; Kalthoff, Theyßen & Schreiber 2016; Köster, Hellmich & Nordmeier 2010; Michalik 2010; Pahl & Lück 2016; Peschel 2010, 2015; Priemer 2011; Waldenmaier, Köster & Müller 2013; Waldenmaier et al. 2015; Windt, Scheuer & Melle 2011). Dies gilt jedoch nicht im gleichen Maß für das außerschulische Lernen. Sucht man im Web (9.740.000 Ergebnisse auf die Suchanfrage Experimente für Kinder), ist das Angebot an meist wenig wissenschaftlich evaluierten Experimenten nicht zu überblicken. Von daher findet sich auch in Science Labs oder Science Centern sowie in Experimentierkästen eine Vielzahl an Experimenten, die für Kinder empfohlen werden, ohne dass geprüft wurde, unter welchen Bedingungen diese im Unterricht zu empfehlen und ob diese Bedingungen im Feld überhaupt herzustellen sind.

Aus theoretischer Sicht ist dabei unstrittig, dass beim naturwissenschaftlichen Experimentieren instruktionale Unterstützung unerlässlich ist (Ramseger 2013). Dennoch wird aus Sicht der Praxis zum einen kritisiert, dass Kinder zu häufig ausschließlich manuell aktiv sind und zum anderen, dass sie zu „Handlangern“ gemacht werden (Elschenbroich 2005) und selbständiges Handeln zu kurz kommt (Murmman et al. 2019; Zahdeh & Peschel 2018; Michalik 2010; Waldenmaier et al. 2013, 2015; Kihm & Peschel 2021). Unklar ist daher weiterhin, wie in Praxissituationen beim Experimentieren ein angemessenes Maß an instruktionaler Stützung (u.a. kognitive Strukturierung und Aktivierung der Lernenden) gesichert werden kann, ohne dass die Eigenregulation beim Lernen (Sinnbildung) zu kurz kommt.

Unklar ist ferner, welche Experimente kindliches Verstehen unter der Bedingung möglichst eigenregulierten Handelns fördern. Instruktionsbefürworter argumentieren, dass durch instruktionale Stützung Lerngrenzen der Kinder überwunden werden können: „Wissenschaftliches Arbeiten erfordert eine der kindlichen Spontaneität geradezu entgegenlaufende Haltung“ (Ramseger 2013, S. 164); „Bei einer Aufgabenbearbeitung sind Lernende je nach Lernausgangslage auf mehr oder weniger Unterstützung angewiesen [...], die den Experimentierprozess der Lernenden strukturiert“ (Stiller & Wilde 2021, o.S.); „discovery tasks, which require children to engage in unguided experimentation, typically do not reveal systematic skills in primary school“ (Osterhaus et al. 2015, S. 56). Die andere Seite betont, dass das Interesse der Kinder und die selbstgewählte Aktivität inklusive der dazu gehörenden Kinderfragen wichtiger für den Lernerfolg sind als das Erreichen von Instruktion abhängiger Lernziele: „bei geöffneten Experimentierangeboten [ist] die Engagiertheit der Kinder signifikant höher“ (Waldenmeyer 2013, S. 140) und weiter (unter Bezug auf Deci & Ryan bzw. Bohl &

Kucharz): „Im Gegenzug bewirkt die engagierte Aktivität des Selbst eine höhere Lernqualität und fördert zugleich die Entwicklung des individuellen Selbst“ (a.a.O., S. 145).

Bei der Konzipierung von Experimentierangeboten sind entwicklungs- und lernpsychologische Aspekte zu beachten. Analog zu den oben gekennzeichneten Niveaustufen der Begriffsentwicklung gehen wir davon aus, dass das Experimentieren durch Entwicklungsniveaus gekennzeichnet ist und vom Explorieren (Erproben, Erzeugen eines Phänomens) über das Versuchen (Prüfen von Vermutungen) zum Experimentieren (Prüfen von Hypothesen) verläuft (Hartinger 2020; Grygier & Hartinger 2009; Giest 2008a, b, 2011, 2016; Giest & Marquardt-Mau 2013).

Wir konzipierten unser Experimentierangebot daher als „Versuche“, bei denen zunächst das Prüfen von Vermutungen (vs. Hypothesen) im Vordergrund steht.

2.3 Außerschulische Lernorte

Das Gestalten einer Lernumgebung, die Kindern ohne fachkundige Unterstützung eine lernende Begegnung mit Natur- und Technikphänomenen gestatten soll, erscheint aus Sicht der Erkenntnisse naturwissenschaftsdidaktischer Forschung als auch aus Sicht eigener praktischer Erfahrungen und wissenschaftlicher Untersuchungen sowie der Evaluation von außerschulischen Lernorten (u. a. Science Centern – Alfieri et al. 2011; Graf & Naschka-Roos 2009; Roth et al. 1997; Lewalter & Geyer 2005, 2009; Itzek-Greulich et al. 2015) als problematisch. Daher müssen beim außerschulischen Lernen (z.B. in Science Labs oder im Museum) jüngeren Kindern entweder durch Erzieher*innen oder durch Aufsichtspersonal (z.B. Student*innen im Rahmen der Ausbildung) ausreichend Hilfen bzw. Orientierungen angeboten werden. Ein gewisses, aber zu lösendes Problem besteht zudem in der Auswahl und Gestaltung der Versuche in einer solchen Weise, dass sie in einer für Kinder ausreichend robusten Präsentationsform vorliegen (Feinmotorik).

3 Fragestellung

Im Rahmen unserer Untersuchung gingen wir der zentralen Frage nach: Lassen sich im Rahmen eines Feldversuchs an außerschulischen Lernorten von uns entwickelte Experimentierstationen so einsetzen, dass über eine adaptive instruktionale Stützung eine möglichst hohe Handlungsregulation (als Grundstock des eigenständig regulierten Experimentierens) zur adäquaten Kenntniserneuerung führt?

Im Einzelnen fragten wir:

- 1) Unterscheidet sich die Handlungsregulation an den Stationen systematisch?
- 2) Gibt es einen Zusammenhang zwischen dem Vorwissen und der Handlungsregulation?
- 3) Wie wirken sich Vorwissen und Stationsbesuch auf den Lerngewinn der Kinder aus?
- 4) Gibt es einen Unterschied zwischen Mädchen und Jungen sowie den Klassenstufen bei Stationsbearbeitung, Handlungsregulation und Lernwerten?
- 5) Moderiert bzw. mediatisiert die Handlungsregulation die Beziehung Vorwissen und Posttestwert/Lerneffekt?
- 6) Lassen sich die Stationen systematisch (ggf. nach unterschiedlicher Dominanz von Hands-on- und Minds-on-Aktivitäten) unterscheiden?

4 Methoden

4.1 Design der Untersuchung

Untersucht wurden Merkmale der Handlungsregulation und Lerneffekte halboffenen Experimentierens an Stationen (freie Wahl der Stationen und der Verweildauer mit Angeboten für eine tutorielle Unterstützung und Nutzung von speziell entwickelten und erprobten Forscher*innenheften). Aus den vielfältigen, nicht mehr zu überschauenden Angeboten an Mate-

rialien zum naturwissenschaftlichen Lernen (Literatur, Internet/Bildungsserver, inzwischen auch spezielle Apps), insbesondere zum Experimentieren, wurden zu den Aufgabenbereichen der Stadtwerke passende Experimente (Wasser, Energie, Umwelt, Mobilität) ausgewählt. Da diese in der Regel nicht wissenschaftlich evaluiert wurden, sondern bestenfalls erprobte Materialien oder auch nur Vorschläge und Ideensammlungen darstellen, wurden sie von uns systematisch bearbeitet (Forscher*innenhefte – enthalten die Beschreibung der Experimente sowie entsprechende Lernaufgaben zur Vermutungsbildung, Beobachtung und Ergebnissicherung; Lehrer*innenhefte zur pädagogischen und fachlichen Reflexion der Experimente – diese waren jedoch nicht Gegenstand der Evaluation).

Die Evaluation der Experimentierstationen erfolgte im Rahmen von „Experimentiertagen“ in der Pädagogischen Werkstatt der Universität Potsdam. Dabei wurde folgendes Forschungsdesign verwendet:

- Nach der Begrüßung und kurzen Einweisung erfolgt ein Pretest im Plenum („Quiz“ mit multiple Choice-Aufgaben zu jeder Station).
- Die Kinder besuchen die Stationen und beginnen mit der Arbeit, wobei das Forscher*innenheft genutzt werden kann. Die Tutor*innen an der Station helfen bei Bedarf. Während der Arbeit an den Stationen werden die Kinder beobachtet.
- Am Ende erfolgt ein Posttest mit identischen Aufgaben im Plenum.

Durch die Gestaltung des Ablaufs entsteht die Situation, dass nicht alle Kinder jede Station bearbeiten und daher nicht beobachtet werden konnten, wodurch je eine Experimentalgruppe (bearbeitet) und Kontrollgruppe (nicht bearbeitet) entstand.

4.2 Abhängige und unabhängige Variablen

Abhängige Variablen sind:

Posttestwert (Begriffsniveau): Die Antworten im Rahmen eines Wissenstests (Quiz zum Begriffsniveau – in der Regel fünf Multiple-Choice-Aufgaben – z.B. „Warum kann ein Insekt auf der Wasseroberfläche stehen, ein Vogel aber nicht?“) wurden je nach Niveaustufe bepunktet:

begriffliches Niveau	Beispiel	Pkt.
sinnlich konkret (Benennung des Phänomens)	„...weil Vögel untergehen, Insekten aber nicht“	1
abstrakt	„...weil Wasser eine dünne Haut bildet, auf der Insekten gerade noch stehen können, Vögel aber nicht“	2
geistig-konkret (Erklärung des Phänomens durch Theorieaussage)	„...weil die Wasserteilchen nur zum Wasser hin angezogen werden, bilden sie so etwas wie eine Haut, die Insekten trägt, Vögel aber nicht“	3
fake – Phänomen- und Theorieebene falsch verbunden [„wissenschaftlich klingender Unsinn“]	„...weil die Wasserteilchen eine Haut bilden“	0

Da der Wissenstest für den Experimentiertag konstruiert war, existiert in der Regel nur eine Niveau-Aufgabe pro Station, was lediglich Trendaussagen gestattet, die näher untersucht werden müssten.

Lerneffekt: Dieser errechnet sich aus der Differenz zwischen Post- und Pretestwerten.

Handlungsregulation: Die Kinder wurden beim Experimentieren an den Stationen von Studierenden beobachtet, wobei die beobachteten Aktivitäten hinsichtlich folgender Items der Eigenständigkeit der Handlungsregulation (Fragebogen) beurteilt wurden:

Merkmal der Handlungsregulation	Item	Punktwert
Orientierung – Arbeitsaufnahme	zielgerichtet	3
	durch Nutzung des Forscher*innenheftes	2
	instruiert	1
	ohne	0
explizite Instruktion	ja	1
	nein	2
Hilfe	ohne/ selbständig/ von anderen Kindern	3
	durch Forscher*innenheft	2
	durch Erwachsene	1
Handlung	selbstständig	4
	miteinander	3
	durch Nachmachen	2
	durch Zusehen	1
Experimentieren	Experiment	3
	Versuch	2
	Exploration	1
Handlungsverlauf	sicher	4
	sicher mit Instruktion	3
	unsicher	2
	unsicher mit Instruktion	1
	ohne	0
Kooperation	allein/Paar	3
	Gruppe	2
	anders	1
	ohne	0
Handlungshilfe	von Kindern/ohne	3
	durch Forscher*innenheft	2
	durch Erwachsene	1
Kontrolle	ja	1
	nein	0
Kontrolle/Feedback durch	Kinder	3
	Forscher*innenheft	2
	Erwachsene	1
	ohne	0

Die Prüfung der Reliabilität der Beobachtungsdaten (10 Items) ergab weitgehend zufriedenstellende Werte (Cronbachs Alpha zwischen 0,511 und 0,701), die stichprobenartig erhobene Interraterreliabilität war leider nicht immer zufriedenstellend (Kappa zwischen 0,4 und 0,5). Die ermittelten Werte lassen sich daher nicht zu Messzwecken inhaltlich nutzen, sondern sind nur in ihrer relativen Höhe im Sinne eines mehr oder weniger eigenregulierten Handelns zu interpretieren.

Unabhängige Variablen sind

- Wert des Pretests = Vorwissen (analog Posttest)
- Bearbeitung der Station (bearbeitet/nicht bearbeitet)
- Station (bzw. Inhalt der Station)
- Gender
- Klassenstufe (3 bzw. 4)

Folgende Experimentierstationen wurden entwickelt:

- *Stationen zum Thema Wasser*: Oberflächenspannung (Wasserhaut – Oberflächenspannung mit Büroklammer testen; „Wasserberg/Überlaufen“ – Kohäsionskraft nutzen, um ein Glas mit Wasser über den Rand hinaus zu füllen; Lösung – Salz und Zucker „verschwinden“ im Wasser; das geschieht jedoch nicht mit allen Stoffen – Gemisch; „Wasser geht in die Luft“ – Aggregatzustände; Wasserkreislauf; Klärwerk – mechanische Klärstufen;

- *Stationen zum Themenkreis Energie:* Intro zum Problem Wärme als Energieform; „Eine Nuss brennt“ – Wärmeenergie aus der Nuss; „Welcher Löffel wird heiß?“ – Wärmeleitung; „Wärme einsperren“ – Wärmeisolation (Thermoskanne bauen und testen); Energieumwandlung (Lasten heben mit einem Windrad); Lichtreflexion und Lichtabsorption („Schwarz wird warm, Weiß nicht“); Elektrostatik (tanzender Stern); Dynamo und Photovoltaik (nicht untersucht);
- *Stationen zum Themenkreis Umwelt:* Müll richtig trennen; Wir trennen Eisen vom Müll (Elektromagnet); Glas richtig sortieren; Schöpfrahmen bauen; Papierrecycling; Müllverwertung (Upcycling mit einem Tetrapack); Müllkreislauf;
- *Stationen zum Themenkreis Mobilität:* Fahr- und Linienpläne richtig nutzen; Reaktionszeiten testen und bezüglich Bremswege bewerten; Wagen – Rollreibung auf verschiedenen Untergründen testen; sehen und gesehen werden – toten Winkel am Fahrzeugmodell testen; Smog erzeugen; Fahrgastsicherheit – Trägheit auf das Verhalten im Bus anwenden – Festhalten).

Aus Platzgründen können die 26 Stationen hier nicht detailliert inhaltlich und bezüglich ihrer Lernanforderungen beschrieben werden. Bei vielen der Stationen sind jedoch die zu explorierenden Effekte und die zu prüfenden Vermutungen auf der Phänomenebene angelegt. Bei einer Reihe von Stationen (z.B. Wasserhaut, Smog, Schwimmen, tanzender Stern) ist das Erzeugen des Phänomens für Kinder zwar attraktiv und eigenreguliert möglich, das Prüfen einer Vermutung/Hypothese wegen des abstrakten oder unanschaulichen Zusammenhangs (ggf. auf einer Modellebene – z.B. Teilchen- bzw. Ladungsmodell) aber schwierig.

5 Ergebnisse

5.1 Datengrundlage

Aus der Gesamtuntersuchung liegen $N(\text{mit/ohne Bearbeitung})=3.120$ (2.181/939) Datensätze zu 26 Stationen vor (Lerndaten nur zu 24 Stationen), 1.635 (1.155/480) der Daten stammen von Jungen und 1.485 (1.026/459) von Mädchen; $N=1.041$ (752/289) aus Klassenstufe 3 und $N=2.079$ (1.429/650) aus Klassenstufe 4.

Die Daten aller abhängigen Variablen sind nicht normalverteilt, daher werden parameterfreie Verfahren zur Analyse verwendet. Die Analyse von Wechselwirkungen muss einzeln erfolgen, denn die Homogenität der Varianzen der abhängigen Variablen ist nur für den Faktor Gender (Levene-Test) gegeben, daher können allgemeine lineare Modelle nicht eingesetzt werden. Regressionsanalysen erfordern die Normalverteilung der Residuen, die im Datensatz gegeben ist (Prüfung von Kurtosis < 3 und Schiefe < 1), unabhängig davon wird Bootstrapping genutzt (da dies bei großem Stichprobenumfang den Verstoß gegen eine ggf. fehlende Normalverteilung kompensiert).

5.2 Ergebnisse der empirischen Untersuchung

F1: Unterscheidet sich die Handlungsregulation an den Stationen systematisch?

H1: Durch die unterschiedlichen inhaltlichen Anforderungen an den Stationen werden verschiedene Bedingungen für die Handlungsregulation gestiftet, weshalb sich die Stationen diesbezüglich systematisch unterscheiden.

Ergebnis: Im Vergleich aller Stationen (Kruskal-Wallis-Test) weist die Handlungsregulation hochsignifikante Unterschiede mit großer Effektstärke (vgl. Lenhard & Lenhard 2016) auf – $H(26, 2173)=628,504$; $p<0,001$; $d=1,251$, was die Bedeutung des Inhalts der Station für die Handlungsregulation unterstreicht (Tab. 1).

Station	M (Handlungsregulation)	SD
alle	1,553	0,42
Mädchen	1,548	0,42
Jungen	1,557	0,42
Klasse 3	1,649	0,38
Klasse 4	1,501	0,43
Überlaufen	0,968	0,18
Licht	0,969	0,38
Löffel	1,042	0,40
Windrad	1,259	0,24
Isolierung	1,264	0,27
Gemisch	1,299	0,24
Nuss	1,341	0,23
Reaktion	1,346	0,43
Aggregat	1,371	0,28
T-Stern	1,470	0,43
Klärwerk	1,482	0,36
E-Magnet	1,489	0,30
Glas	1,505	0,40
Intro_Energie	1,519	0,45
Schöpfrahmen	1,572	0,33
Müllkreis	1,603	0,52
Papier	1,612	0,36
Smog	1,625	0,26
Wagen	1,647	0,33
Schwimmen	1,664	0,36
Wasserhaut	1,670	0,39
Upcycling	1,721	0,32
Boot	1,830	0,43
Trägheit	1,861	0,32
T-Winkel	1,884	0,31
Fahrplan	1,925	0,33

Tab. 1: Rangreihe der Handlungsregulation/Testwerte (Mittelwerte und Standardabweichungen) nach Stationen und zentralen abhängigen Variablen

H1 ist anzunehmen: Der Unterschied der Handlungsregulation an den Stationen weist eine starke inhaltliche Komponente auf, die sich auf das Maß der Eigenständigkeit des Experimentierens auswirkt. Diese inhaltliche Komponente wirft allerdings die Frage nach der Wissensabhängigkeit der Handlungsregulation auf.

F2: Gibt es einen Zusammenhang zwischen dem Vorwissen und der Handlungsregulation?

H2: Bei höherem Vorwissen müsste eine höhere Handlungsregulation (vor allem bezogen auf das Erzeugen des beabsichtigten Effekts) beobachtet werden.

Ergebnis: Das Vorwissen korreliert generell schwach aber signifikant positiv mit der Handlungsregulation ($r=0,094^{**}$). Bezogen auf die einzelnen Stationen gibt es nicht signifikante positive Korrelationen bei E-Magnet, Wasserhaut, Smog, Papier, Schwimmen, Boot, T-Winkel, Nuss, Aggregatzustände, Gemisch, Reaktion, Trägheit und negative bei T-Stern, Wagen, Überlaufen, Isolierung ($r=-0,602^{**}$), Löffel, Windrad, Licht, Intro-Energie, Glas, Müllkreis, Fahrplan.

Eine lineare Regressionsanalyse (JASP 0.14.1¹) ergab sehr kleine Moderationseffekte bezüglich der Interaktion von Vorwissen (Tab. 2) und Station hinsichtlich der Voraussage der Handlungsregulation. Das Gesamtmodell klärt aber nur 1% der Varianz auf. Der Effekt ist $\Delta R^2=0,3\%$; $f(1,1478)=4,347$; $p=0,037$; 95%CI [-0,005; -1,451e-4], allerdings schließt das Vertrauensintervall beim Bootstrapping sehr knapp Null ein. Für andere Faktoren sind keine Moderationseffekte festzustellen.

Die Prüfung der Mediation bezüglich der Beziehung Vorwissen → Handlungsregulation (direkter Effekt, d.E.= 0,035) erbrachte nur für Klassenstufe → Handlungsregulation einen signifikanten indirekten Effekt (i.E.=0,003; $z=-1,988$; $p=0,047$), der zu einem totalen Effekt von t.E.=0,032 führte – Werte Klasse 3>4 – s.u.).

H2 ist von daher nur mit Vorsicht anzunehmen, da die Effekte erwartungswidrig gering sind: Das Vorwissen spielt, anders als der Inhalt der Stationen, hinsichtlich der Handlungsregulation eine geringe Rolle.

F3: Wie wirken sich Vorwissen und Stationsbesuch auf den Lerngewinn der Kinder aus?

H3: Der Stationsbesuch sollte sich, zwar entsprechend der Erkenntnislage mit geringen Effektstärken, dennoch in Abhängigkeit vom Vorwissen, in höheren Lernwerten niederschlagen, wobei größere Unterschiede zwischen den Stationen zu beobachten sein sollten (Domänenspezifik des Wissens).

Ergebnis: Die verschiedenen (Inhalte der) Stationen weisen z.T. hochsignifikante Unterschiede bei Vorwissen und Lernwerten auf. Die jeweiligen Effektstärken liegen bei Bearbeitung: Vorwissen: $d=1,059$; Posttest: $d=1,151$; Lerneffekt: $d=0,406$; ohne Bearbeitung: Vorwissen: $d=1,399$; Posttest: $d=1,391$; Lerneffekt: $d=0,545$. Der Inhalt der Station erweist sich als wichtiger Faktor mit Blick auf Vorwissen und Lerneffekte, gleichzeitig führt der zum Pretest identische Posttest zu einem z.T. deutlichen Lerneffekt (Tab. 2 – fett markierte Stationen).

Station	N (bearbeitet/ nicht bearbeitet)	Pretest	Posttest	Pretest	Posttest
		bearbeitet		nicht bearbeitet	
alle	1494/705	1,555	1,729	1,524	1,626
Mädchen	711/340	1,527	1,721	1,505	1,595
Jungen	784/365	1,571	1,735	1,542	1,655
Klasse 3	505/216	1,482	1,656	1,461	1,574
Klasse 4	989/489	1,592	1,766	1,552	1,649
Überlaufen	9/1	1,78	1,67	0,00	0,00
Licht	52/0	1,89	1,26	-/-	-/-
Löffel	81/32	1,92	2,40	1,91	2,43
Windrad	63/6	1,24	1,71	1,00	1,00
Isolierung	47/46	1,64	1,92	1,87	1,56
Gemisch	11/0	1,36	1,64	-/-	-/-
Nuss	131/28	1,21	1,42	1,76	1,61
Reaktion	26/54	1,65	2,08	2,09	1,85
Aggregat	79/0	1,64	1,91	-/-	-/-
T-Stern	78/86	1,61	1,58	0,77	1,25
Klärwerk	146/81	mehrheitlich nicht vergleichbar			
E-Magnet	166/105	1,72	2,39	1,78	2,33
Glas	39/63	0,26	0,41	0,11	0,25
Intro_energie	72/30	1,50	3,00	2,00	2,25

¹ Die Analyse wurde mit SPSS gegengeprüft und erbrachte analoge Werte.

Schöpfrahmen	18/40	-/-	-/-	-/-	-/-
Müllkreis	74/28	0,16	0,30	0,43	0,25
Papier	109/71	2,23	2,23	2,37	2,22
Smog	129/46	1,87	1,96	1,85	1,98
Wagen	93/13	1,65	1,85	1,58	1,08
Schwimmen	73/20	0,93	0,91	0,95	1,20
Wasserhaut	230/43	1,72	2,39	1,78	2,33
Upcycling	76/26	-/-	-/-	-/-	-/-
Boot	173/33	1,12	1,70	1,54	1,65
Trägheit	52/28	2,06	2,13	2,06	2,00
T-Winkel	96/28	2,25	2,33	1,98	2,30
Fahrplan	49/31	1,66	1,85	1,85	1,82

Tab. 2: Testwerte (Mittelwerte) der Stationen – nach Rangreihe der Handlungsregulation geordnet

Bei identischen Pre- und Posttests muss ein möglicher Lerneffekt mitbedacht werden, wie die Daten ohne Bearbeitung nahelegen, dennoch liegt der Posttestwert bei Bearbeitung etwas höher als ohne, was die aus der Literatur bekannten geringen Lerneffekte an außerschulischen Lernorten bestätigt.

Anders als bei der Handlungsregulation spielt das Vorwissen für die Lerneffekte eine wichtige Rolle (Tab. 3).

Vergleich Pre-Post-Test (Wilcoxon-Test)

mit Bearbeitung, für alle Stationen (positiver Lerneffekt)	$z(1,1473)=-5,243; p<0,001; d=0,276$
ohne Bearbeitung, für alle Stationen (positiver Lerneffekt)	$z(1,693)=-2,342; p=0,019; d=0,179$
für einzelne Stationen (positiver Lerneffekt)	E-Magnet, Wasserhaut, Boot, Nuss, Löffel, Windrad, Intro-Energie, Reaktion
Vergleich der Testwerte (Mann-Whitney-Test – ohne vs. mit Bearbeitung)	
alle Stationen – Vorwissen (n.s.) Posttest (B.>o.B.)	$z(1,2199)=-2,058; p=0,040; d=0,088$
Nuss: Vorwissen (B.<o.B.)	$z(1,143)=-2,11; p=0,035; d=0,359$
T-Stern: Vorwissen (B.>o.B.) Posttest (B.>o.B.) Lerneffekt (B.<o.B.)	$z(1,150)=-5,353; p<0,001; d=0,972$ $z(1,143)=-2,279; p=0,023; d=0,388$ $z(1,143)=-2,569; p=0,010; d=0,44$
Wagen: Posttest (B.>o.B.) Lerneffekt (B.>o.B.)	$z(1,77)=-2,526; p=0,012; d=0,601$ $z(1,77)=-2,125; p=0,034; d=0,499;$
Isolierung: Lerneffekt (B.>o.B.)	$z(1,65)=-2,876; p=0,004; d=0,764;$
Reaktion: Vorwissen (B.<o.B.) Lerneffekt (B.>o.B.)	$z(1,80)=-3,089; p=0,002; d=0,736$ $z(1,80)=-2,889; p=0,004; d=0,683$

Tab. 3: Signifikante Unterschiede bei den Testwerten

Die Interaktion von Vorwissen und Bearbeitung der Station sagt die Posttestwerte signifikant voraus: das Gesamtmodell ist signifikant und weist eine Varianzaufklärung von 19,3% auf. Der Moderationseffekt der Bearbeitung der Station auf die Beziehung Vorwissen und Posttestwert ist zwar klein, aber signifikant: $\Delta R^2=0,5\%$; $f(1,2162)=7,227$; $p<0,001$; 95% CI[-0,214; -0,051]. Für die Interaktion von Vorwissen und Station bezüglich der Vorhersage der Posttestwerte gilt: mit Bearbeitung klärt das signifikante Gesamtmodell 17,8% der Varianz für Posttestwerte auf, für Lerneffekte 30,5%. Der Moderationseffekt ist in beiden Fällen signifikant, aber identisch klein: Posttestwert – $\Delta R^2=2,0\%$ /Lerneffekt – $\Delta R^2=1,7\%$; $f(1,1469)=35,936$; $p<0,001$; 95% CI [0,012; 0,023]. Ohne Bearbeitung gilt analog: das Modell ist sign., bei Lerneffekt mit 30%, bei Posttestwert mit 33,5% der Varianzaufklärung: Post-

testwert – $\Delta R^2=2,3\%$ /Lerneffekt – $\Delta R^2= 2,4\%$; $f(1,689)=23,979$; $p<0,001$; 95% CI[0,010; 0,024].

Insgesamt lässt sich keine signifikante indirekte Wirkung des Faktors Bearbeitung auf den Effekt Vorwissen→Posttestwert/Lerneffekt nachweisen, hinsichtlich der Mediation² des Effekts Vorwissen→Posttestwert/Lerneffekt durch den Faktor Station – vgl. die nachfolgende Übersicht:

untersuchter Effekt	direkter Effekt	indirekter Effekt (Faktor Station)	totaler Effekt
mit Bearbeitung		0,022 (z=5,169; p<0,001)	
Vorwissen→Posttestwert	0,379		0,397
Vorwissen→Lerneffekt	-0,621		-0,603 ³
ohne Bearbeitung		0,029 (z=3,382; p<0,001)	
Vorwissen→Posttestwert	0,501		0,530
Vorwissen→Lerneffekt	-0,499		-0,470

Bezüglich der begrifflichen Niveaustufen sind die Effektstärken für das Projekt zwar enttäuschend, aber – der Erkenntnislage entsprechend – nur in Ausnahmefällen nicht klein, wobei die Effekte der identischen Pre- und Posttests und der Bearbeitung der Stationen zusammenwirken. Die Bedeutung von Vorwissen und Inhalt der Station für den Lerneffekt wurde entsprechend H3 bestätigt.

F4: Gibt es einen Unterschied zwischen Mädchen und Jungen sowie den Klassenstufen bei Stationsbearbeitung, Handlungsregulation und Lernwerten?

H4: Aus etlichen (für einen Überblick vgl. Oppermann & Keller 2018) auch eigenen Untersuchungen (Giest 2002) ist bekannt, dass bezüglich des naturwissenschaftlichen Lernens die Unterschiede zwischen den Mädchen und Jungen in der Grundschule (mit Ausnahme der absoluten Leistungsspitze – Bos et al. 2012) noch relativ gering sind: „Auf der Basis von TIMSS 2015 lassen sich für Viertklässlerinnen und Viertklässler in Deutschland für die Gesamtskala Naturwissenschaften keine signifikanten leistungsbezogenen Disparitäten feststellen“ (a.a.O., S. 22). In der betrachteten Klassenstufe könnten bei außerschulischen Lernangeboten dennoch gewisse Interessensunterschiede bei den Geschlechtern Einfluss auf Stationsbearbeitung, Handlungsregulation und Lernwerte haben. Bezüglich der Klassenstufen sind keine großen Unterschiede zu erwarten. Einzelne Klassen unterscheiden sich in der Regel mehr als Klassenstufen, ferner war keine Randomisierung des Samples möglich und die Untersuchungen fanden im Sommer (Schuljahresende) und im Winter statt, sodass die Differenz zwischen der Sommeruntersuchung in Klasse 3 kein ganzes Schuljahr zur Winteruntersuchung in Klasse 4 darstellt.

Ergebnis:

- a) Stationsbearbeitung: Es lassen sich keine wesentlichen Unterschiede im Verhältnis von Mädchen und Jungen mit und ohne Stationsbearbeitung feststellen: 71% Jungen und 69% Mädchen bearbeiteten die Stationen, 29% Jungen bzw. 31% Mädchen nicht. Bei Mädchen liegt das Minimum von 28% Bearbeitung bei Isolierung, dann folgen Trägheit, Schwimmen, Boot (38-42%), das Maximum von 59% erreicht der Tanzende Stern. Mädchen bevorzugen tendenziell Stationen, an denen gebastelt oder sorgfältig beobachtet wird (T-Stern, Intro-Energie, Müllkreislauf, Gemisch), Jungen solche, an denen ein Wettbewerbscharakter auszumachen ist (Isolierung, Trägheit, Schwimmen, Boot (72-58%). Die teilnehmende Beobachtung sowie Videoaufzeichnungen ergaben, dass die Kinder (weitge-

² Moderation bedeutet, dass der Effekt einer unabhängigen Variable X auf eine abhängige Variable Y für verschiedene Werte einer dritten Moderator-Variablen unterschiedlich ist. Mediation bedeutet, dass der Effekt einer Variablen X auf eine Variable Y (partiell) indirekt ist: er wird auch durch eine dritte Variable M verursacht.

³ Mehr Vorwissen wirkt sich negativ auf den Lerneffekt aus, weil mit hohem Vorwissen, wegen der begrenzten Niveaustufen, weniger hinzulernt werden kann.

hend unabhängig vom Geschlecht) vor allem jene Stationen besuchten, die noch freie Plätze boten. Analoges gilt auch für die Klassenstufen.

- b) Handlungsregulation: Der Faktor Gender ist hier unauffällig. Die Klassenstufe 3 weist erwartungswidrig eine leicht höhere Handlungsregulation auf – $z(1,2173)=-6,940$; $p<0,001$; $d=0,301$), die Effektstärke ist jedoch relativ klein. Auch die Mediationsanalyse ergibt für den Zusammenhang Vorwissen→Handlungsregulation eine Mediation des Effekts durch den Faktor Klassenstufe ($-0,003$; $z=-1,988$; $p=0,047$), der zu einer sehr kleinen negativen Effektveränderung (d.E.= $0,035$ →t.E.= $0,032$) führt, da die Klassenstufe 4 hier niedrigere Werte erzielt.
- c) Lernwerte: Bezüglich des Vergleichs von Pre- und Posttest ergaben sich einige kleine, aber signifikante positive Effekte, die im Sinne der Ergebnisse zu Frage F3 zu interpretieren sind (vgl. die folgende Übersicht).

Bearbeitung	Mädchen	$z(2,699)=-3,716$; $p<0,001$; $d=0,284$
	Jungen	$z(2,774)=-3,742$; $p<0,001$; $d=0,271$
	Klasse 3	$z(2,504)=-2,637$; $p=0,008$; $d=0,237$
	Klasse 4	$z(2,969)=-4,783$; $p<0,001$; $d=0,311$
ohne Bearbeitung	Jungen	$z(1,360)=-1,996$; $p=0,046$; $d=0,160$

Die Moderations- und Mediationsanalysen bleiben hier unauffällig.

Bezüglich des Vergleichs mit vs. ohne Bearbeitung gibt es nur bei Klassenstufe 4 einen signifikanten Unterschied beim Lerneffekt (B.>o.B.) – $z(1,1489)=-2,031$; $p=0,042$; $d=0,105$.

Eine Regressionsanalyse ergab bei Bearbeitung eine sign. Interaktion von Vorwissen und Klassenstufe hinsichtlich der Vorhersage der Posttestwerte/Lerneffekte: $\Delta R^2=0,8\%/ \Delta R^2=0,6\%$; $f(1,1469)=12,129$; $p<0,001$; 95% CI[0,078; 0,278], wobei das Modell Posttestwert 16% und Lerneffekt 30% der Varianz aufklärt, ohne Bearbeitung: $\Delta R^2=0,8\%/ \Delta R^2=0,7\%$; $f(1,689)=7,357$, $p=0,007$; 95% CI[0,052; 0,327], wobei das Modell Posttestwert 29% und Lerneffekt 27% der Varianz aufklärt.

H4 ist damit tendenziell bestätigt.

F5: Moderiert bzw. mediatisiert die Handlungsregulation die Beziehung Vorwissen und Posttestwert/Lerneffekt?

H5: Die Handlungsregulation wäre moderierende Variable, wenn sie (im Sinne einer abhängigen Variablen) vom Stationsbesuch/Inhalt der Station und vom Vorwissen beeinflusst wird sowie (im Sinne einer unabhängigen Variablen) den Wert des Posttests/Lerneffekts als abhängige Variable beeinflusst. Eine Mediation der Variablen ist anzunehmen, wenn indirekte Effekte (Vorwissen→Handlungsregulation→Posttestwert) existieren.

Ergebnis: Eine Moderation der Handlungsregulation hinsichtlich der Beziehung zwischen Vorwissen und Posttestwert/Lerneffekt konnte insgesamt nicht festgestellt werden, außer bei der Station Trägheit/Festhalten: hier moderiert die Interaktion zwischen Handlungsregulation und Vorwissen die Beziehung Vorwissen-Posttestwert/Lerneffekt: $\Delta R^2=15,2\%/ \Delta R^2=9,4\%$; $f(1,45)=9,555$; $p=0,003$; 95% CI[0,389; 1,842], wobei das Modell für Posttestwert 24%, für Lerneffekt 53% der Varianz aufklärt. Die Mediationsanalyse erbrachte insgesamt keine signifikanten indirekten Wirkungen (Vorwissen→Handlungsregulation→Posttest), auch nicht bei den einzelnen Stationen außer bei Isolierung ($Z=-2,182$; $p=0,029$; Effektveränderung: direkter Effekt: 0,413; indirekter Effekt: -0,266; totaler Effekt: 0,148 – hier hängen Handlungsregulation und Vorwissen sign. negativ, Handlungsregulation und Posttestwert positiv, also genäufig zusammen).

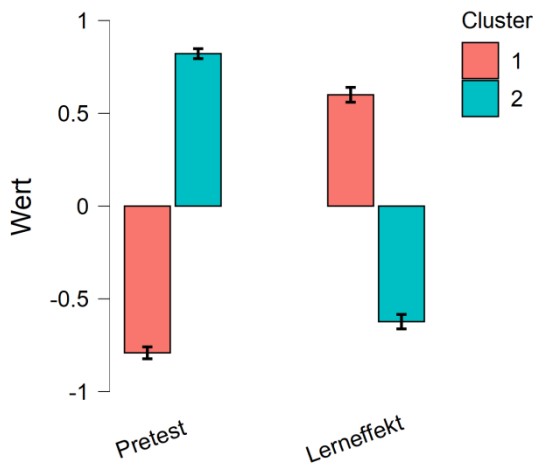
H5 ist zurückzuweisen, Handlungsregulation moderiert generell nicht signifikant die Beziehung Vorwissen Posttest/Lerneffekt, ist also diesbezüglich nicht als moderierende Variable anzusehen.

F6 Lassen sich die Stationen systematisch (ggf. nach unterschiedlicher Dominanz von Hands-on- und Minds-on-Aktivitäten) unterscheiden?

H6 In der Literatur (s.o.) finden sich immer wieder Hinweise auf eine Unterscheidung von „Hands-on“- und „Minds-on“-Aktivitäten, d.h. hinsichtlich einer unterschiedlichen Betonung manueller und mentaler Aktivitäten beim Experimentieren. Ggf. regen die verschiedenen Inhalte der Stationen diese beiden Aspekte unterschiedlich stark an, sodass sich die Stationen diesbezüglich systematisch unterscheiden lassen.

Ergebnis: Um den Hands-on vs. Minds-on-Aspekt zu prüfen, wurde eine Clusteranalyse mit den Variablen zum Wissensaspekt und zur Handlungsregulation durchgeführt. Diese erbrachte nur bei Einbeziehung der Variablen Vorwissen und Posttestwerte bzw. Lerneffekte mehr als ein, nämlich zwei Clusterzentren (Abb. 1), was im Rückschluss bedeutet, dass der Hands-on-Aspekt (Handlungsregulation bildet vor allem diesen Aspekt ab) kein bedeutendes Kriterium für eine gruppierende Unterscheidung der Stationen darstellt, obwohl diesbezüglich ein mehr oder weniger kontinuierlicher Unterschied zwischen allen Stationen existiert (Tabelle 1). Da das Modell (Pre- Posttest) nur 15,8% der Varianzaufklärung erfüllt, jenes (Pretest-Lerneffekt) aber 56% und im ersten Fall das Cluster 1 für alle Stationen dichter besetzt ist als Cluster 2, wird die Clusteranalyse Pretest-Lerneffekt genutzt (Abb. 1).

Abb. 1: Clustermittelwerte



Cluster 1 ist durch jeweils relativ niedriges Vorwissen und höhere Lerneffekte gekennzeichnet, der Anteil der Aufklärung der Clusterheterogenität beträgt hier 56%, bei Cluster 2 44%, welches durch ein relativ hohes Vorwissen aber geringe Lerneffekte charakterisiert ist. Wird die Zuordnung der Fälle zu den Clustern nach Stationen geordnet, kann die jeweilige Besetzung hinsichtlich der Größe verglichen werden. Die Besetzung in Cluster 1 ist bei folgenden Stationen größer als in Cluster 2: E-Magnet, Nuss, T-Stern, Wagen, Aggregatzustandsänderung, Löffel, Windrad, Glas, Reaktion, Trägheit. Die Besetzung in Cluster 2 ist größer als in Cluster 1 bei: Wasserhaut, Smog, Papier, T-Winkel, Licht, Müllkreislauf, Fahrplan.

Der Vergleich beider Stations-Gruppen (Mann-Whitney-Test) zeigt signifikante Unterschiede bei allen Variablen außer bei der Handlungsregulation (vgl. die folgende Übersicht):

Mann-Whitney-U-Test	W	n1/n2	p	d
Handlungsregulation	198270,5	1038/963	0,509	
Lerneffekt	301867,5	1101/1022	<0,001	0,875
Posttest	288484,0	1124/1032	0,007	0,966
Vorwissen	232807,5	1117/1023	<0,001	1,194

Inhaltlich lassen sich die Gruppen unter Berücksichtigung der Erfahrungen aus der teilnehmenden Beobachtung folgendermaßen interpretieren:

Die 1. Gruppe (Cluster 1) ist gekennzeichnet durch geringeres Vorwissen aber höhere Lerneffekte (Posttestwerte): Anleitung führt (zur adäquaten Handlung und) zum Lernzuwachs, Phänomene werden neu entdeckt und können auf der Phänomenebene erklärt werden. Die 2. Gruppe (Cluster 2) kennzeichnet ein höheres Vorwissen ohne Lernzuwachs und relativ geringe Posttestwerte. Hier werden vor allem Phänomene erzeugt, aber nicht tiefer verstanden (das Vorwissen auf der Phänomenebene kann nicht erweitert werden).

Bei Klärwerk, Upcycling, Schöpfrahmen fehlen Lernwerte oder sind nicht vergleichbar, bei Überlaufen, Gemisch, Isolierung und Intro-Energie liegen nicht ausreichend Lerndaten vor, sodass eine aussagekräftige Zuordnung zu den Clustern nicht möglich ist.

H6 ist dahingehend zu präzisieren, dass sich in der Tat zwei Hauptgruppen identifizieren lassen, die sich vor allem in den Wissens- und Lernwerten und weniger in der Handlungsregulation unterscheiden. Eine systematische Unterscheidung der Stationen nach der Dominanz von Hands-on- oder Minds-on-Aktivitäten lässt sich durch die Daten nicht stützen, die Handlungsorientierung weist aber dennoch eine starke inhaltliche Komponente auf (F1).

6 Fazit

Experimentierstationen stellen in der Regel kaum vergleichbare und insgesamt sehr unterschiedliche Anforderungen an das Handeln, die Handlungsregulation sowie Wissen und Lernen/Verstehen, wobei hinsichtlich der Geschlechtergruppen und Klassenstufen relativ geringe Unterschiede festgestellt wurden. Generell sind die Lerneffekte des Experimentierangebots hinsichtlich des Conceptual Change-Aspekts bescheiden, obwohl (oder gerade weil) die beobachteten Hands-on Aktivitäten sowie Spaß und Freude am Experimentierangebot gegeben waren.

Die angebotenen Stationen, obwohl alle für die Klassenstufen in der einschlägigen Literatur bzw. in Medien (z.B. Löwenzahn) empfohlen, unterscheiden sich hinsichtlich der Handlungsregulation, des Vorwissens und Lernzuwachses bzw. der Orientierung auf das begriffliche Niveau erheblich.

Für den Einsatz an außerschulischen Lernstationen wäre es sinnvoll, Experimentierstationen hinsichtlich der ermittelten zwei Cluster zu klassifizieren und entsprechende Konsequenzen für den Unterricht zu ziehen. An außerschulischen Lernorten sollten angeboten werden:

- Stationen/Experimente, bei denen es mehr um den Wissenserwerb geht, was mit bzw. ohne Anleitung verbunden sein kann. Hier führt adaptive (bedarfsgerechte) Anleitung in der Regel zur adäquaten Handlung und zum Lernzuwachs, wobei Phänomene vor allem auf der Oberflächen-/Phänomenebene erklärt werden können bzw. die Bewältigung höherer kognitiver Anforderungen (auf Tiefen-/Modellebene) an eine adäquate unterrichtliche Vor- und Nachbereitung gebunden ist.
- Stationen, bei denen Handeln mit hoher Eigenregulation möglich ist (mit erfolgreicher bzw. ohne Anleitung und wenig abhängig vom Vorwissen), bei denen vor allem eigenen Fragen nachgegangen, viel Neues entdeckt (vor allem beobachtet) werden kann, wobei nicht notwendig jeder zugrunde liegende naturwissenschaftliche Zusammenhang immer verstanden werden muss.

Bei der Interpretation und Bewertung der Untersuchung ist nochmals zu betonen, dass es sich nicht um eine Laborstudie handelt, sondern das eigentliche Ziel des Projekts in der Entwicklung und Evaluation der Experimentierstationen bestand. Diesem Ziel mussten sich die wissenschaftlichen Begleituntersuchungen unterordnen, was die saubere Variablenkontrolle und Randomisierung behinderte.

Literatur

- Alberts, Sonja & Giest, Hartmut (2011): Es hat Spaß gemacht. Über das Lernen in Science Centern. In: Hartmut Giest, Astrid Kaiser & Claudia Schomaker (Hrsg.): Sachunterricht – auf dem Weg zur Inklusion. Bad Heilbrunn, S. 151-156.
- Alberts, Sonja & Giest, Hartmut (2012): Lernen an Experimentierstationen im Science Center. In: Hartmut Giest, Eva Heran-Dörr & Carmen Archie (Hrsg.): Lernen und Lehren im Sachunterricht. Bad Heilbrunn, S. 79-86.
- Alfieri, Louis, Brooks, Patricia J., Aldrich, Naomi J. & Tenenbaum, Harriet R. (2011): Does discovery-based instruction enhance learning? In: *Journal of Educational Psychology*, 103, S. 1-18.
- Anderson, David, Lucas, Keith B. & Ginns, Ian S. (2003): Theoretical Perspectives on Learning in an Informal Setting. In: *Journal of Research in Science Teaching*, 40, 2, S. 177-199.
- Bundesministerium für Bildung und Forschung (BMBF) (2012): Perspektive MINT. Bonn.
- Bohrmann, Mareike, Todorova, Maria & Möller, Kornelia (2016): Welchen Einfluss hat der bereichsspezifische Aus- und Fortbildungshintergrund von Sachunterrichtslehrkräften auf die Bewertung und Entwicklung von Experimenten bei Grundschulkindern? In: Hartmut Giest, Thomas Goll & Andreas Hartinger (Hrsg.): Sachunterricht – zwischen Kompetenzorientierung, Persönlichkeitsentwicklung, Lebenswelt und Fachbezug. Bad Heilbrunn, S. 99-107.
- Bos, Wilfried, Wendt, Heike, Köller, Olaf. & Selter, Christoph (Hrsg., 2012): TIMSS 2011. Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich. Münster.
- Bos, Wilfried, Wendt, Heike, Köller, Olaf, Selter, Christoph Schwippert, Knut & Kasper, Daniel (2016): TIMSS 2015: Wichtige Ergebnisse im Überblick. In: Heike Wendt, Wilfried Bos, Christoph Selter, Olaf Köller, Knut Schwippert & Daniel Kasper (Hrsg.): TIMSS 2015: Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich. Münster, S. 13-30.
- Dunker, Nina (2016): Überzeugungen von Sachunterrichtslehrkräften zum Experimentieren im Unterricht. In: Hartmut Giest, Thomas Goll & Andreas Hartinger (Hrsg.): Sachunterricht – zwischen Kompetenzorientierung, Persönlichkeitsentwicklung, Lebenswelt und Fachbezug. Bad Heilbrunn, S. 107-115.
- Edelsbrunner, Peter A., Schalk, Lennart, Schumacher, Ralph, & Stern, Elsbeth (2018): Variable control and conceptual change. A large-scale quantitative study in elementary school. In: *Learning and Individual Differences*, 66, pp. 38-53.
- Elschenbroich, Donata (2005): Weltwunder. Kinder als Naturforscher. München.
- Evaluationsbericht (2016): Erstellt von Giest, H., Egbert, B., Heiden, S., Lachmann, F. & Wandel, K., unveröffentlichtes Material. Potsdam.
- Falk, John H. (2004): The Director's Cut: Toward an Improved Understanding of Learning from Museums. In: *Science Education*, 88, 1, pp. 83-96.
- Freericks, Renate (2011): Außerschulische Lernorte: Typologie und Entwicklungsstand. In: Renate Freericks & Dieter Brinkmann (Hrsg.): *Zukunftsfähige Freizeit. Analysen – Perspektiven – Projekte*. 1. Bremer Freizeitkongress Hochschule Bremen; Dokumentation der Fachtagung 12./13. November 2010. Bremen, S. 11-22.
- Gesellschaft für Didaktik des Sachunterrichts (GDSU) (Hrsg., 2013): *Perspektivrahmen Sachunterricht*. Vollständig überarbeitete und erweiterte Ausgabe. Bad Heilbrunn.
- Giest, Hartmut (2002): *Entwicklungsfaktor Unterricht*. Landau.
- Giest, Hartmut (2008a): Experimentieren und Problemlösen als Lernhandlungen. In: *Grundschulunterricht/Sachunterricht*, 2, S. 4-9.
- Giest, Hartmut (2008b): Problemlösen und Experimentieren in der Grundschule. Probleme, Perspektiven und Beispiele. In: *Grundschulunterricht/Sachunterricht*, 2, S. 15-19.
- Giest, Hartmut (2011): Wissensaneignung, Conceptual Change und die Lehrstrategie AK. In: *Tätigkeitstheorie*, 4, 65-100. http://www.ich-sciences.de/media/journal/Ausgabe_4/4_4.pdf [15.01.2022].
- Giest, Hartmut (2016): *Zur Didaktik des Sachunterrichts*. Berlin.
- Giest, Hartmut (2020): *Vorlesungen über Didaktik des Sachunterrichts. Ein Beitrag zur Konkretisierung kulturhistorischer Didaktik*. Berlin.
- Giest, Hartmut & Alberts, Sonja (2010): Explorieren und Experimentieren im Science Center: Ein Beitrag zur naturwissenschaftlichen Grundbildung im Primarbereich? In: Karl-Heinz Arnold, Kathrin Hauenschild, Britta Schmidt & Birgit Ziegenmeyer (Hrsg.): *Zwischen Fachdidaktik und Stufendidaktik*. Wiesbaden, S. 169-172.
- Giest, Hartmut & Marquardt-Mau, Brunhild (2013): Anschlussfähigkeit sichern – Übergänge gestalten. In: *Grundschulunterricht/Sachunterricht*, 2, S. 4-7.
- Graf, Bernhard & Noschka-Roos, Annette (2009): Stichwort Lernen im Museum oder: Eine Kamerafahrt mit der Besucherforschung. In: *Zeitschrift für Erziehungswissenschaft*, 12, 1, S. 7-27.

- Grygier, Patricia & Hartinger, Andreas (2009): Gute Aufgaben Sachunterricht. Naturwissenschaftliche Phänomene begreifen. Berlin.
- Guderian, Pascal (2007): Wirksamkeitsanalyse außerschulischer Lernorte. Berlin.
- Haider, Michael (2015): Physikalische Aspekte. In: Joachim Kahlert, Maria Fölling-Albers, Margarete Götz, Andreas Hartinger, Susanne Miller & Steffen Wittkowske (Hrsg.): Handbuch Didaktik des Sachunterrichts. 2. Aufl.. Bad Heilbrunn, S. 122-128.
- Hardy, Ilonca, Kleickmann, Thilo, Koerber, Susanne, Mayer, Daniela, Möller, Kornelia, Pollmeier, Judith, Schwippert, Knut & Sodian, Beate (2010): Die Modellierung naturwissenschaftlicher Kompetenz im Grundschulalter. In: Zeitschrift für Pädagogik, 56. Beiheft, S. 115-125.
- Hartinger, Andreas (2020): Experimente und Versuche. In: Dietmar von Reeken (Hrsg.): Handbuch Methoden im Sachunterricht. Band 3: Dimensionen des Sachunterrichts. Baltmannsweiler, S. 73-80.
- Itzek-Greulich, Heike, Flunger, Barbara, Vollmer, Christian, Nagengast, Benjamin, Rehm, Markus & Trautwein, Ulrich (2015): Effects of a science center outreach lab on school students' achievement – Are student lab visits needed when they teach what students can learn at school? In: Learning and Instruction, 38, pp. 43-52.
- Kalthoff, Britta, Theyßen, Heike & Schreiber, Nico (2016): Vergleich von expliziter und impliziter Instruktion im Experimentalpraktikum für Sachunterrichtsstudierende. In: Hartmut Giest, Thomas Goll & Andreas Hartinger (Hrsg.): Sachunterricht – zwischen Kompetenzorientierung, Persönlichkeitsentwicklung, Lebenswelt und Fachbezug. Bad Heilbrunn, S. 132-139.
- Kaltman, Gwen S. (2010): Hands-on learning. Washington.
- Kihm, Pascal & Peschel, Markus (2021): Demokratie lernen durch Experimentieren?! – Aushandlung eines selbstbestimmten Vorgehens beim Offenen Experimentieren im Sachunterricht. In: Toni Simon (Hrsg.): Demokratie im Sachunterricht – Sachunterricht in der Demokratie. Beiträge zum Verhältnis von Demokratie(lernen) und Sachunterricht(sdidaktik). Wiesbaden, S. 195-205.
- Koerber, Susanne, Mayer, Daniela, Osterhaus, Christopher, Schwippert, Knut & Sodian, Beate (2015): The development of scientific thinking in elementary school. A comprehensive inventory. In: Child Development, 86, 1, pp. 327-336.
- Koerber, Susanne, Sodian, Beate, Kropf, Nicola, Mayer, Daniela & Schwippert, Knut (2011): Die Entwicklung des wissenschaftlichen Denkens im Grundschulalter. Theorieverständnis, Experimentierstrategien, Dateninterpretation. In: Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 43, 1, S. 16-21.
- Köster, Hilde, Hellmich, Frank & Nordmeier, Volkhardt (Hrsg., 2010): Handbuch Experimentieren. Baltmannsweiler.
- Lenhard, Wolfgang & Lenhard, Alexandra (2016): Berechnung von Effektstärken. Dettelbach. <https://www.psychometrica.de/effektstaerke.html> [15.01.2022].
- Lewalter, Doris & Geyer, Claudia (2005): Evaluation von Schulklassenbesuchen im Museum. In: Zeitschrift für Pädagogik, 51, 6, S. 774-785.
- Lewalter, Doris & Geyer, Claudia (2009): Evaluation von Schulklassenbesuchen im Museum. In: Zeitschrift für Erziehungswissenschaft, 3, S. 775-785.
- Lucas, Keith B. (2000): One Teacher's Agenda for a Class Visit to an Interactive Science Center. In: Science Education, 84, pp. 524-544.
- Marquardt-Mau, Brunhild & Hoffmann, Yvonne (2009): Naturwissenschaften in altersgemischten Lernsituationen. In: Barbara Berthold & Heike Hahn (Hrsg.): Altersmischung als Lernressource – Impulse aus Fachdidaktik und Grundschulpädagogik. Baltmannsweiler, S. 268-284.
- Michalik, Kerstin (2010): Didaktische Konzepte für die naturwissenschaftliche Grundbildung von Kindern im Elementarbereich. In: Hans-Joachim Fischer, Peter Gansen & Kerstin Michalik (Hrsg.): Sachunterricht und frühe Bildung. Bad Heilbrunn, S. 93-108.
- Möller, Kornelia, Hardy, Ilonca, Jonen, Angela, Kleickmann, Thilo & Blumberg, Eva (2006): Naturwissenschaften in der Primarstufe – Zur Förderung konzeptuellen Verständnisses durch Unterricht und zur Wirksamkeit von Lehrerfortbildungen. In: Manfred Prenzel & Lars Allolio-Näcke (Hrsg.): Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms BiQua. Münster, S. 161-193.
- Murmann, Lydia, Pech, Detlef, Schomaker, Claudia & Stiller, Jurik (2019): „Aus der Perspektive von Kindern“ – Inwiefern kann der Forschungsansatz der Phänomenographie ein Impuls für die Beschreibung kindlicher Lernentwicklung und die Formulierung von Kompetenzniveaus sein? In: GDSU-Journal, Heft 9, S. 82-94.
- Oppermann, Elisa & Keller, Lena (2018). Geschlechtsunterschiede in der frühen MINT-Bildung – Forschungsüberblick. Berlin. www.haus-der-kleinen-forscher.de [15.01.2022]
- Osterhaus, Christoph, Koerber, Susanne & Sodian, Beate (2015): Children's understanding of experimental contrast and experimental control: an inventory for primary school. In: Frontline Learning Research, 3, 4, pp. 56-94.

- Pahl, Angelika & Lück, Gisela (2016): Naturwissenschaftliches Experimentieren – Kinder frühzeitig und kompetenzorientiert. In: Hartmut Giest, Thomas Goll & Andreas Hartinger (Hrsg.): Sachunterricht – zwischen Kompetenzorientierung, Persönlichkeitsentwicklung, Lebenswelt und Fachbezug. Bad Heilbrunn, S. 58-65.
- Peschel, Markus (2010): Grundschullabor für Offenes Experimentieren – Grundschultransfer? In: Hartmut Giest & Detlef Pech (Hrsg.): Anschlussfähige Bildung im Sachunterricht. Bad Heilbrunn, S. 49-56.
- Peschel, Markus (2015): Offenes Experimentieren – das Projekt SelfPro. In: Hans-Joachim Fischer, Hartmut Giest & Kerstin Michalik (Hrsg.): Bildung im und durch Sachunterricht. Bad Heilbrunn, S. 59-64.
- Priemer, Burkhard (2011): Was ist das Offene beim offenen Experimentieren? In: Zeitschrift für Didaktik der Naturwissenschaften, 17, S. 315-337.
- Ramseger, Jörg (2013): Prozessbezogene Qualitätskriterien für den naturwissenschaftlichen Unterricht – zehn Kriterien für wirksames didaktisches Handeln im Elementar- und Primarbereich. In: Stiftung Haus der kleinen Forscher (Hrsg.): Wissenschaftliche Untersuchungen zur Arbeit der Stiftung „Haus der kleinen Forscher“, Bd. 5. Schaffhausen, S. 147-171.
- Roth, Wolff-Michael, McRobbie, Campbell J., Lucas, Keith B. & Boutonné, Sylvie (1997): The Logical Production of Order in Traditional Science Laboratories: A Phenomenological Analysis. In: Learning and Instruction, 7, 2, pp. 107-136.
- Scharfenberg, Franz-Josef, Möller, Andrea, Kaufmann, Kathrin & Bogner, Franz X. (2019): Schülerlabore und Lehr-Lern-Labore. In: Jorge Groß, Marcus Hammann, Philipp Schmiemann & Jörg Zabel (Hrsg.): Biologiedidaktische Forschung: Erträge für die Praxis. Berlin, S. 229-249.
- Stiller, Cornelia & Wilde, Matthias (2021): Einfluss gestufter Lernhilfen als Unterstützungsmaßnahme beim Experimentieren auf den Lernerfolg im Biologieunterricht. In: Zeitschrift für Erziehungswissenschaft, 24, S. 743-763.
- Vosniadou, Stella (Hrsg., 2015): International Handbook of Research on Conceptual Change. 2. Aufl. New York u.a.
- Vygotskij, Lev Semjonowitsch (2002): Denken und Sprechen. Weinheim & Basel.
- Waldenmaier, Christine, Köster, Hilde & Müller, Bernhard (2013): Unterschiede bezüglich der Engagiertheit von Kindergruppen bei geöffneten und geschlossenen Experimentierangeboten im naturwissenschaftsbezogenen Sachunterricht. In: Hans-Joachim Fischer, Hartmut Giest & Detlef Pech (Hrsg.): Der Sachunterricht und seine Didaktik. Bad Heilbrunn, S. 137-146.
- Waldenmaier, Christine, Müller, Bernhard, Köster, Hilde & Körner, Hans-Dieter (2015): Engagiertheit und Motivation in unterschiedlichen Experimentiersituationen im Sachunterricht. In: Hans-Joachim Fischer, Hartmut Giest & Kerstin Michalik (Hrsg.): Bildung im und durch Sachunterricht. Bad Heilbrunn, S. 87-92.
- Waltner, Christine & Wiesner, Hartmut (2006): Physik lernen im Deutschen Museum – eine explorative Studie. Vortrag auf der GDCP-Jahrestagung 2006 in Bern.
- Wendt, Heike, Bos, Wilfried, Selter, Christoph, Köller, Olaf, Schwippert, Knut & Kasper, Daniel (Hrsg., 2016): TIMSS 2015: Mathematische und naturwissenschaftliche Kompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich. Münster.
- Windt, Anna, Scheuer, Rupert & Melle, Insa (2011): Methoden beim naturwissenschaftlichen Experimentieren im Elementarbereich. In: Hartmut Giest, Astrid Kaiser & Claudia Schomaker (Hrsg.): Sachunterricht – auf dem Weg zur Inklusion. Bad Heilbrunn, S. 157-162.
- Zadch, Mahsa Vali & Peschel, Markus (2018): SelfPro-Entwicklung von Selbstkonzepten beim Offenen Experimentieren. In: Ute Franz, Hartmut Giest, Andreas Hartinger, Anja Heinrich-Dönges & Bernd Reinthoffer (Hrsg.): Handeln im Sachunterricht. Bad Heilbrunn, S. 183-190.