

Experimentelle Kompetenz messen?!

Nico Schreiber*, Heike Theyßen*, Horst Schecker⁺

*Lehrstuhl für Didaktik der Physik, TU Dortmund,

⁺Institut für Didaktik der Naturwissenschaften, Abt. Physikdidaktik, Universität Bremen

Kurzfassung

Experimentelle Kompetenz als Teil des Kompetenzbereichs „Erkenntnisgewinnung“ soll gemäß der Bildungsstandards im Unterricht gefördert werden. Für die Diagnose experimenteller Kompetenz müssen daher auch geeignete Werkzeuge zur Verfügung stehen. Es ist zu bezweifeln, dass schriftliche Wissenstests valide genug sind, um die ganz Breite experimenteller Fähigkeiten, einschließlich der tatsächlichen Durchführung von Experimenten, zu erfassen. Experimentaltests mit Realexperimenten hingegen sind in Durchführung und Auswertung für den großflächigen Einsatz in Lernstandserhebungen oder PISA-Tests sehr aufwändig. Eine Alternative könnten Experimentaltests mit virtuellen Medien, wie Simulationsbaukästen für experimentelle Handlungssituationen, sein. Ob solche Verfahren zur Diagnose geeignet sind, soll durch das hier beschriebene Vorhaben überprüft werden. Den Erhebungen liegt ein Modell experimenteller Kompetenz zugrunde, welches durch Indikatoren operationalisiert ist.

Für die Kompetenztests wurden zwei experimentelle Aufgabenstellungen in je drei Testversionen (schriftlich, Realexperiment und Simulationsbaukasten) entwickelt. In einer Pilotstudie bearbeiten Schüler der 10. Klassenstufe die Aufgabenstellungen in jeweils unterschiedlichen Testversionen. Die Vorstudie diente der Erprobung und Optimierung der Übungs- und Begleiterhebungsmaterialien, der Kompetenztests, sowie der Auswertungsmethoden. Im Rahmen der Hauptstudie werden die mit den verschiedenen Testversionen erzielten Kompetenzzuweisungen auf Korrelationen hin untersucht. In diesem Artikel werden der theoretische Hintergrund, speziell das zugrunde gelegte Modell experimenteller Kompetenz, sowie das Untersuchungsdesign vorgestellt.

1. Motivation und Zielsetzung

Sowohl in den nationalen Bildungsstandards [1] als auch in daran anschließenden fachdidaktischen Kompetenzmodellen (Schecker & Parchmann [2]; Neumann et al. [3]) wird experimentelle Kompetenz explizit als eine Komponente naturwissenschaftlicher Kompetenz im Bereich „Erkenntnisse gewinnen“ oder „naturwissenschaftliches Arbeiten“ ausgewiesen. Neben Methoden zur Erhebung von Fachwissen werden also in zunehmendem Maße Methoden zur Erhebung experimenteller Kompetenz benötigt. Dies gilt sowohl für die Kompetenzdiagnostik auf Individualebene durch die einzelne Lehrperson als auch für das Assessment zur Bilanzierung von Lernergebnissen mit Schulleistungstudien. In Studien wie TIMSS und PISA wurden in Deutschland bisher experimentelle Kompetenzen nicht durch Experimentaltests erhoben, anders als z.B. in der Schweiz mit den TIMSS-Experimentiertests (vgl. [4]); oder im Projekt HarMoS, (vgl. [5]). Ein praktischer Grund ist der logistische Aufwand, die entsprechenden Versuchsmaterialien einheitlich bereit zu stellen. Deshalb wird oft versucht, die Experimentierfähigkeiten der Schülerinnen durch schriftliche Wissenstests zu messen. Wir gehen davon aus, dass diese Diagnostik nicht hinreichend valide ist, weil prozessbezogene Aspekte, die wesentlich für die experimentelle Vorgehensweise sind, nicht erfasst werden. Wünschenswert wäre eine alternative Erhebungsmethode, die

eine Erfassung der prozessbezogenen Aspekte zulässt, aber weniger aufwändig in Durchführung und Auswertung ist als Experimentaltests mit Realexperimenten.

Das zentrale Ziel des Projekts „Diagnostik experimenteller Kompetenz“ (eXkomp) ist daher die Entwicklung und Validierung von Testverfahren, die

- handlungs- und prozessbezogene Aspekte experimenteller Kompetenz erfassen,
- empirisch abgesichert und
- gleichzeitig potenziell großflächig einsetzbar sind.

Simulationsbaukästen (vgl. z.B. [6]), die vielfältige experimentelle Handlungsmöglichkeiten erlauben, könnten die Grundlage für eine solche Alternative sein.

Voraussetzung für die Entwicklung und Validierung eines solchen Messverfahrens ist ein ausdifferenziertes und operationalisiertes Konstrukt experimenteller Kompetenz. Auf dieser Basis kann überprüft werden, welche Erhebungsmethoden geeignet sind, um im Rahmen des theoretischen Konstrukts experimentelle Kompetenz zu messen.

2. Modell experimenteller Kompetenz

Experimentieren ist sehr vielfältig und es gibt verschiedene Vorstellungen, was unter dem Begriff des Experimentierens zu verstehen ist [7]. Damit verbunden ist die Frage, welche experimentelle Kompe-

tenz eigentlich notwendig ist, um erfolgreich experimentieren zu können. Wir wollen uns hier an der sehr spezifischen Sicht bereits bekannter Kompetenzmodelle aus der (fachdidaktischen) Forschung orientieren und die Situation des typischen selbstständigen Experimentierens, z.B. in der Schule, berücksichtigen.

Eine Grundlage für die Ausdifferenzierung der experimentellen Kompetenz stellt ein von Hammann [8] vorgeschlagenes Modell dar. Das Modell orientiert sich an der Darstellung des naturwissenschaftlichen Erkenntnisweges bei Klahr [9] und baut gleichzeitig auf empirischen Befunden auf. Hammann unterscheidet nach Klahr drei Teilkompetenzen: das Suchen und Aufstellen von Hypothesen, die Planung und Durchführung von Experimenten, sowie die Analyse der experimentellen Ergebnisse. Innerhalb dieser Bereiche differenziert er zwischen vier nach der Systematik der Vorgehensweise gestuften Kompetenzniveaus.

Ein ähnliches Modell mit vier Teilkompetenzen schlägt Mayer [10] vor. Es enthält als vierte Teilkompetenz die Formulierung naturwissenschaftlicher Fragestellungen. Die Durchführung von Experimenten wird mit schriftlichen Wissenstests als nicht erfassbar aufgefasst und in diesem Modell nicht als eigene Teilkompetenz berücksichtigt. Möller et al. [11] finden innerhalb der Teilkompetenzen fünf gestufte Kompetenzniveaus.

Im Unterschied zu diesen Arbeiten legen wir den Fokus auf die Durchführung von Experimenten als zentralen und fachbezogenen Teil des Experimentierens. Die der Durchführung vorausgehende Zielklärung und die Schlussfolgerungen aus den experimentellen Ergebnissen werden als Randbereiche mit erfasst und ergänzen den Dreischritt nach Hammann und Klahr.

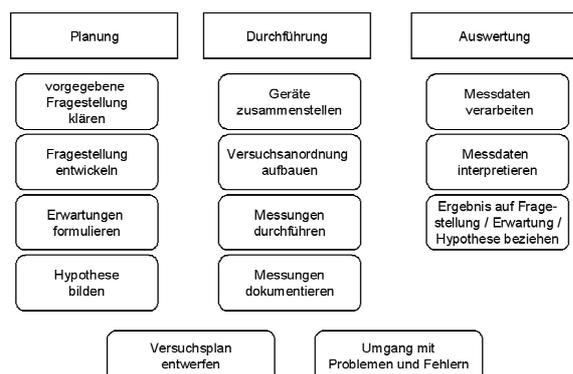


Abb. 1 Modell experimenteller Kompetenz

Abbildung 1 zeigt die drei übergeordneten Bereiche „Planung“, „Durchführung“ und „Auswertung“ mit den zugeordneten Merkmalen des Experimentierprozesses. Bereits an der Anzahl der Merkmale ist zu erkennen, dass der Schwerpunkt hier im Bereich der Durchführung liegt. Die drei Bereiche sowie die zwei nicht eindeutig zugeordneten Merkmale an den Schnittstellen werden im Folgenden erläutert.

Planung:

Wenn Forscher ein unbekanntes Phänomen entdecken, werden sie zunächst *Fragestellungen entwickeln*, die eine genauere Klärung bzw. Präparation des eigentlichen Phänomens ermöglichen. Erst nach intensiver Auseinandersetzung und auf Grundlage theoretischen Wissens ist die Formulierung begründeter Annahmen über Wirkungszusammenhänge („*Hypothesen*“) möglich. Bei im Grundsatz bekannten Phänomenen kommt es in der Wissenschaft oftmals gar nicht darauf an, Fragestellungen oder Hypothesen selbst zu entwickeln, sondern Aufträge mit gegebenen Fragestellungen und Hypothesen sorgfältig abzuarbeiten und gegebenenfalls begrenzt weiterzuentwickeln, z.B. hinsichtlich der experimentellen Verfahren. Dies gilt noch für viele Diplomarbeiten oder auch Promotionsvorhaben. Hier muss sich der Durchführende lediglich in die Fragestellungen einarbeiten und das Ziel der Untersuchung klären, um es zu einem eigenen Ziel machen zu können.

In Lehr-Lernsituationen wird der Generierung von Hypothesen und eigenen Fragestellungen oftmals in einer idealisierten Vorstellung von (normal-) wissenschaftlicher Forschung ein zu großer Stellenwert eingeräumt. In unserem Modell, das stärker als die Ansätze von Klahr, Hammann oder Mayer auf den Prozess des Experimentierens ausgerichtet ist, wird daher der Klärung vorgegebener Fragestellungen durch den Schüler eine mindestens ebenso große Bedeutung zugemessen wie der Entwicklung eigener Untersuchungsziele oder sogar Hypothesen. Erwartungen über den Ausgang eines Experiments sind allerdings notwendig, um „Fehler“ erkennen zu können (s.u.). Erwartungen sind jedoch niederschwelliger als theoretisch begründete Vorhersagen des Versuchsausgangs.

Zwischen der Planung und der Durchführung einer experimentellen Untersuchung steht im Modell zunächst der gedankliche und dann der reale Entwurf einer Versuchsanordnung und der Entwurf eines Versuchsplans. Sie sind als wichtige Schnittstelle beiden übergeordneten Bereichen des klassischen naturwissenschaftlichen Erkenntnisweges zu zuordnen. Ein wesentlicher Teilaspekt der Entwicklung eines Versuchsplans ist die Parameterkontrolle. Diese kann (und sollte) sowohl explizit in der Planung berücksichtigt werden, als auch implizit in der Durchführung.

Durchführung:

Wenn ein physikalisches Experiment durchgeführt werden soll, müssen — auf Basis des Umsetzungsplans — geeignete Geräte zusammengestellt sowie die Versuchsanordnung aufgebaut und getestet werden. Der Entwurf des Versuchsplans und die Schritte bis zu einem funktionstüchtigen Experiment bilden keine lineare Abfolge. Es handelt sich i.d.R. um spiralförmig durchlaufene (Teil-) Prozesse mit gegenseitigen Modifikationen. Schließlich werden die Messungen durchgeführt und dokumentiert.

Auswertung:

Wenn die Messungen abgeschlossen sind, müssen die Messdaten verarbeitet, interpretiert und schließlich auf die Fragestellung, Erwartung oder Hypothese bezogen werden.

Der Umgang mit Fehlern ist wieder eine Schnittstelle und sowohl der Durchführung als auch der Auswertung zuzuordnen. Hierzu gehört der Umgang mit nicht funktionierenden Aufbauten bei der Durchführung ebenso wie das Erkennen von fehlerhaften Messwerten, Interpretationen oder Ansätzen, die nicht zur Beantwortung der Fragestellung geeignet sind.

Um es noch einmal zu betonen: Die obige Darstellung soll nicht den Eindruck vermitteln, dass es eine strenge Reihenfolge innerhalb des Dreischritts gäbe. Das Modell geht vielmehr davon aus, dass alle Merkmale in einem Experimentierprozess in verschiedenen Reihenfolgen, ggf. iterativ und auch nicht immer vollständig auftreten können.

3. Testformate

3.1. Bislang übliche Testformate

Tests zur Erhebung experimenteller Kompetenz in den Naturwissenschaften liegen vor, greifen jedoch bislang fast ausschließlich auf zwei Verfahren zurück: Experimentaltests mit Realexperimenten und schriftliche Wissenstests. Bei den Experimentaltests ist zu unterscheiden, ob die Auswertung *prozessbezogen*, d.h. durch Handlungsanalysen, oder *produktbezogen*, z.B. anhand schriftlicher Arbeitsergebnisse erfolgt. Experimentaltests mit prozessbezogener Auswertung liegen von Walpuski [12] und Neumann [13] vor. Walpuski erstellt aus Videoanalysen Prozessgrafiken. Er orientiert sich wesentlich an einem Dreischritt (Idee, Experiment, Schlussfolgerung), der Parallelen zu Hammanns Strukturierung experimenteller Kompetenz enthält. Neumann filmt und analysiert kategoriegeleitet die Experimentaltests. Hierbei verwendet Neumann ein Kategoriensystem nach C. von Aufschnaiter [14], das zwischen explorativer, intuitiv regelbasierter und explizit regelbasierter Vorgehensweise unterscheidet. Diese Differenzierung für den Bereich der experimentellen Durchführung weist Parallelen zu der bei Hammann [8] vorgenommenen Unterscheidung zwischen systematischem und unsystematischem Umgang mit Variablen auf.

In den TIMSS-Experimentiertests [15] wurden einfache experimentelle Aufgaben von den Schülerinnen bearbeitet (z.B. Kraft-Dehnungs-Zusammenhang bei einem Gummiband). Es wurden nur die Ergebnisse („Produkte“) ausgewertet. Ergänzend zu schriftlichen Wissenstests werden die Experimentaltests bei „HarmoS“ [4] ebenfalls produktbezogen ausgewertet.

Mit Ausnahme von TIMSS und HarmoS erfolgt die Untersuchung großer Probandengruppen bisher in der Regel durch schriftliche Wissenstests. Zu diesen

werden hier auch solche Varianten gezählt, bei denen Experimente nicht schriftlich beschrieben, sondern vorgeführt werden und anschließend Fragen zu Vorgehensweise und Interpretation zu beantworten sind ([16]; vergleichbar auch mit Experimentalaufgaben in Abiturklausuren). Zur Erfassung naturwissenschaftlicher Arbeitsweisen mit schriftlichen Wissenstests wurde von Henke und Sumfleth [17] ein Testinstrument entwickelt und in zahlreichen Untersuchungen erprobt und weiterentwickelt (z.B. [18], [19]). Zur Erhöhung der Auswertungsobjektivität wurde der Test als Multiple-Choice-Test konzipiert. Um primär Kenntnisse über naturwissenschaftliche Arbeitsweisen zu erfassen, wird das Fachwissen ähnlich wie bei den PISA-Aufgaben [20] in der Aufgabe vorgegeben. Auch die bei „HarmoS“ eingesetzten schriftlichen Wissenstests (vgl. [4], S. 207) sind nach PISA Vorbild konstruiert.

Ein alternatives Testformat kann ein Simulationsbaukasten sein. Abbildung 2 zeigt einen Screenshot des Simulationsbaukastens für den Bereich der Elektrizitätslehre, der im Projekt eXkomp entwickelt wurde und eingesetzt wird.

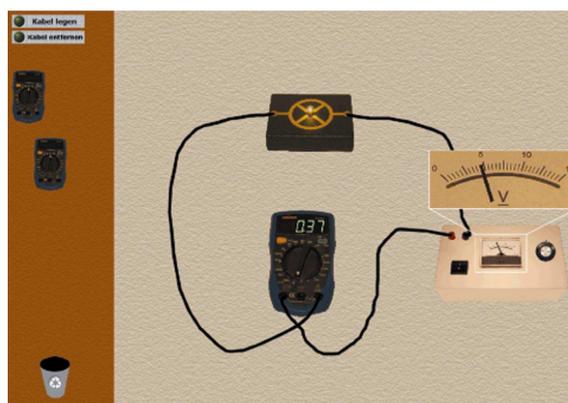


Abb. 2 Simulationsbaukasten zur Elektrizitätslehre

Das Experimentiermaterial des Simulationsbaukastens ist dem Material aus dem Realexperiment in Funktionsweise und Aussehen nachempfunden. Im Regal (dunkelbrauner Bereich) liegt, wie beim Realexperiment, das Experimentiermaterial bereit. Mit dem ausgewählten Material kann dann auf der weißen Arbeitsfläche experimentiert werden. Dabei muss z.B. die Spannungsquelle eingeschaltet und die gewünschte Spannung eingestellt werden. Die Spannungsanzeige wird bei Bedarf vergrößert dargestellt, um deutliches Ablesen zu ermöglichen (s. Abb. 2). Auch das Multimeter kann mit Kabeln versehen und der Messbereich eingestellt werden. Dabei wird der entsprechende Bereich am Multimeter vergrößert, um eine gute Ablesbarkeit zu ermöglichen. Kabel werden verlegt, wenn der Button „Kabel legen“ angeklickt wurde und die Maus vom gewünschten Startpunkt bei gedrückter Maustaste zum gewünschten Zielpunkt bewegt wird. Während der Bewegung wird das Kabel entlang der Mausspur verlegt (gezeichnet). Der Verlauf des Kabels auf der Arbeitsfläche kann anschließend (bei festen Anschlüssen)

noch mit der Maus verändert werden. Durch Anklicken des Buttons „Kabel löschen“ und anschließend eines Kabels wird dieses gelöscht. Geräte können auch durch falsche Handhabung kaputt gehen. Sie werden durch neue Geräte im Regal ersetzt, wenn ein defektes Gerät erkannt und in den Mülleimer (unten links) verschoben wurde.

3.2. Austauschbarkeit der Testformate

Shavelson et al. [21] reanalysieren empirische Daten von Baxter & Shavelson [22] mit dem Ziel, verschiedene Formen von Experimentaltests, darunter auch eine Version mit einem Simulationsbaukasten, und schriftliche Wissenstests hinsichtlich ihrer „Austauschbarkeit“ zur Erhebung experimenteller Fähigkeiten zu überprüfen. Dazu verglichen sie die erzielten Kompetenzzuweisungen aus Experimentaltests mit einem Simulationsbaukasten mit denen aus schriftlichen Wissenstests und Experimentaltests mit Realexperimenten, welche anhand von Stichwortprotokollen der SchülerInnen oder anhand direkter Beobachtung durch trainierte Rater bewertet worden sind. Das Kategoriensystem für die Experimentaltests ist, ähnlich wie bei den TIMSS Experimentiertests (vgl. [15]), produktbezogen und sehr grobstufig. Beispielsweise wird die Bearbeitung so genannter „Black-Box-Aufgaben“ zur Elektrizitätslehre nur dann als richtig gewertet, wenn der Inhalt richtig erkannt und die richtige Beschaltung für die Ermittlung des Inhaltes verwendet wurde. Teilerfolge oder die Schritte, die zur richtigen oder falschen Lösung führten, werden nicht bewertet. Die Bewertung anhand der Beobachtung des Realexperiments wird als Maßstab für die anderen Verfahren angesetzt. Sie weist, gemittelt über verschiedene Aufgaben, die höchste Korrelation mit der Bewertung von Realexperimenten anhand der Stichwortprotokolle ($r = 0,78$) auf, die geringste Korrelation mit den schriftlichen Wissenstests ($r = 0,46$) und ebenfalls eine unerwartet geringe Korrelation mit den Experimentaltests basierend auf Computersimulation ($r = 0,50$). Die geringe Korrelation mit den schriftlichen Wissenstests führen die Autoren darauf zurück, dass die SchülerInnen bei den Experimentaltests eine Rückmeldung aus dem Realexperiment bzw. der Computersimulation erhalten, die in schriftlichen Wissenstests fehlt. Bei den Experimentaltests von Shavelson et al. ist zu betonen, dass die Auswertung bei allen drei Varianten produktorientiert erfolgte. Als Fazit ihrer Studien halten sie fest: *“(a) direct observation, notebook, and computer simulation are equally exchangeable, (b) paper-and-pencil methods are not exchangeable for performance assessments, and (c) student performance over occasions is volatile.”* ([21], S. 70). Teil (c) bezieht sich darauf, dass die Korrelation zwischen den Bewertungen anhand direkter Beobachtung bei Realexperimenten und Protokollen von Realexperimenten basierend auf Testdurchführungen auf 0.52 sank, nachdem zu jeweils unterschiedlichen Zeitpunkten über mehrere Monate hinweg bei einer zufällig ausgewählten

Testgruppe die Tests durchgeführt worden sind. Damit konnte der Einfluss des Messzeitpunktes nachgewiesen werden.

3.3 Testformate für die Vergleichsstudie eXkomp

Vor dem Hintergrund des oben beschriebenen Modells experimenteller Kompetenz sollen in der Vergleichsstudie eXkomp drei verschiedene Testformate auf ihre Eignung zur Diagnose experimenteller Kompetenz überprüft werden: Experimentaltest mit Realexperiment, schriftlicher Test sowie Experimentaltest mit Simulationsbaukasten.

Experimentaltest mit Realexperiment

Bei diesem Testformat erhalten die Probanden eine experimentell lösbare Aufgabenstellung (z.B. die Erfassung der Kennlinie einer Glühlampe), zugehörige grundlegende Fachinformationen (Definition der Kennlinie), einen Pool an Experimentiermaterial sowie Protokollbögen.

Die Datenerfassung erfolgt über Videoaufzeichnungen der experimentellen Handlungen sowie über die Protokollbögen. Dieses Testformat wird, wie bei Shavelson et al., als „Referenz“ für die Kompetenzzuweisung angesetzt, weil es dem realen Experimentieren (bei vorgegebener Fragestellung) sehr nahe kommt.

Schriftlicher Test

Hier wird die gleiche experimentelle Fragestellung mit den gleichen Fachinformationen vorgegeben. Die Probanden müssen sich in zwei fiktive Schüler, Alina und Bodo, die diese Fragestellung experimentell bearbeiten sollen, hinein versetzen. Die Schritte sind durch Multiple-Choice Fragen, z.B. die Auswahl der richtigen Schaltung aus vier vorgegebenen Schaltungen, und einzelne offene Fragen stark vorkonstruiert.

Der Test wird html-basiert am PC bearbeitet. Die Eingaben werden bis auf eine von Hand anzufertigende Versuchsskizze direkt in einer Datenbank erfasst. Diese Variante bietet den Vorteil, dass an bestimmten Stellen die korrekten Zwischenergebnisse, wie z.B. die richtige Schaltung, eingeblendet werden können, damit die Probanden auf dieser Basis weiter arbeiten können (Vermeidung von Folgefehlern).

Experimentaltest mit Simulationsbaukasten

Analog zum ersten Testformat erhalten die Probanden die gleiche experimentell lösbare Aufgabenstellung, die Fachinformationen sowie den Protokollbogen. Anstelle des realen Experimentiermaterials wird nun ein Simulationsbaukasten (vgl. Abb. 2) eingesetzt, in dem die gleichen Geräte wie im Realexperiment und weitgehend die gleichen Handlungsmöglichkeiten zur Verfügung stehen.

Die Daten werden zunächst über eine Bildschirmaufzeichnung sowie die Protokollbögen erfasst. Im nächsten Schritt werden von dem Simulationsbaukasten Navigationsdaten erzeugt, die eine Rekonstruktion der für die Kompetenzdiagnostik relevanten

ten Handlungen ohne Bildschirmaufzeichnung ermöglichen sollen (siehe unten).

Die Tabelle in Abbildung 3 stellt die drei Testformate, deren charakteristische Merkmale sowie die Erhebungsmethoden gegenüber.

Realexperiment 	schriftlich 	Simulationsbaukasten 
• echtes Experimentieren	• in experimentelle Situation hineinversetzen	• Experimentieren innerhalb der Simulation
• Handlungsmöglichkeiten	• rein gedankliches Experimentieren	• Handlungsmöglichkeiten
• Rückkopplung	• Abwägen zwischen Distraktoren	• Rückkopplung
• Lösungsweg / Antwortformat: offen	• Lösungsweg / Antwortformat: geschlossen	• Lösungsweg / Antwortformat: offen
Erhebungsmethoden Video & Protokoll	Erhebungsmethoden html basiert & Papier	Erhebungsmethoden Bildschirmaufzeichnung & Protokoll, Logdaten

Abb. 3 Gegenüberstellung der Testformate

Die Testformate sollen an drei Kriterien gemessen werden (vgl. 1.):

- (i) Sie sollen *prozessbezogene* Aspekte erfassen,
- (ii) einen geringen logistischen Aufwand erfordern (Versuchsmaterialien) und
- (iii) möglichst einfach auswertbar sein.

Beim Realexperiment ist eine prozessbezogene Auswertung möglich, Material- und Analyseaufwand sind für den großflächigen Einsatz aber groß. Diesen Nachteil hat der schriftliche Test nicht, allerdings ist hier keine handlungsorientierte Auswertung möglich. Eine Lösung könnte der Simulationskasten sein. Wenn hier eine teilautomatisierte Auswertung mit Hilfe von Navigationsdaten [23] möglich wird, ist eine prozessbezogene Auswertung mit verhältnismäßig geringen Material- und Analyseaufwand möglich.

4. Hypothesen und Forschungsfrage

Schriftliche Wissenstests sind zwar eine sehr effiziente Erhebungsmethode, nach Shavelson et al. [18] jedoch aufgrund der fehlenden Rückkopplung aus Experiment bzw. Computersimulation nicht austauschbar gegen produktbezogen ausgewertete Experimentaltests. Daraus ergibt sich die erste Hypothese:

1. *Zwischen Kompetenzzuweisungen auf Basis von Papier & Bleistifttests und prozessbezogenen Kompetenzzuweisungen auf Basis von Realexperimenten besteht kein hinreichend großer Zusammenhang.*

Shavelson et al. erklären produktbezogen ausgewertete Experimentaltests mit Realexperiment bzw. Computersimulation für austauschbar. Experimentaltests mit virtuellen Medien werden bei uns gemäß der Konstruktion der Kompetenztests den Handlungsmöglichkeiten mit den Realexperimenten weitgehend entsprechen. Bei beiden Varianten können daher sehr ähnliche Handlungen durchgeführt und

von Ratern basierend auf detaillierten Aufzeichnungen analysiert werden.

2. *Zwischen Kompetenzzuweisungen auf Basis der Experimentaltests mit Simulationsbaukästen und Zuweisungen, die auf Experimentaltests mit Realexperimenten beruhen, besteht ein hinreichend großer Zusammenhang.*

Die Studien von Shavelson et al. zeigen zwar eine mittlere Korrelation zwischen der Bewertung von Experimentaltest anhand von Stichwortprotokollen bzw. direkter Beobachtung, jedoch sind die Beobachtungskriterien hier produktbezogen und nicht prozessbezogen. Es ist zu vermuten, dass produktbezogen ausgewertete Experimentaltests andere Aspekte experimenteller Kompetenz erfassen als prozessbezogen ausgewertete. Die Qualitäten der Wege, auf denen ein bestimmtes experimentelles Ergebnis gefunden wird, können sich deutlich unterscheiden. Die Analyse der Handlungsaufzeichnungen aus Experimentaltests sollte demnach andere Ergebnisse erbringen können als die Analyse der Stichwortprotokolle. Das führt zu folgender Hypothese:

3. *Bei Realexperiment und Simulationsbaukasten besteht zwischen Kompetenzzuweisungen auf Basis von Handlungsaufzeichnungen im Vergleich zu einer Zuweisung alleine auf Basis der Stichwortprotokolle kein hinreichend großer Zusammenhang.*

Zusätzlich wird eine Forschungsfrage formuliert, die notwendig ist, um die Validität der Navigationsdaten im Hinblick auf teilautomatisierte Auswertungen zu überprüfen. Sie lautet:

1. *Wie valide erfasst bei Experimentaltests mit Simulationsbaukästen die Auswertung detaillierter Navigationsdaten die gleichen Konstrukte wie die Auswertung der Bildschirmaufzeichnungen?*

5. Untersuchungsdesign

5.1 Zwei Sets von Kompetenztests

Die Hauptstudie (Frühjahr 2010) dient dem Vergleich der drei Testformate der Kompetenztests.

Die Kompetenztests wurden in allen drei Versionen (vgl. 3.3; schriftlich (P), Realexperiment (R), Simulationsbaukasten (S)) mit jeweils zwei Sets von Aufgabenstellungen entwickelt. Die Aufgabenstellungen der beiden Sets beziehen sich auf die Unterrichtsgegenstände der Elektrizitätslehre der Mittelstufe und haben grundsätzlich den gleichen Aufbau. Eine mögliche Aufgabenstellung wäre, die Kennlinie einer Glühlampe aufzunehmen. In Abbildung 4 ist die Aufgabenstellung abgebildet.

Experiment: Kennlinie einer Glühlampe

Information:

Die Eigenschaften einer Glühlampe werden durch eine Kennlinie beschrieben. Das ist ein Diagramm, bei dem auf der x-Achse die Stromstärke in der Glühlampe und auf der y-Achse die Spannung an der Glühlampe aufgetragen ist.

Deine Aufgabe:

Messe und zeichne die Kennlinie einer Glühlampe.

Abb. 4 Aufgabenbeispiel „Kennlinie einer Glühlampe“

Das notwendige fachliche Wissen wird bereits zu Beginn der Aufgabe bereitgestellt. Die Probanden klären zunächst die Fragestellung, entwickeln Ideen zur Durchführung, welche bei den experimentellen Varianten im Protokollbogen notiert werden sollen und führen das Experiment durch. Ein richtiger Versuchsaufbau im Realexperiment ist in Abb. 5, ein richtiger Versuchsaufbau im Simulationsbaukasten ist in Abb. 6 zu sehen.



Abb. 5 mögliche Schaltung zur Bestimmung der Kennlinie im Realexperiment

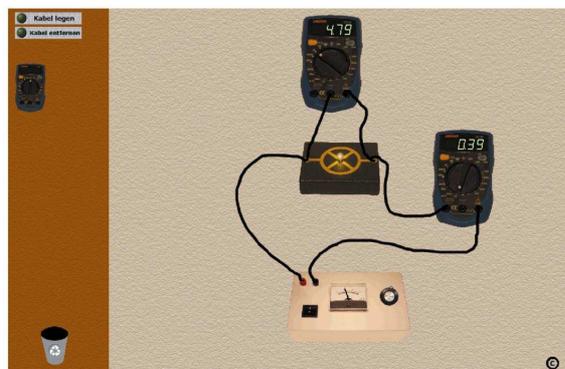


Abb. 6 mögliche Schaltung zur Bestimmung der Kennlinie mit Simulationsbaukasten

Es wird deutlich, dass sich die Aufbauten sehr ähnlich sind.

Der Protokollbogen ist weiterhin grob vorstrukturiert, indem Messwerte, Berechnungen, Ergebnisse und Begründungen verlangt werden.

Bei der rein schriftlichen Variante ist die Aufgabenstellung analog, aber noch stärker vorstrukturiert. Anstelle des Schaltungsaufbaus erfolgt hier beispielsweise die Abfrage der richtigen Schaltung (vgl. Abb. 7)



3) Alina und Bodo wollen nun ein Experiment durchführen um die Kennlinie einer Glühlampe zu bestimmen.

- a) Sie haben sich die folgenden vier möglichen Schaltungen überlegt. Welche können sie verwenden? (Denke daran, dass eine oder mehrere der vorgegebenen Antwortmöglichkeiten richtig sein können.)

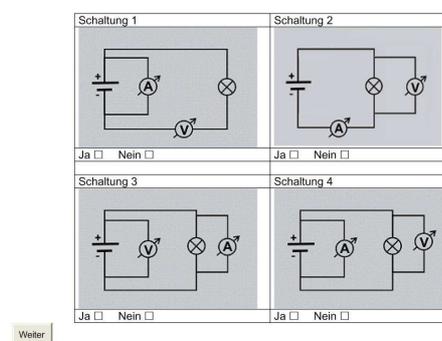


Abb. 7 Abfrage der richtigen Schaltung beim schriftlichen Test

5.2 Begleiterhebung und Übungsphase

Vor der Durchführung der Kompetenztests ist eine Begleiterhebung notwendig, um Personenparameter zu erfassen, von denen ein Zusammenhang mit experimenteller Kompetenz erwartet wird oder bekannt ist. Das sind kognitive Fähigkeiten [22], inhaltsbezogenes Fachwissen [24] sowie Aspekte des physikbezogenen Selbstkonzeptes, die sich speziell auf die Einschätzung der eigenen experimentellen Fähigkeiten beziehen. Außerdem werden Computerkenntnisse erfragt (Selbsteinschätzung). Zur Erfassung der kognitiven Fähigkeiten werden ausgewählte Subskalen des Kognitive- Fähigkeiten-Tests (KFT; [25]) eingesetzt. Da bei den Kompetenztests mit dem Simulationsbaukasten pseudodreidimensionale experimentelle Anordnungen in zweidimensionaler Repräsentation am PC zu bearbeiten sind, wird insbesondere das räumliche Vorstellungsvermögen berücksichtigt. Darüber hinaus werden mit einer weiteren Subskala die sprachlichen Fähigkeiten als Voraussetzung für das Verständnis der Aufgabenstellungen erfasst. Für den modifizierten Fachwissenstest werden Tests von v. Rhöneck [26] und Beichner & Engelhardt [27] zugrunde gelegt, die speziell auf bekannte Lernschwierigkeiten im Bereich „elektrischer Stromkreis“ eingehen. Das Selbstkonzept bezüglich experimenteller Fähigkeiten und die Selbsteinschätzung wird in einem modifizierten Fragebogen nach Brell [28] erhoben. Die Daten dienen der Randomisierung der Probandengruppen für die Kompetenztests.

Neben der Begleiterhebung ist eine Übungsphase notwendig, in der die zukünftigen Probanden mit dem Simulationsbaukasten und den Experimentiermaterialien vertraut gemacht werden.

5.3 Pilotstudien

Eine erste Pilotstudie dient der Erprobung und Optimierung der Übungsmaterialien, der Begleiterhebung, der Kompetenztests und der Auswertungsmethoden. Für die Datenerhebung (März/April 2009) waren ursprünglich 30 Probanden der Jahrgangsstufe 10 (Gymnasium) eingeplant. Da sich die Aufgabenstellungen aus dem Themenbereich „elektrische Stromkreise“ als unerwartet schwierig erwiesen, wurden zusätzlich Schülerinnen und Schüler der 12. Jahrgangsstufe einbezogen.

Die zweite Pilotstudie dient der weiteren Erprobung und Optimierung der Kompetenztests, sowie der Erprobung und Validierung der Verfahren zur Aufzeichnung und zur Auswertung der Navigationsdaten, die bei der Computersimulation erzeugt werden.

5.4 Hauptstudie

Das Design der Hauptstudie ist in Abbildung 8 schematisch dargestellt.

Schule	Begleiterhebung			
Didaktik-labor	Übungsphase			
	Tests experimenteller Kompetenz			
	Gruppe 1	Gruppe 2	Gruppe 3	Gruppe 4
Uni Bremen	R1	S1	R1	P1
	S2	R1	P2	R2

R: Realexperiment, P: schriftlich, S: Simulation

1: Aufgabenset 1, 2: Aufgabenset 2

Abb. 8 Design der Hauptstudie

Die Hauptstudie gliedert sich in drei Phasen: die Begleiterhebung an der Schule und die Übungsphase sowie die Durchführung der Kompetenztests im Physikdidaktiklabor der Universität Bremen. Die Begleiterhebung (5.2) wird im Rahmen einer Schulstunde in Klassen- bzw. Kursverbänden mit insgesamt ca. 180 Probanden durchgeführt. Der Überhang von 50% gegenüber der Probandenzahl der Kompetenztests ermöglicht es, Probanden mit extremen Personenparametern bei der Zusammenstellung der Testgruppen auszuschließen und Ausfälle, z.B. durch Krankheit oder Klassenwechsel, zu kompensieren.

Die Übungsphase zur Einführung der Medien und Arbeitsformen wird ca. zwei Wochen vor der Durchführung der Kompetenztests im Didaktiklabor durchgeführt.

Anhand der erhobenen Personenparameter werden für die Kompetenztests vier Gruppen mit je ca. 30 Probanden randomisiert zusammengestellt. Die Einteilung erfolgt auf Basis der Individuen. Gleichzeitig ist sicherzustellen, dass jede Klasse auf alle vier Gruppen aufgeteilt wird

Die Kompetenztests werden mit beiden zuvor erprobten Aufgabensets (5.1) durchgeführt. Sie finden im Didaktiklabor statt, wo die für die Experimentaltests notwendigen apparativen Voraussetzungen in hinreichender Anzahl zur Verfügung stehen (Experimentiermaterial, PC-Arbeitsplätze, Kameras). Um den Einfluss der Umgebung für alle Probanden gleich zu halten, finden auch die schriftlichen Wissenstests dort statt.

Da der Experimentaltest mit Realexperiment (R) als Referenz für die Qualität der Testformate gewählt wurde, bearbeitet jeder Proband eine der beiden Aufgabenstellungen in diesem Testformat (R) und die zweite Aufgabenstellung in einem der anderen beiden Testformate (S oder P). Aufgrund des von Shavelson et al. nachgewiesenen Einflusses des Messzeitpunktes wird zusätzlich ein Kreuzdesign gewählt (vgl. Abb. 8).

5.5 Teilstudien zur Auswertung

Zur Überprüfung der Hypothesen und Beantwortung der Forschungsfrage werden im Rahmen der Pilotstudien und der Hauptstudie entsprechende Teilstudien durchgeführt.

1. Zur Überprüfung der ersten Hypothese wird für alle Probanden der Gruppen drei und vier anhand der schriftlichen Wissenstests zum Experimentieren eine Kompetenzzuweisung getroffen. Eine weitere Kompetenzzuweisung wird anhand der Analysen der Handlungsaufzeichnungen beim Realexperiment getroffen. Die damit erzielten Kompetenzzuweisungen werden auf Zusammenhänge untersucht.
2. Zur Überprüfung der zweiten Hypothese werden für alle Probanden der Gruppen eins und zwei anhand der Analysen der Handlungsaufzeichnungen für jede Testvariante des Realexperiments und des Simulationsbaukastens Kompetenzzuweisungen getroffen. Diese werden auf Zusammenhänge untersucht.
3. Zur Überprüfung der dritten Hypothese wird zunächst für alle Probanden für alle Testdurchführungen mit Realexperiment oder Simulationsbaukasten anhand der Stichwortprotokolle eine Kompetenzzuweisung getroffen. Basierend auf der Handlungsaufzeichnung wird eine weitere Kompetenzzuweisung getroffen. Beide Zuweisungen werden auf Zusammenhänge untersucht.
4. Zur Beantwortung der Forschungsfrage werden für alle Probanden, die Kompetenztests mit dem Simulationsbaukasten bearbeitet haben, zum einen anhand der Bildschirmaufzeichnungen und zum anderen anhand der Navigationsdaten Kompetenzzuweisungen getroffen. Die mit den verschiedenen Verfahren erzielten Kompetenzzuweisungen werden auf Zusammenhänge untersucht.

6. Literatur

- [1] KMK, Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (Hrsg.) (2004). Bildungsstan-

- dards im Fach Physik (Chemie & Biologie) für den Mittleren Schulabschluss. München: Luchterhand.
- [2] Schecker, H., Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften* 12 (2006), 45-66.
- [3] Neumann, K., Kauertz, A., Lau, A., Notarp, H. & Fischer, H. E. (2007). Die Modellierung physikalischer Kompetenz und ihrer Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 125-143.
- [4] Labudde, P. & Stebler, R. (1999): Lern- und Prüfungsaufgaben für den Physikunterricht. Erträge aus dem TIMSS-Experimentiertest. In: *Unterricht Physik* 10 (1999) Nr. 54, 23-31.
- [5] Zeyer, A., Adamina, M., Gingins, F., Labudde, P. (2007). HarmoS Naturwissenschaften – Entwicklung, Umsetzung und Assessment von Standards im naturwissenschaftlichen Unterricht der Schweiz. In Höttecke, D. (Hrsg.), *Kompetenzen, Kompetenzmodelle, Kompetenzentwicklung*. Münster: LIT, 206-208
- [6] Goldberg, F., Bendall, S. (1995). Making the invisible visible: A teaching and learning environment that builds on a new view of the physics learner. *American Journal of Physics* 63, 11, 978-991. (Siehe auch URL <http://cpucips.sdsu.edu/web/CPU/>)
- [7] Höttecke, D. (2007). Fachliche Klärung des Experimentierens. In Höttecke, D. (Hrsg.), *Kompetenzen, Kompetenzmodelle, Kompetenzentwicklung*. Münster: LIT, 293-295
- [8] Hammann, M. (2004). Kompetenzentwicklungsmodelle: Merkmale und ihre Bedeutung – dargestellt anhand von Kompetenzen beim Experimentieren. *Der mathematischer und naturwissenschaftliche Unterricht* 57, Nr. 4, 196-203.
- [9] Klahr, D. (2000). *Exploring Science. The cognition and development of discovery processes*. Massachusetts.
- [10] Mayer, J. (2007). Erkenntnisgewinnung als wissenschaftliches Problemlösen. In D. Krüger, H. Vogt (Hrsg.), *Theorien in der biologiedidaktischen Forschung*, Berlin: Springer.
- [11] Möller, A., Grube, C. & Mayer J. (2007). Kompetenzniveaus der Erkenntnisgewinnung bei Schülerinnen & Schülern der Sekundarstufe I. In Bayrhuber et. al. (Hrsg.), *Ausbildung und Professionalisierung von Lehrkräften, Internationale Tagung der Fachgruppe Biologiedidaktik im Verband Biologie, Biowissenschaften & Biomedizin*; Essen.
- [12] Walpuski, M. (2006). Optimierung von Kleingruppenarbeit durch Strukturierungshilfen und Feedback. Band 49. In H. Niedderer, H. Fischler, E. Sumfleth (Hrsg), *Studien zum Physik- und Chemielernen*. Berlin: Logos.
- [13] Neumann, K. (2004). Didaktische Rekonstruktion eines physikalischen Praktikums für Physiker. In H. Niedderer, H. Fischler (Hrsg.), *Studien zum Physiklernen*, Bd. 38, Berlin: Logos.
- [14] Aufschnaiter, C. von (2002). Ich weiß was rauskommt, aber ich kann es nicht erklären. In A. Pitton (Hrsg.), *Außerschulisches Lernen in Physik und Chemie* (S. 114–116). Münster: LIT.
- [15] Stebler, R., Reusser, K., Ramseier, E. (1998). Praktische Anwendungsaufgaben zur integrierten Förderung formaler und materialer Kompetenzen - Erträge aus dem TIMSS Experimentiertest. *Bildungsforschung und Bildungspraxis*, 20(1), 28-54.
- [16] Berge, O. E., Volkmer, M. (2002). Lernerfolgskontrolle mit Experimenten – Didaktische Aspekte. *Naturwissenschaften im Unterricht - Physik*, 71/72/2002, 206-210.
- [17] Henke, C. und Sumfleth, E. (2005). Leistungsmessung in der Oberstufe mit chemischen Experimentalaufgaben. In A. Pitton (Hrsg.), *Lehren und Lernen mit neuen Medien*. Münster: LIT, 340-342.
- [18] Klos, S. & Sumfleth, E. (2006). Einfluss des Fachs Naturwissenschaft auf Lernerfolg und Interesse. In A. Pitton (Hrsg.), *Lehren und Lernen mit neuen Medien. Gesellschaft für Didaktik der Chemie und Physik*. Münster: LIT, 343-345.
- [19] Ruman, S. (2005). *Kooperatives Experimentieren im Chemieunterricht. Entwicklung und Evaluation einer Interventionsstudie zur Säure-Base-Thematik*. Berlin: Logos.
- [20] Prenzel, M., Baumert, J. Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrum, R., Rolff, H.-G., Rost, J., Schiefele, U. (2004). *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland - Ergebnisse des zweiten internationalen Vergleichs*. Münster: Waxmann.
- [21] Shavelson, R. J., Ruiz-Primo, M. A., Wiley, E. W. (1999). Note on Sources of Sampling Variability in Science Performance Assessments. *Journal of Educational Measurement*, 36 (1), 61-71.
- [22] Baxter, G. P., Shavelson, R. J. (1994). Science Performance Assessments: Benchmarks and Surrogates. *International Journal of Educational Research*, 21, 233-350.
- [23] Zastrow, M. (2001). Interaktive Experimentieranleitungen. Entwicklung und Evaluation eines Konzeptes zur Vorbereitung auf das Experimentieren mit Messgeräten im Physikalischen Praktikum. In H. Niedderer, H. Fischler (Hrsg.), *Studien zum Physiklernen*, Band 18. Berlin: Logos.

-
- [24] Henke, C. (2007). Experimentell – naturwissenschaftliche Arbeitsweisen in der Oberstufe. In H. Niedderer, H. Fischler (Hrsg.), Studien zum Physiklernen, Bd. 59, Berlin: Logos.
- [25] Heller, K. A. & Perleth, C. (2000). Kognitiver Fähigkeitstest für 4.-12. Klassen, Revision (KFT 4-12+ R). Göttingen: Hogrefe.
- [26] Rhöneck, C. v. (1988): Aufgaben zum Spannungsbegriff. Naturwissenschaften im Unterricht - Physik/Chemie 36 Nr. 31, S. 38-41.
- [27] Engelhardt, P.V. & Beichner, R.J. (2004): DIRECT - Determining and Interpreting Resistive Electric Circuits Concepts Test. In: American Journal of Physics, Vol. 72, 1, 98-115.
- [28] Brell, C. (2008). Lernmedien und Lernerfolg – reale und virtuelle Materialien im Physikunterricht. In H. Niedderer, H. Fischler (Hrsg.), Studien zum Physiklernen, Bd. 38, Berlin: Logos.