

Auswertungsmethodik im Projekt „Diagnostik experimenteller Kompetenz“

Nico Schreiber*, Heike Theyßen*, Horst Schecker[†]

*Universität Duisburg-Essen, Didaktik der Physik, Universitätsstr. 2, 45141 Essen, nico.schreiber@uni-due.de und heike.theysen@uni-due.de; [†]Universität Bremen, Didaktik der Physik, Otto-Hahn-Allee 1, 28334 Bremen, schecker@physik.uni-bremen.de

Kurzfassung

Im Forschungsvorhaben „Diagnostik experimenteller Kompetenz“ (eXkomp) werden mit verschiedenen Testverfahren (Realexperiment, Computersimulation, schriftliches Verfahren) experimentelle naturwissenschaftliche Fähigkeiten gemessen und die Ergebnisse aus den drei Verfahren miteinander verglichen. Für die Daten aus den Testverfahren, bei denen die Probanden mit Realexperimenten bzw. Computersimulationsbaukästen arbeiten, wurden zwei neue Auswertungsmethoden entwickelt. Bei der Beurteilung des experimentellen Prozesses fokussiert die eine Methode auf Kriterien der fachlichen Richtigkeit von Zwischenschritten einer Handlungssequenz beim Experimentieren, die andere auf Kriterien der Arbeitsqualität. Validiert wurden beide Methoden durch ein hoch inferentes Expertenrating. In diesem Beitrag werden die Auswertungsmethoden und die Ergebnisse der Validierung vorgestellt.

1. Rahmenbedingungen

Ziel des DFG-Projekts „Diagnostik experimenteller Kompetenz“ (eXkomp) ist die Entwicklung und Validierung von Testverfahren, die handlungs- und prozessbezogene Aspekte experimenteller Kompetenz erfassen und gleichzeitig großflächig einsetzbar sind [1]. Neben schriftlichen Tests werden hierzu Experimentaltests mit Realexperimenten und Simulationsbaukästen eingesetzt, die prozessbezogen ausgewertet werden. Auf diese Testverfahren und die hierfür entwickelten Auswertungsmethoden bezieht sich der vorliegende Beitrag.

Um eine prozessorientierte Kompetenzzuweisung zu erreichen, ist eine Beurteilung des experimentellen Vorgehens vorzunehmen. Dafür stehen Videoaufzeichnungen, Navigationsdaten (nur bei dem Simulationsbaukasten) und Protokollbögen zur Verfügung. Es liegen keine verbalen Informationen vor, weil die Schülerinnen und Schüler jeweils alleine experimentieren. „Lautes Denken“ wurde nicht eingesetzt, weil es untypisch für das selbstständige Experimentieren ist und zu einer zusätzlichen kognitiven Belastung führen könnte. Im Folgenden werden verschiedene Methoden zur Auswertung der Prozessdaten vorgestellt.

2. Handlungen im experimentellen Prozess

Ausgangspunkt für die Entwicklung der Auswertungsmethoden sind beobachtbare Handlungen im experimentellen Prozess, die in Anlehnung an ein normativ entwickeltes Modell experimenteller Kompetenz (Abb. 1) eingeordnet werden können (vgl. [1] und [2]).

Mit den verwendeten Aufgabenstellungen werden die in Abbildung 1 fett umrandeten Modellkomponenten abgedeckt. Um die Komponente „mit Prob-

lemen und Fehlern umgehen“ zu untersuchen, wurden zusätzliche Testaufgaben konzipiert, bei der geplant Fehler in die Versuchsanordnung eingebaut wurden.

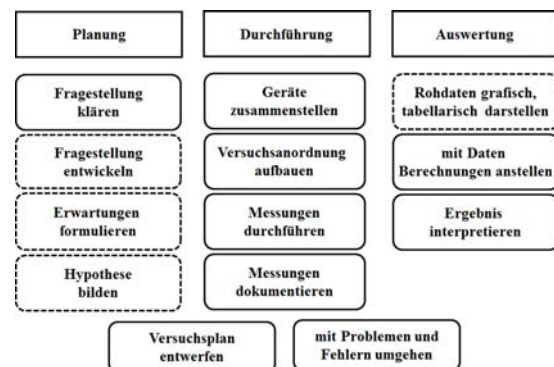


Abb. 1: Modell experimenteller Kompetenz

Handlungsbilder visualisieren den experimentellen Verlauf (Abb. 2). Dazu werden die Handlungskomponenten auf der Ordinate und die Zeit auf der Abszisse aufgetragen. Die Kodierung der Handlungen erfolgt in Intervallen von fünf Sekunden anhand eines ausführlichen Kodiermanuals. Die Doppelkodierung wird von zwei Ratern, die durch Beispiele geschult sind, durchgeführt. Die Inter-Rater-Reliabilität (Cohens Kappa) der Handlungskategorisierung beträgt beim Realexperiment 0.83 und bei der Computersimulation 0.87 (errechnet über die Kategorisierungen von fünf bzw. zwei vollständigen Versuchsdurchführungen).

Auf Grundlage von Daten aus zwei Pilotstudien, an denen 75 Bremer Oberstufenschülerinnen und -schüler teilgenommen haben, zeigt sich, dass die Handlungskomponenten sinnvoll zu Aufbau-, Mess-

und Auswertungsvorgängen zusammengefasst werden können, so wie es in Abbildung 2 angedeutet wird.

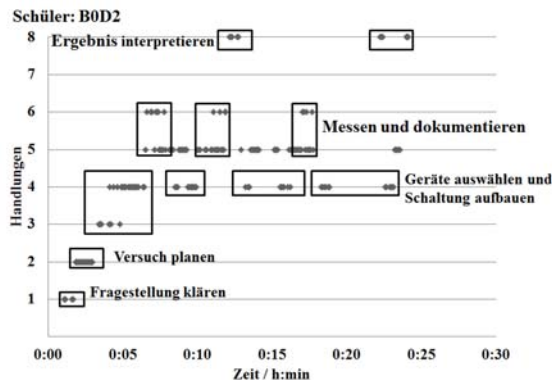
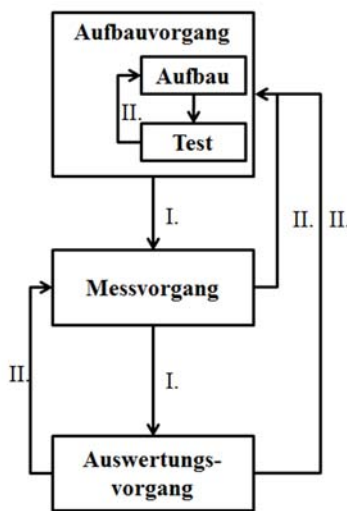


Abb. 2: Handlungsbild eines Schülers

Das Flussdiagramm in Abbildung 3 stellt grafisch den Ablauf des Experimentierens dar, wie er sich aus der Analyse der Schülerhandlungen ergibt.

Experimenteller Prozess



- I. Vorheriger Vorgang wird angenommen.**
- II. Vorheriger Vorgang wird abgebrochen.**

Abb. 3: Experimenteller Prozess im Flussdiagramm

Der experimentelle Prozess beginnt mit einem Aufbauvorgang. Dieser besteht aus den Komponenten „Geräte auswählen“ und „Versuchsanordnung aufbauen“ sowie einem Test der Versuchsanordnung. Der „Test einer Versuchsanordnung“ wird nicht explizit im Modell genannt, ist aber im Sinne einer Probemessung in der Komponente „Messungen durchführen“ enthalten. Das Testen der Versuchsanordnung geht fließend in das Messen im Sinne des systematischen Generierens von Daten über, wenn die ermittelten Daten vom Durchführenden angenommen und dokumentiert werden (I). Der Aufbauvorgang gilt ab dann als abgeschlossen, bis gegebenenfalls ein neuer Aufbauvorgang einen bereits

begonnenen Messvorgang beendet. Wenn die Messergebnisse aus dem Testen der Versuchsanordnung nicht dokumentiert werden, folgt in der Regel ein Umbau der Versuchsanordnung (II).

Ein Messvorgang besteht aus den Komponenten, „Messungen durchführen“ und „Messungen dokumentieren“. Werden die Messdaten nicht akzeptiert (d.h. nicht in die Auswertung einbezogen), gilt der Messvorgang als abgebrochen (II) und die Versuchsanordnung wird umgebaut (neuer Aufbauvorgang). Wenn die Messdaten akzeptiert werden (I), folgt ein Auswertungsvorgang, der aus den Komponenten „mit Daten Berechnungen anstellen“ und „Ergebnis interpretieren“ besteht. Sollte ein Auswertungsvorgang abgebrochen werden, folgt darauf entweder ein Aufbau- oder ein Messvorgang (II). Die Handlungskomponenten, „Fragestellung klären“ und „Versuchsplanung“ finden sich sowohl am Beginn der experimentellen Untersuchung als auch während oder ganz am Ende des Gesamtprozesses. Gedanklich werden die Klärung und die Planung vermutlich vor dem Aufbau durchgeführt oder zumindest begonnen; dies ist jedoch nicht beobachtbar.

Die Handlungskomponenten und deren Zusammenfassung zu Aufbau- und Messvorgängen stellen die Grundlage der weiteren Auswertungsschritte dar.

3. Handlungssequenzanalyse

Die Handlungssequenzanalyse ist eine niedrig inferente Beurteilung des experimentellen Prozesses unter fachlichen Gesichtspunkten. Es handelt sich dabei um ein niedrig inferentes Verfahren, weil die Beurteilungskriterien eng umrissen und der Beobachtung gut zugänglich sind. Die Handlungssequenzanalyse erfolgt getrennt für Aufbau- und Messvorgänge. Dabei wird jeder abgeschlossene Vorgang anhand eines Kodiermanuals beurteilt. Es kommt durchaus vor, dass mehrere Aufbau- und Messvorgänge auftreten. Durch die Beurteilungen wird für die Aufbauphase (= Gesamtheit aller Aufbauvorgänge) und die Messphase (= Gesamtheit aller Messvorgänge) eine Kennzahl ermittelt.

1. Stufe	2. Stufe	3. Stufe	Kennzahl
ein Aufbauvorgang	richtig	richtig	7
		fehlerhaft	2
	fehlerbehaftet	falsch	1
mehrere Aufbauvorgänge	Verbesserung	richtig	6
		fehlerhaft	5
	keine Verbesserung	richtig	7
fehlerhaft		4	
falsch		3	

Abb. 4: Analyseschema zur Handlungssequenzanalyse

Die Vorgehensweise wird hier am Beispiel der Aufbauphase erläutert. In Abbildung 4 ist das dreistufige Einordnungsschema für die Aufbauphase dargestellt. Die erste Stufe differenziert danach, ob ein Aufbau-

vorgang oder mehrere Aufbauvorgänge auftreten. Sollte nur ein Vorgang auftreten, kann dieser fachlich korrekt oder mit einem Fehler behaftet sein (2. Stufe). Wenn mehrere Aufbauvorgänge in einem experimentellen Prozess auftauchen, besteht die Möglichkeit der Verbesserung (2. Stufe). Anschließend (3. Stufe) wird für den *letzten auftretenden Vorgang* die fachliche Beurteilung (richtig, fehlerhaft oder falsch) vorgenommen. Bei einem Rating nach dem Flussdiagramm in Abbildung 4 wird also eine Folge von fachlichen Beurteilungen vorgenommen. Der letzten Stufe im Flussdiagramm wird eine Kennzahl für die Beurteilung der experimentellen Kompetenz bezüglich des Aufbaus zugeordnet. Die Kennzahl „1“ steht für die schlechteste Beurteilung, die Kennzahl „7“ für die beste Beurteilung. Dabei können Kennzahlen doppelt vergeben werden, sofern die Handlungssequenzen als gleichwertig betrachtet werden. Dieser Reihenfolge liegt eine normative Setzung zu Grunde, die durch eine Expertenbefragung zurzeit überprüft wird. In der aktuellen Version wird beispielsweise positiv gewertet, wenn ein Versuch unternommen wurde, das experimentelle Handeln zu verbessern.

4. Arbeitsqualitätsanalyse

Bei der Arbeitsqualitätsanalyse handelt es sich um eine hoch inferente Beurteilung des experimentellen Prozesses anhand von Kategorien, die die Arbeitsqualität beschreiben. Aus vorliegenden Videos wurden elf Kategorien explorativ generiert. Die Qualitätsaspekte sind in Tabelle 1 und 2 mit den möglichen Beurteilungskategorien aufgelistet.

Die Kategorien „Sicherheitsmaßnahmen beachtet“, „vorbereitete und nachvollziehbare Dokumentation der Messergebnisse“, ein „aufgeräumter Arbeitsplatz“ und das „effektive Arbeiten“ (entspricht „Vorgehen beim Aufbau“) finden sich auch bei [3].

Die prozessorientierte Beurteilung der experimentellen Handlungen orientiert sich an einem Ansatz von Kulgemeyer [4]. Kulgemeyer erzeugt einen heuristischen Index zur Rangbildung von Probanden hinsichtlich ihrer Fähigkeit physikalisch zu kommunizieren, indem er die Einzelkennwerte für Beurteilungskategorien aufsummiert.

Der Arbeitsqualitätsindex im eXkomp Projekt besteht für die Aufbauphase aus den in Tabelle 1 angeführten Aspekten. Für jeden Aspekt sind Beurteilungskategorien vorgegeben, die mit einer Wertung verbunden sind. Der Arbeitsqualitätsindex für die Aufbauphase ist dann die Summe der Wertungen der Arbeitsqualitätsaspekte.

Der Arbeitsqualitätsindex für die Messphase wird analog zu dem in der Aufbauphase ermittelt. Er ist dann die Summe der Wertungen der Aspekte, die in Tabelle 2 aufgeführt sind.

Arbeitsqualitätsanalyse in der Aufbauphase		
Aspekte der Arbeitsqualität	Beurteilungskategorien	Wertung
versierter Umgang mit Experimentiermaterial?	• ja	1
	• nein	-1
Sicherheitsmaßnahmen beim Aufbau beachtet?	• eher ja	1
	• eher nicht	0
	• nein	-1
Topologie der Schaltung übersichtlich?	• ja	1
	• nein	-1
aufgeräumter Experimentiertisch?	• ja	1
	• nein	-1
Vorgehen beim Aufbauen?	• zielstrebig, schnell und planvoll	1
	• bedächtig und planvoll	0
	• chaotisch	-1
Aufgabenstellung verinnerlicht?	• ja	1
	• nein	-1

Tab. 1: Aspekte der Arbeitsqualität, Beurteilungskategorien und Wertungen für die Aufbauphase

Arbeitsqualitätsanalyse in der Messphase		
Aspekte der Arbeitsqualität	Beurteilungskategorien	Wertung
Sicherheitsmaßnahmen beim Messen beachtet?	• immer	1
	• nicht immer	0
	• nie	-1
Messprotokoll vorbereitet?	• ja	1
	• nein	-1
Messen erfolgt in sinnvoller Reihenfolge?	• ja	1
	• nein	-1
Dokumentation nachvollziehbar?	• ja	1
	• nein	-1
Kontrollmessungen durchgeführt?	• ja	1
	• nein	-1

Tab. 2: Aspekte der Arbeitsqualität, Beurteilungskategorien und Wertungen für die Messphase

5. Überprüfung der Methoden

Dickmann [5] vergleicht in seiner Studie die beiden Auswertungsmethoden jeweils mit hoch inferenten Einschätzungen durch zwei Rater. Beide Rater sind mit dem Modell experimenteller Kompetenz vertraut, kennen aber die Schemata der formalisierten Handlungssequenzanalyse (Abb. 4) und der Arbeitsqualitätsanalyse (Tab. 1 und 2) nicht. Für ihre Einschätzungen der Handlungsabfolgen und der Arbeitsqualität betrachten beide Rater unabhängig voneinander Paare von Probanden (zwei Videos von Durchführungen mit Realexperimenten). Jeder Rater

entscheidet auf Basis des Modells pauschal, welche der beiden experimentellen Vorgehensweisen als besser bezeichnet werden kann. Zum Vergleich werden die mit den entsprechenden formalisierten Methoden ermittelten Ränge herangezogen.

In der Studie von Dickmann [5] werden 18 Paarvergleiche berücksichtigt. Bei 17 der 18 Paarvergleiche stimmen beide Rater überein, was zu einer relativen Übereinstimmung von 94%¹ führt. Nach diesem Ergebnis scheint ein Expertenrating mit Paarvergleichen eine geeignete Möglichkeit zu sein, experimentelle Fähigkeiten einzuschätzen. Ein Nachteil ist, besonders bei großen Stichproben, der hohe Durchführungsaufwand.

Die Übereinstimmung der Ergebnisse der formalisierten Handlungssequenzanalyse mit der Einschätzung von Rater 1 beträgt 88%, mit Rater 2 94%. Rater 1 stimmt mit den Einschätzungen der Arbeitsqualitätsanalyse in 72% der Fälle überein, Rater 2 in 67%. Für die Intercoder-Reliabilität, die über beide Rater und die Ergebnisse der Handlungssequenzanalyse berechnet wurde, ergibt sich in dieser Studie ein Cronbachs α von 0.87. Cronbachs α für die Intercoderreliabilität zwischen beiden Ratern und den Einschätzungen der Arbeitsqualitätsanalyse beträgt 0.72. Die Ergebnisse sind insgesamt zufriedenstellend, besonders für ein hoch inferentes Rating. Ein wichtiger Grund für die schlechtere Übereinstimmung mit der Beurteilung der Arbeitsqualitätsanalyse ist nach [5], dass den Ratern eine strikte Trennung zwischen fachlicher Korrektheit und Aspekten der Arbeitsqualität nicht gelingt.

Den Ergebnissen von Dickmann [5] zufolge sind beide Auswertungsmethoden ein geeignetes ordinales Maß, um experimentelle Fähigkeiten zu messen. Die Handlungssequenzanalyse weist dabei eine höhere Übereinstimmung zum Expertenrating auf, als die Arbeitsqualitätsanalyse. Außerdem dürfte bei der Handlungssequenzanalyse die Inter-Rater-Reliabilität² höher sein, weil es sich hier um ein niedrig inferentes Beurteilungsverfahren handelt. Insbesondere die Handlungssequenzanalyse erscheint daher als ein geeignetes Instrument zur Diagnostik der Fähigkeit zur Durchführung von Experimenten.

6. Literatur

- [1] Schreiber, N., Theyßen, H., Schecker, H. (2009): Experimentelle Kompetenz messen?!. *Physik und Didaktik in Schule und Hochschule* 3/8, 92 -101.
- [2] Schreiber, N., Theyßen, H., Schecker, H. (2011): Diagnostik experimenteller Kompetenz: Ergebnisse von Pilotstudien. In: D. Höttecke (Hrsg.): *Naturwissenschaftliche Bildung als Beitrag zur Gestaltung partizipativer Demokratie*. Jahrestagung in Potsdam 2010, Münster: LIT, 244-246.
- [3] Wittig, C., Everson, T. (1999): Lead It Be – Laboratory Rubric. Homepage: <http://ed.fnal.gov/lincon/w99/projects/leaditbe/abrubric.html> (Stand: 04/2011)
- [4] Kulgemeyer, C. (2010): *Physikalische Kommunikationskompetenz. Modellierung und Diagnostik*, Berlin: Logos.
- [5] Dickmann, M. (2011): *Vergleich prozessorientierter Auswertungsverfahren für Experimentaltests*. Masterarbeit, Technische Universität Dortmund.

¹ Dickmann [5] nutzt als Maß der Übereinstimmung die relative Übereinstimmung, weil die Reihenfolge der vorgelegten Experimentdurchführungen zufällig bestimmt worden ist (vgl. [5], S. 29).

² Ergebnisse zur Inter-Rater-Reliabilität liegen für beide Auswertungsverfahren noch nicht vor, sondern nur für die Kategorisierung der Handlungen.