

Problemlösen in der Mechanik: eine Untersuchung mit Studierenden

Martina Brandenburger*, Silke Mikelskis-Seifert*, Peter Labudde[†]

*Pädagogische Hochschule Freiburg, Abteilung Physik [†]Pädagogische Hochschule Fachhochschule Nordwestschweiz

Martina.Brandenburger@ph-freiburg.de, silke.mikelskisseifert@ph-freiburg.de, peter.labudde@fhnw.ch

Kurzfassung

Sowohl im schulischen als auch universitären Umfeld nehmen Probleme in Form von Übungsaufgaben oder Prüfungen einen wichtigen Platz ein, da so die Leistungen der Lernenden in Physik eingeordnet werden können (vgl. z.B. Fischer & Draxler [1] oder Kühn [2]).

Um die Leistung von Studierenden im Bereich „Mechanik“ zu messen, wurde ein Papier-und-Bleistift-Test entwickelt, der auf dem Modell des wissenszentrierten Problemlösens (nach Friege [11]) basiert. Nach diesem Modell lässt sich das Problemlösen in vier Phasen unterteilen, die unterschiedliche Anforderungen an den Bearbeitenden stellen. Die einzelnen Items des Tests wurden so entworfen, dass jeweils nur eine Phase eines Problems bearbeitet wird. Um der Komplexität des Problemlösens mit einer großen Variation von Aufgaben gerecht werden zu können, wurde ein spezielles Testheft-Design verwendet, das die Anzahl der zu bearbeitenden Aufgaben pro Testheft auf ein praktikables Maß reduziert.

Ausgewertet wurde der Test mit Hilfe eines Rasch-Modells, das die Leistungen der Studierenden gut beschreibt. Über die Itemschwierigkeit lassen sich mehrere, qualitativ unterschiedliche Stufen finden, welche die Fähigkeiten der Studierenden mit entsprechendem Personenparameter widerspiegeln. Es wurden Gruppenvergleiche durchgeführt, die im Hinblick auf ein Modell zur Erklärung des Erfolgs beim Problemlösen erste Hinweise liefern.

1. Einleitung

1.1. Ausgangssituation

Die Bearbeitung von Probleme bzw. Aufgaben ist in der Physik von zentraler Bedeutung (vgl. z.B. Fischer & Draxler [1] oder Kühn [2]). Sie werden, in Form von Übungsaufgaben, zum Lernen benutzt, dienen aber auch als Prüfungsaufgaben zur Leistungskontrolle.

Fragt man Studierende, wie sie ihre Leistungen bei diesen wichtigen Tätigkeiten einschätzen, so ergibt sich ein überwiegend positives Selbstkonzept in Physik und beim Problemlösen. Auf einer Likert-Skala von 1 (geringes Selbstkonzept) bis 4 (hohes Selbstkonzept) ergibt sich für die im Abschnitt 4.3 vorgestellte Population ein Durchschnitt von 2.83 zum allgemeinen Selbstkonzept in Physik und einen Durchschnitt von 2.67 beim Selbstkonzept zum Problemlösen (für eine nähere Beschreibung der verwendeten Skalen s.h. Brandenburger & Mikelskis-Seifert [3]).

Vergleicht man diese positive Selbsteinschätzung mit den Erfahrungen über tatsächliche Leistungen von Studierenden im Lehrbetrieb, ist jedoch eine Diskrepanz festzustellen. Typischerweise fällt es Studierenden, beispielsweise in Übungen, schwer, Probleme zu bearbeiten – sie scheitern sowohl bei etwaigen Ansätzen, als auch bei der (mathematischen) Durchführung.

1.2. Forschungsfragen

Die oben beschriebene Ausgangssituation stellt den Rahmen für ein Dissertationsprojekt dar, das zur Klärung der Frage, was den Erfolg beim Problemlösen beeinflusst, beitragen möchte. Hierbei werden folgende Forschungsfragen und Hypothesen untersucht:

(F1) *Wie kann die Fähigkeit zum Problemlösen in der Physik differenziert gemessen werden?* (H1) Da die vier Phasen des Problemlösens verschiedene Anforderungen an den Bearbeitenden stellen, wird eine unterschiedliche Schwierigkeit der einzelnen Phasen vermutet. Es erscheint somit sinnvoll, ein Testformat zu entwickeln, das die Phasen des Problemlösens getrennt und unabhängig voneinander untersucht.

(F2) *Mit welchen (personenbezogenen) Faktoren steht der Erfolg beim Problemlösen in Verbindung und wie groß ist jeweils dieser Einfluss?* (H2a) Aus den Ergebnissen der Expertiseforschung ist bekannt, dass das domänenspezifische Vorwissen in Physik und allgemein die Erfahrungen mit Problemlösen einen großen Einfluss auf den Erfolg beim Problemlösen haben. (H2b) Von geringerem Einfluss werden weitere personenbezogene Faktoren wie das Selbstkonzept der Probanden und das Interesse sein. (H2c) Es wird angenommen, dass auch die Kenntnisse in Mathematik einen Einfluss auf den Erfolg beim

Problemlösen haben. Erwartet wird, dass dieser Einfluss in den verschiedenen Phasen des Problemlösens unterschiedlich groß sein wird.

Der Fokus der folgenden Ausführungen liegt auf der Forschungsfrage (F1), auch mit Hinblick auf eine Untersuchung der Leistungen der untersuchten Studierenden.

2. Theoretischer Rahmen

Theoretisch eingebettet ist die Arbeit in die Expertiseforschung. Deren zentrales Anliegen ist die Untersuchung herausragender menschlicher Leistungen und die Bedingungen ihres Zustandekommens (vgl. Gruber [4]). Typischerweise werden Probleme aus Lehrbüchern der entsprechenden Domäne herangezogen, um den Expertisegrad einer Person über Leistungstests zu ermitteln (vgl. Reinhold et al. [5]).

2.1. Was ist ein Problem?

Auch wenn in vielen Untersuchungen mit Problemen gearbeitet wird, herrscht Uneinigkeit darüber, wie der Begriff „Problem“ zu definieren ist (vgl. z.B. Smith [6] oder Maloney [7]). Im Folgenden wird dargelegt, was im Rahmen dieser Arbeit unter „Problem“ verstanden wird.

„Klassisch“ für die Definition von „Problem“ sind sogenannte „gap“ (Lücken) Definitionen. Ein Beispiel aus dem deutschsprachigen Raum ist die Definition nach Dörner [8], die besagt, dass ein Problem als „Denkanforderung“ sich durch drei Komponenten kennzeichnet: (1) ein unerwünschter Anfangszustand (2) ein erwünschter Endzustand und (3) eine Barriere, welche die Transformation von (1) zu (2) verhindert. Charakteristisch für diese Art von Definitionen ist die Barriere oder Lücke zwischen (1) und (2). Die Barriere ist von der Person, welche die Anforderung bearbeitet, abhängig. Nach Dörner [8] gibt es somit einen Unterschied zwischen einem „Problem“ und einer „Aufgabe“ – für eine „Aufgabe“ sind die zur Transformation von (1) zu (2) benötigten Mittel der Person bekannt. Für Untersuchungen zum Problemlösen ist es aber nicht vorteilhaft, wenn es vom jeweiligen Probanden abhängt, ob es sich bei einer Anforderung noch um ein Problem oder schon um eine Aufgabe handelt. Weitere (internationale) „gap“ Definitionen, die ähnlich angelegt sind, finden sich zum Beispiel bei Newell & Simon [9] oder Hayes [10].

Smith [6] hat die oben genannte Problematik aufgegriffen und eine erweiterte Definition formuliert. Als Problem ist nun jede Anforderung anzusehen, die das Analysieren und Schlussfolgern auf ein Ziel (oder eine „Lösung“) hin benötigt. Hierbei muss das Schlussfolgern auf dem Verständnis des bearbeiteten Themenbereichs beruhen. Daraus folgt unmittelbar, dass Probleme nicht algorithmisch gelöst werden können und zur Lösung mehr nötig ist, als das bloße Erinnern von Fakten.

Diese weiter gefasste Definition von Smith stellt die Basis der hier vorgestellten Arbeiten dar. Dement-

sprechend werden „Problem“ und „Aufgabe“ im Folgenden als synonyme Begriffe aufgefasst.

2.2. Typen von Problemen

Wendet man die in Abschnitt 2.1 dargelegte Auffassung von „Problem“ auf die Domäne Physik an, lassen sich in gängigen Lehrbüchern zwei verschiedene Arten von Problemen unterscheiden (nach Friege [11]).

Der Typ der „Einsetzaufgabe“ umfasst alle Probleme, die das bloße Einsetzen von Werten in eine Formel beinhalten. Ein Beispiel hierfür wäre: *„Wie weit fällt ein Stein im freien Fall (ohne Luftwiderstand) in 10 Sekunden?“*. Solche Probleme sind nicht sehr komplex und es wird nur wenig physikalisches Wissen benötigt, das nicht weiter verknüpft oder angewandt werden muss.

Der zweite Typ sind sogenannte „wissenszentrierte Probleme“. Bei dieser Art von Problemen wird die Kenntnis über viele physikalische Wissens Elemente vorausgesetzt, die zur Lösung nicht nur reproduziert, sondern auch zueinander in Beziehung gesetzt werden müssen. Ein Beispiel hierfür ist: *„Eine Person stößt einen Körper an, sodass er zunächst über den Tisch rutscht und dann liegen bleibt. Erarbeiten Sie eine Formel, mit der Sie bestimmen können, wie weit der Körper rutscht.“* (Problem nach Sherin [12]). Bei den in Abschnitt 3 vorgestellten Problemen handelt es sich um wissenszentrierte Probleme.

Die Untersuchung der bei der Bearbeitung von (wissenszentrierten) Problemen ablaufenden Prozesse ist im Rahmen der Lehr-Lern-Forschung ein traditionelles Forschungsgebiet¹. Bereits sehr frühe, nicht empirische, Arbeiten (z.B. Dewey [13] oder Polya [14]) geben einen geordneten Ablauf des Problemlösens in mehreren, aufeinander aufbauenden Phasen an. Das im nächsten Abschnitt vorgestellte Modell des wissenszentrierten Problemlösens von Friege [11] baut auf diesen Ergebnissen der Forschung zum Problemlösen auf und beschreibt Problemlöseprozess im Rahmen der Domäne „Physik“.

Das Modell des wissenszentrierten Problemlösens wurde als Basis für die hier vorgestellte Arbeit ausgewählt, da es gut in der physikdidaktische Forschung eingebettet ist und sich durch seine klare Struktur angemessen operationalisieren lässt. Dadurch eignet es sich in besonderer Weise für die empirische Prüfung der in Abschnitt 1.2 vorgestellten Forschungsfragen.

2.3. Modell des wissenszentrierten Problemlösens

Die Bearbeitung von wissenszentrierten Problemen erfolgt (nach Friege [11]) formal nach vier Phasen, die unterschiedliche Anforderungen an den Bearbeitenden stellen (s.h. Abb. 1).

¹ Einen guten Überblick zur Forschung zum Problemlösen liefert z.B. Maloney [15].

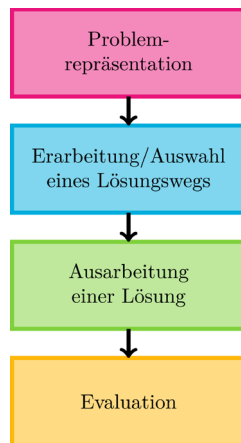


Abb. 1: Modell des wissenszentrierten Problemlösens nach Friege (2001) [11]

(1) *Problemrepräsentation*: Die physikalischen Inhalte eines Problems werden erkannt und mit Fachtermini ausgedrückt. Zur Repräsentation gehören Skizze, physikalische Konzepte und Idealisierungen. (2) *Auswahl / Erarbeitung eines Lösungswegs*: Ist ein ähnliches Problem wie das zu bearbeitende bereits bekannt, kann dies als Beispielproblem für den Lösungsweg herangezogen werden. Ist kein ähnliches Problem bekannt, wird durch Rückgriff auf Faktenwissen (z.B. $F = m \cdot a$) und Beziehungen zwischen Wissens-elementen (Kräfte addieren sich) ein neuer Lösungsweg entwickelt. (3) *Erarbeitung einer Lösung*: Unter Verwendung des Lösungswegs wird eine Lösung ausgearbeitet. Hierbei stehen Probleme mathematischer Natur im Vordergrund. (4) *Evaluation*: Nachdem eine Lösung ermittelt wurde, wird versucht, diese auf ihre Richtigkeit zu prüfen.

Für die Entwicklung des Fragebogens wurde die Phase der Evaluation ausgeklammert. Stattdessen wurde das Nachvollziehen von ausgearbeiteten Lösungen mit aufgenommen, da vermutet wird, dass es sich hierbei um eine grundlegende Voraussetzung zum selbstständigen Problemlösen handelt. Die anderen drei Phasen wurden in der Untersuchung berücksichtigt.

Im nächsten Abschnitt wird der entwickelte Problemlösetest vorgestellt.

3. Instrumente und Methoden

3.1. Der Problemlösetest

Betrachtet man die in Abschnitt 2.3 beschriebenen Phasen des Problemlösens, so stellt man fest, dass diese unterschiedliche Anforderungen an den Bearbeitenden stellen. Deshalb erschien es nicht angemessen, die Fähigkeit zum Problemlösen lediglich über die Betrachtung vollständiger Problemlöseprozesse zu untersuchen. Scheitert ein Student nämlich bereits bei der Repräsentation, können die weiteren Schritte nicht beobachtet werden, auch wenn der Student diese, mit Hilfe von Hinweisen, vielleicht korrekt durchführen könnte. Die Problemlösefähig-

keit wird durch den Test als Summe ihrer Teile, der Phasen des Problemlösens, untersucht.

Es wurde ein Papier-und-Bleistift-Test entwickelt, der die zur Bearbeitung einer Phase des Problemlösens benötigten vorangegangenen Phasen vorgibt. Soll beispielsweise eine Lösung durchgeführt werden, so sind Repräsentation und Lösungsweg angegeben.

Es wurden für die Untersuchung vier Probleme entsprechend vorbereitet, sodass sich daraus zwölf Items für den Problemlösetest ergeben. Die Probleme stammen aus den Bereichen „Energie“ und „Kinematik“ im Rahmen der Mechanik.

3.2. Beispielitems

Im Folgenden werden zwei Beispielitems vorgestellt, die den Charakter des Problemlösetests verdeutlichen sollen. Beide Probleme wurden aus einem Physiklehrbuch (Giancoli [18]) entnommen.

P1 – Sprung an ein Seil: Ziel dieses Problems ist es, herauszufinden, in welchem Winkel ein Seil, das an einem Baum hängt, ausgelenkt wird, wenn eine Person mit einer gewissen Anfangsgeschwindigkeit an es springt. Das Item zur Repräsentation dieses Problems (im Folgenden als „Seil“ bezeichnet) beinhaltet das Ausfüllen eines Lückentexts. Dieses geschlossene Format wurde gewählt, da Pilotierungen mit einem offenen Format Probleme mit Bodeneffekten zeigten (s.h. Brandenburger & Mikelskis-Seifert [3]). Ebenfalls soll eine Skizze angefertigt werden.

P3 – Die gelbe Ampel: Ziel dieses Problems ist es, zu berechnen, wie lange das gelbe Licht einer Ampel leuchten muss, damit ein Auto rechtzeitig Anhalten oder doch noch über die Kreuzung fahren kann. Das Item zum Lösungsweg (im Folgenden als „Ampel“ bezeichnet) gibt eine geeignete Repräsentation inklusive Skizze des Problems vor. Es sollen nun das relevante physikalische Faktenwissen erklärt und zwischen mehreren möglichen Lösungsansätzen (verbal und als Formel) via Multiple-Choice ausgewählt werden.

3.3. Das Testheft-Design

Um der Komplexität des Problemlösens mit einer größeren Variation von Aufgaben, bei gleichzeitiger angemessener Anzahl von Aufgaben pro Testheft, gerecht werden zu können, wurde ein spezielles Design verwendet, das die Probleme (mit Überlappungen) durchrotiert. Insgesamt besteht ein Testheft aus jeweils vier Items, die die vier Phasen des Problemlösens abdecken – jedoch von vier verschiedenen Problemen (s.h. Tabelle 1).

	Repräsen- tation	Lösungs- weg	Lösung	Nachvoll- ziehen
Testheft A	P1	P2	P3	P4
Testheft B	P1	P3	P4	P2
Testheft C	P4	P3	P2	P1
Testheft D	P2	P1	P2	P4

Tabelle 1: Design des Problemlösetests

Die Testzeit für ein Testheft beträgt 45 Minuten.

Die Auswertung des Problemlösetest erfolgt mit Hilfe eines Rasch-Modells. Dies wird im nächsten Abschnitt vorgestellt.

4. Auswertung

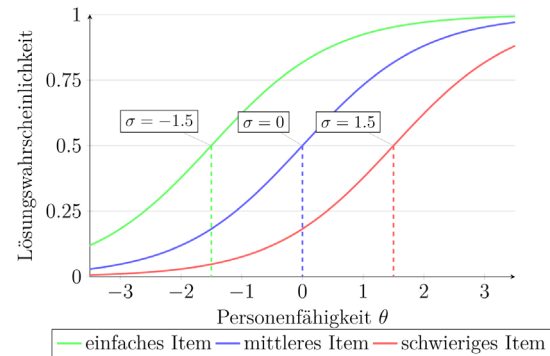
Da es sich bei dem Problemlösetest um einen Leistungstest handelt, bei dem anzunehmen ist, dass nicht alle Items die gleiche Schwierigkeit besitzen (was eine Annahme der klassischen Testtheorie ist), wurden die Daten mit einem probabilistischen Modell – dem Rasch-Modell – ausgewertet. Des Weiteren ist es Rasch-Modellen möglich, mit einer Vielzahl „fehlender Werte“ umzugehen, die bei einem Testheft-Design, wie oben beschrieben, zwangsläufig auftreten.

Im Folgenden werden einige Grundannahmen zum Rasch-Modell zusammengefasst. Eine vertiefte Einführung liefert z.B. Rost [19].

4.1. Einführung in das Rasch-Modell

Eine Grundannahme der probabilistischen Testtheorie ist, dass Items eine unterschiedliche Schwierigkeit σ besitzen, weshalb sie von Personen mit unterschiedlicher latenter Personenfähigkeit θ mit einer unterschiedlichen Wahrscheinlichkeit gelöst werden können. Hat eine Person beispielsweise die Personenfähigkeit 1.3 und das Item die Schwierigkeit 1.3, so kann dieses Item mit einer Wahrscheinlichkeit von 50% gelöst werden. Schwierigere Items ($\sigma > 1.3$) können mit einer Wahrscheinlichkeit kleiner 50%, einfachere Items ($\sigma < 1.3$) größer 50% gelöst werden. Einem Item wird diejenige Itemschwierigkeit σ zugeordnet, die dem Wert der Personenfähigkeit θ entspricht, bei der eine Lösungswahrscheinlichkeit von 50% vorliegt.

Die Funktion der Lösungswahrscheinlichkeiten für verschiedene Personenwerte bei einem bestimmten Item hat, einen charakteristischen Verlauf (s.h. Abb. 2), der nach den Annahmen des Rasch-Modells für alle Itemschwierigkeiten gleich ist.


Abb. 2: Theoretischer Verlauf der Itemfunktion

Die Itemschwierigkeit und Personenfähigkeit werden auf der gleichen Skala aufgetragen, der Logit-Skala. Diese ist eine logarithmische Transformation der Wettquotienten („Odds“) für die Wahrscheinlichkeit einer Lösung. Hätte eine Person für ein Item (mit der Schwierigkeit $\sigma = 0$) eine Lösungswahrscheinlichkeit von 25%, so ist der Wettquotient hierfür 1:3. Logarithmiert man diesen Wert, erhält man für die Personenfähigkeit θ den Logit -1.1.

Itemschwierigkeit und Personenfähigkeit werden während der Berechnung des Modells so geschätzt, dass die Wahrscheinlichkeit der gegebenen Daten („Likelihood“) unter den geschätzten Modellparametern maximal ist. Hierbei wird eine Summenorientierung festgelegt. Ist die Itemschwierigkeit summenorientiert, so ist die Summe aller Itemschwierigkeiten gleich 0. Ist die Personenfähigkeit summenorientiert, so ist die Summe aller Personenfähigkeiten gleich 0. Dies bedeutet, dass es sich bei der Logit-Skala um keine absolute Skala handelt – von Bedeutung sind weniger die einzelnen Werte, als die Differenzen zwischen Werten.

Über ihren Likelihood-Wert können verschiedene Rasch-Modelle, die auf den gleichen Daten beruhen, miteinander verglichen werden.

Das einfachste Rasch-Modell, anhand dessen die vorherigen Annahmen vorgestellt wurden, ist das eindimensionale dichotome Rasch-Modell. Die Daten zu jedem Item liegen hierbei dichotom (0 oder 1 – falsch oder richtig) vor. Die Grundideen dieses Modells können allerdings einfach auf nicht-dichotome Daten übertragen werden. Ein Vertreter hierfür ist das Partial-Credit-Modell (Schwellenmodell). Ein Anwendungsbeispiel ist ein Test, bei dem jedes Item nicht nur komplett richtig oder komplett falsch sein kann, sondern es auch teilweise richtige Lösungen gibt. Das Erreichen eines Teilpunktes oder der vollständigen Punktzahl wird als Überschreiten einer Schwelle gesehen, der eine Schwierigkeit zugeordnet wird (und somit auch eine benötigte Personenfähigkeit und das Item lösen zu können). Beim Partial-Credit-Modell müssen die Differenzen zwischen den Schwellen bei allen Items nicht gleich sein (im Gegensatz zum Ratingskalen-Modell).

Zur Verdeutlichung anhand des oben vorgestellten Problemlösetests: bei jedem Item werden die Ant-

worten der Studierenden als 0 (nicht gelöst), 1 (teilweise richtig gelöst / Ansatz), 2 (komplett richtig gelöst / vollständig) kodiert. Nach der Berechnung des Modells ergibt sich für das Item „Seil“, dass eine teilweise richtige Lösung eine Itemschwierigkeit von -0.96 und eine komplett richtige Lösung eine Schwierigkeit von 0.25 besitzt. Für das Item „Ampel“ liegen die Schwellen bei -0.10 und 1.05 .

4.2. Gütekriterien des Rasch-Modells

Um die Ergebnisse eines Rasch-Modells nutzen zu können, muss gesichert werden, dass alle Items (inkl. der Schwellen) Rasch-homogen sind. Anschaulich ausgedrückt wird geprüft, ob der tatsächliche Verlauf der Itemfunktionen dem theoretisch geforderten Verlauf (s.h. Abb. 2) entspricht. Hierfür gibt es zwei wichtige Prüfmaße, den wMNSQ (weighted mean square – gewichtete mittlere quadratische Abweichung) und den dazugehörigen t-Wert. Der Erwartungswert des wMNSQ ist 1.00 (perfekte Übereinstimmung). Der tatsächliche Wert muss nach Adams & Wu [20] für jedes Item zwischen 0.80 und 1.20 liegen und der kritische Werte der t-Verteilung von 1.96 darf für kein Item überschritten werden. Darüber hinaus wird die klassische Trennschärfe jedes Items berechnet. Diese sollte über 0.30 liegen.

4.3. Rasch-Modell des Problemlösetests

Für den in Abschnitt 3.1 vorgestellten Problemlösetest wurde zur Auswertung ein eindimensionales Rasch-Modell mit Partial-Credit gewählt. Die Berechnung erfolgte mit conquest [21]. Es wurden die in Abschnitt 3.1 genannten zwölf Items mit jeweils zwei Schwellen einbezogen. Die Kodierung der Antworten ist entsprechend 0 (nicht gelöst), 1 (teilweise richtig gelöst / Ansatz), 2 (komplett richtig gelöst / vollständig). Diese Kodierung entspricht gleichzeitig der Bepunktung der Items, weshalb pro Testheft maximal 8 Punkte (je zwei pro Phase) erreicht werden können. Der vorläufigen Datensatz beinhaltet 174 Studierende aus verschiedenen Studienrichtungen der Physik (Lehramt, Nebenfach, Fach). Für eine endgültige Auswertung werden mindestens 200 Personen (50 pro Testheft) angestrebt.

Nach den in Abschnitt 4.2 genannten Kriterien ist der Problemlösetest Rasch-homogen; alle wMNSQ liegen zwischen 0.80 und 1.20 , es findet keine Überschreitung des kritischen t-Werts statt und die klassische Trennschärfe aller Items liegt zwischen 0.40 und 0.84 .

Es wurden Modellvergleiche mit anderen potenziellen Modellen (Ratingskalen-Modell, mit Hintergrundmodell, mehrdimensionale Modelle) durchgeführt. Es stellt sich (mit Hilfe von χ^2 -Tests) heraus, dass das eindimensionale Partial-Credit Modell unter den gegebenen Daten als das beste Modell anzusehen ist.

Trägt man die Personenfähigkeit θ und die Itemschwierigkeit σ auf der Logit-Skala auf, so ergibt sich folgende Grafik (Abb. 3).

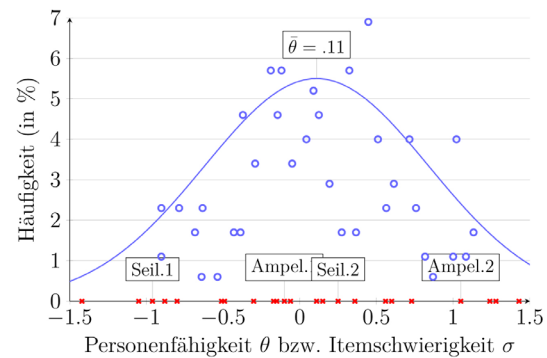


Abb. 3: Personen- und Itemwerte auf der Logit-Skala

Es lässt sich erkennen, dass die Schwierigkeit der Items (rote Kreuze auf der x-Achse) breit gestreut ist, ohne allzu große Lücken aufzuweisen. Es wird also ein Schwierigkeitsbereich gut abgedeckt. Betrachtet man die Beispielitems aus Abschnitt 3.2, so stellt man fest, dass „Seil.1“ (Ansatz zu „Seil“) ein relativ einfaches Item ist ($\sigma = -0.96$). Dieses Item vollständig zu bearbeiten („Seil.2“) liegt bei einem mittleren Schwierigkeitsgrad ($\sigma = 0.25$).

Die Personenwerte (blaue Kreise) streuen über die Bandbreite der möglichen Werte. Sie haben ihren Mittelwert bei 0.11 und eine Standardabweichung von 0.72 . Da die Itemwerte auf 0 summenorientiert sind, wäre bei einem perfekt angemessenen Test, der nicht zu leicht und nicht zu schwer für die entsprechende Population ist, ein Mittelwert von 0 zu erwarten. Der Test ist zwar etwas zu leicht ($\bar{\theta} > 0$), kann jedoch trotzdem als passend angesehen werden.

5. Ergebnisse

5.1. Schwierigkeitsprofil und Stufenmodell

Ordnet man die Items der Schwierigkeit nach an, ergibt sich ein Schwierigkeitsprofil (s.h. Abb. 4).

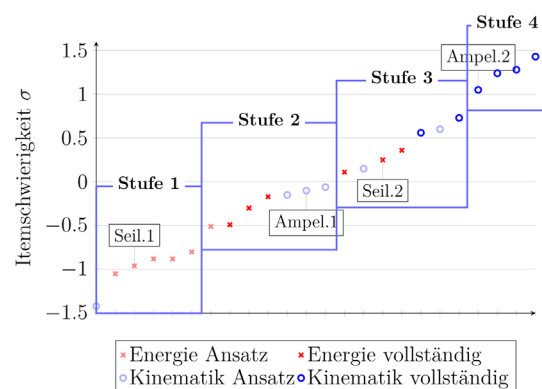


Abb. 4: Schwierigkeitsprofil der Items

Für eine Komplexitätsreduzierung wird nun versucht, Gruppen von Items („Stufen“) zu finden, die von der Schwierigkeit her vergleichbar sind und ähnliche Anforderungen an den Bearbeitenden stellen. Aus statistischer Sicht sind hierbei besonders „Sprünge“ in der Itemschwierigkeit interessant. Die in Abb. 4 eingezeichneten Stufen orientieren sich an

solchen Sprüngen – die Differenzen zwischen zwei Items sind an den Stufengrenzen besonders groß. Des Weiteren spiegelt sich auch die Verteilung der Population in den Stufen wider. An der Grenze zwischen Stufe 2 und Stufe 3 liegt der Mittelwert von 0.11 und Stufen 2 und 3 sind jeweils eine Standardabweichung von 0.72 breit.

Über diese statistische Betrachtung hinaus, lässt sich die Anordnung der Items auch aus inhaltlichen Gründen erklären. Insbesondere der Bereich der Items (Energie oder Kinematik) scheint einen großen Einfluss auf die Schwierigkeit zu haben. Um dies graphisch darzustellen, wurden für die Items aus dem Bereich „Energie“ als rote Kreuze dargestellt, die Items aus dem Bereich „Kinematik“ als blaue Kreise (jeweils noch mit der Abstufung hell und dunkel für die Schwellen „Ansatz“ und „vollständig“). Es lässt sich so erkennen, dass sich die Items aus dem Bereich „Energie“ tendenziell weiter links anordnen, also einfacher sind, wohingegen die Items zur „Kinematik“ schwieriger sind und somit weiter rechts im Schwierigkeitsprofil zu finden sind. Bei einer detaillierteren Betrachtung zeigt sich, dass die Phasen des Problemlösens unterschiedlich schwierig sind. Tendenziell sind die Lösung und das Nachvollziehen eines Problems einfacher als das Erstellen einer Repräsentation und die Ausarbeitung eines Lösungswegs. Dies deckt sich mit Ergebnissen der Forschung zum Problemlösen (s.h. z.B. Chi et al. [16] oder Hayes [10]). Diese legen nahe, dass die Repräsentation und die Erarbeitung eines Lösungswegs als zentral für die Lösung eines Problems anzusehen sind und große Anforderungen an den Bearbeitenden stellen. Die richtige Lösung hingegen ergibt sich, quasi zwangsläufig, aus der Korrektheit der Repräsentation und des erarbeiteten Lösungswegs (vgl. z.B. Heller & Reif [17] oder Newell & Simon [9]).

Zusammengefasst können Personen, die einen Personenwert innerhalb einer der gefundenen Stufen besitzen, folgende Anforderungen erfüllen:

Stufe 1: Die Bearbeitenden können Ansätze in einfachen Bereichen (Energie) finden, wenn es sich nicht um Lösungen handelt. Im Schnitt werden 2.2 von 8 Punkten erreicht. 16% der Population befinden sich in dieser Stufe.

Stufe 2: Die Bearbeitenden können einfache Phasen des Problemlösens (Lösung und Nachvollziehen) in einfachen Bereichen (Energie) vollständig bearbeiten und in schwierigen Bereichen (Kinematik) Ansätze finden. Im Schnitt werden 3.3 von 8 Punkten erreicht. 29% der Population befinden sich in dieser Stufe.

Stufe 3: Die Bearbeitenden können schwierigere Phasen des Problemlösens (Repräsentation und Lösungsweg) in einfachen Bereichen (Energie) vollständig bearbeiten und in schwierigen Bereichen (Kinematik) Ansätze finden. Im Schnitt werden 4.9

von 8 Punkten erreicht. 42% der Population befinden sich in dieser Stufe.

Stufe 4: Die Bearbeitenden können schwierige Bereiche (Kinematik) vollständig bearbeiten. Im Schnitt werden 7.0 von 8 Punkten erreicht. 13% der Population befinden sich in dieser Stufe.

5.2. Gruppenvergleiche

Für eine erste Auswertung wurden einfache Gruppenvergleiche der erreichten Punkte² (von maximal 8 Punkten) durchgeführt (s.h. Abb. 5). Da die Punkte über die verschiedenen Populationen nicht normalverteilt sind, wurden die Mittelwertvergleiche mit Mann-Whitney-U-Tests durchgeführt.

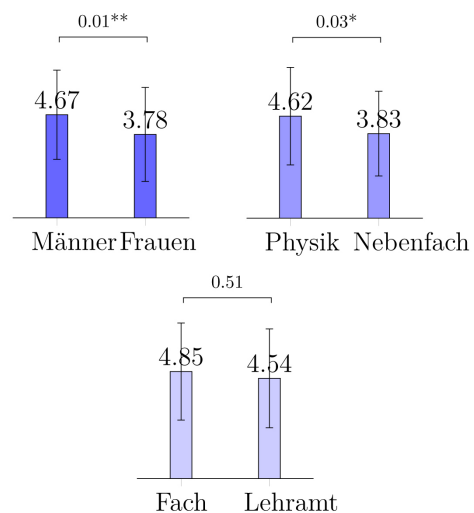


Abb. 5: Gruppenvergleiche der erreichten Punkte

Wie aus zahlreichen Untersuchungen zu Leistungsunterschieden zwischen Männern und Frauen zu erwarten war, ergeben sich zwischen beiden Gruppen signifikante Unterschiede. Es bleibt anzumerken, dass für diesen Vergleich jedoch keine Kovariate, wie zum Beispiel der Studienschwerpunkt, mit einbezogen wurden, von denen jedoch ein stärkerer Einfluss als das Geschlecht an sich zu erwarten ist.

Die Unterteilung der Population nach Studienschwerpunkt, d.h. ob Studierende Physik als Hauptfach oder als Nebenfach (z.B. im Studium der Biologie) belegen, führt ebenfalls zu signifikanten Unterschieden zugunsten der Hauptfachstudierenden. Auch dieser Unterschied war zu erwarten, da davon ausgegangen werden kann, dass Studierende des Hauptfachs zum einen über mehr Fachwissen verfügen und sich zum anderen im Studium länger/intensiver mit dem Problemlösen beschäftigt haben.

Ausgehend davon ist überraschend, dass sich keine signifikanten Unterschiede zwischen Studierenden, die Physik als Fachstudium angeben, und Studierenden des Lehramts finden lassen.

² Da alle Items Rasch-homogen sind, können die die Personenfähigkeit θ und die erreichte Punktzahl (Summenscore) als äquivalent angesehen werden (vgl. Rost [19]).

Diese Mittelwertvergleiche sind ein erster Schritt, um Unterschiede beim Problemlöseerfolg zu erklären. Die Hauptstudie, die im folgenden Abschnitt kurz vorgestellt wird, geht dem weiter nach.

6. Ausblick auf die Hauptstudie

Ziel des Dissertationsvorhabens ist es, wie in Abschnitt 1.2 dargelegt, eine Antwort auf die Frage zu finden, was den Erfolg beim Problemlösen beeinflusst.

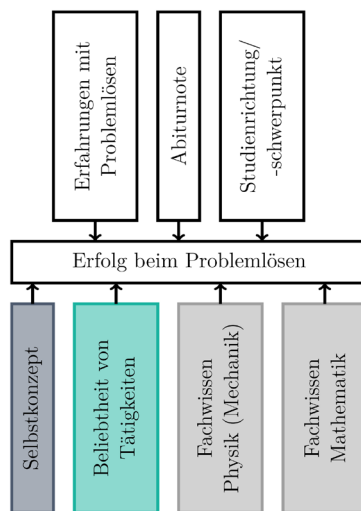


Abb. 6: Überblick über die Hauptstudie

Als zentrales Element für dieses Vorhaben wurde der hier vorgestellte Problemlösetest entwickelt. Darüber hinaus werden weitere ausgewählte Prädiktorvariablen erhoben (s.h. Abb. 6): Angaben zur Person (Abiturnote, Studienrichtung, Erfahrungen mit Problemlösen), Likert-Skalen zum problemlösebezogenen Selbstkonzept und zur Beliebtheit von Tätigkeiten (s.h. [3]) und Fachwissenstests in Mathematik (s.h. [22]) und Mechanik. Die Testzeit für den gesamten Test beträgt 90 Minuten. Ausgewertet wird die Untersuchung mit einer Mischung aus Varianzanalysen, Rasch-Modellen, Klassenbildung mit LCA und linearer Regression. Angestrebt wird eine Stichprobengröße von mindestens 200 Personen.

Abschließendes Ziel der Untersuchung ist die Aufstellung eines Modells, das weitere Erkenntnisse über Zusammenhänge zwischen den verschiedenen Prädiktorvariablen und dem Erfolg beim Problemlösen ermöglicht.

7. Literatur

- [1] Fischer, Hans E.; Draxler, Dennis (2001): Aufgaben und naturwissenschaftlicher Unterricht. In: MNU 54 (7), S. 388–393.
- [2] Kühn, Svenja Mareike (2011): Weiterentwicklung der Aufgabenkultur im naturwissenschaftlichen Unterricht der gymnasialen Oberstufe und im Abitur. In: Zeitschrift für Didaktik der Naturwissenschaften 17, S. 35–55.
- [3] Brandenburger, Martina; Mikelskis-Seifert, Silke (2013): Was beeinflusst den Erfolg beim

- Problemlösen in der Physik? In: Sascha Bernholt (Hg.): Inquiry-based Learning - Forschendes Lernen. GDCP Jahrestagung 2012 Hannover. Kiel: IPN, S. 761–763.
- [4] Gruber, Hans (2010): Expertise. In: Detlef H. Rost (Hg.): Handwörterbuch pädagogische Psychologie. 3 // 4. Weinheim: Beltz PVU, S. 183–189.
- [5] Reinhold, Peter; Lind, Gunter; Friege, Gunnar (1999): Wissenszentriertes Problemlösen in Physik. In: Zeitschrift für Didaktik der Naturwissenschaften 5 (1), S. 41–62.
- [6] Smith, Mike U. (1991): A View from Biology. In: Mike U. Smith (Hg.): Toward a unified theory of problem solving. Views from the content domains. Hillsdale NJ: Lawrence Erlbaum Associates, S. 1–19.
- [7] Maloney, David P. (2011): An Overview of Physics Education Research on Problem Solving. In: C. Henderson und K. A. Harper (Hg.): Getting Started in PER (Reviews in PER, 2). Online verfügbar unter <http://www.compadre.org/Repository/document/ServeFile.cfm?ID=11457&DocID=2427>, zuletzt geprüft am 02.03.2014.
- [8] Dörner, Dietrich (1976): Problemlösen als Informationsverarbeitung. 1. Aufl. Stuttgart [u.a.]: Kohlhammer (Kohlhammer Standards Psychologie : Studententext : Teilgebiet Denkpsychologie).
- [9] Newell, Allen; Simon, Herbert A. (1972): Human problem solving. Englewood Cliffs N.J.: Prentice-Hall.
- [10] Hayes, John R. (1981): The complete problem solver. Philadelphia Pennsylvania: Franklin Institute Press.
- [11] Friege, Gunnar (2001): Wissen und Problemlösen. Eine empirische Untersuchung des wissenszentrierten Problemlösens im Gebiet der Elektrizitätslehre auf der Grundlage des Experten-Novizen-Vergleichs. Berlin: Logos-Verlag.
- [12] Sherin, Bruce L. (2001): How Students Understand Physics Equations. In: Cognition and Instruction 19 (4), S. 479–541.
- [13] Dewey, John (2002): Wie wir denken. Zürich: Verlag Pestalozzianum (John-Dewey-Reihe, 2).
- [14] Pólya, George (1985): How to solve it. A new aspect of mathematical method. 2. Aufl. Princeton N.J.: Princeton University Press.
- [15] Maloney, David P. (1994): Research on problem solving: physics. In: Dorothy L. Gabel (Hg.): Handbook of research on science teaching and learning. New York, Toronto: Macmillan. Maxwell Macmillan Canada. Maxwell Macmillan International, S. 327–354.
- [16] Chi, Michelene T. H.; Feltovich, Paul J.; Glaser, Robert (1981): Categorization and representation of physics problems by experts and novices. In: Cognitive Science 5, S. 121–152.

- [17] Heller, Joan I.; Reif, Frederick (1984): Prescribing Effective Human Problem-Solving Processes: Problem Description in Physics. In: *Cognition and Instruction* 1 (2), S. 177–216.
- [18] Giancoli, Douglas C. (2006): *Physik*. 3. Aufl. München u.a.: Pearson Studium (Ph physik).
- [19] Rost, Jürgen (2004): *Lehrbuch Testtheorie - Testkonstruktion*. 2. Aufl. Bern [u.a.]: Hans Huber (Psychologie Lehrbuch).
- [20] Adams, Raymond J.; Wu, Margaret (2002): *PISA 2000 technical report*. Paris: OECD.
- [21] Wu, Margaret L. (2007): *ACER ConQuest version 2.0. Generalised item response modeling software*. Camberwell: ACER Press.
- [22] Brandenburger, Martina; Mikelskis-Seifert, Silke; Labudde, Peter (im Druck): Einfluss der Kenntnisse in Mathematik auf das Problemlösen in Physik. In: Sascha Bernholt (Hg.): *GDCP Jahrestagung 2013 München*. Kiel: IPN