

Eine Studie zu den Merkmalen der Aufgabenschwierigkeit am Beispiel eines Informatik-Schülerwettbewerbs Erster Teil: Aufgabenklassifizierung

Kirsten Schlüter

Didaktik der Informatik
Universität Erlangen-Nürnberg
Martensstraße 3, 91058 Erlangen
schlueter@informatik.uni-erlangen.de

Abstract: Als erster Schritt einer Studie mit dem Ziel, den Kompetenzanspruch von Informatik-Aufgaben zu modellieren, wurde im Rahmen einer Umfrage unter Informatik-Didaktikern an Schulen und Universitäten eine Klassifizierung der Aufgaben eines Informatik-Schülerwettbewerbs für die Sekundarstufe erarbeitet. In diesem Artikel werden die Auswertungsmodalitäten und die Ergebnisse beschrieben, auf deren Basis im weiteren Verlauf der Studie die relevanten Kriterien der Aufgabenschwierigkeit bestimmt und kategorisiert werden.

1 Zielsetzung

Die Erfassung und Beurteilung des theoretischen Konstruktes informatischer Kompetenz mittels Aufgaben verlangt ein Schema zur Klassifizierung der Aufgaben, das dazu geeignet ist, die charakterisierenden Merkmale der Aufgabenschwierigkeit sichtbar zu machen. Das Schema soll sich an Bildungsstandards und an bewährten Taxonomien wie der von Bloom für den Bereich kognitiver Lernziele orientieren. Es soll anwendbar sein auf den Aufgabenbestand in Schulbüchern, Wettbewerben, Lehrportalen etc. und mit überschaubarem Einarbeitungsaufwand für Lehrpersonen mit Unterrichtspraxis und einer Fachausbildung in Informatik handhabbar sein.

Das Vorhaben der Entwicklung eines solchen Kriterienschemas wird in einer dreiteiligen Studie realisiert: Aufgabenklassifizierung, empirische Überprüfung und Kategorisierung der Merkmale. Über den ersten Teil wird hier berichtet. Ein bisher theoretisch begründeter Kriterienkatalog wird im Rahmen einer Umfrage unter Didaktikern an Schulen und Universitäten zur Klassifizierung eines Satzes von Informatikaufgaben herangezogen. Die Aufgabenklassifizierung ist der Ausgangspunkt für die empirische Überprüfung der Kriterien im zweiten Teil. Dazu werden die Aufgaben durch Schüler der Sekundarstufe bearbeitet und ihre empirische Schwierigkeit wird ermittelt. Im Abgleich der Resultate mit der Klassifizierung soll der Einfluss der Kriterien auf die Aufgabenschwierigkeit untersucht werden. Der dritte Teil umfasst die Anwendung kategorisierender statistischer Verfahren zur Auswahl und Gruppierung der relevanten Merkmale.

1.1 Ausgangssituation

Die Situation der Schulinformatik in Deutschland stellt sich auf der einen Seite so dar, dass Informatikunterricht noch längst nicht in allen Bundesländern und allen Schularten angeboten wird oder sogar verpflichtend ist. Bildungsstudien wie PISA konzentrieren sich auf die Kernfächer. Instrumente der Leistungserhebung wie etwa die zentralen Jahrgangsstufentests in Bayern in den Fächern Mathematik, Deutsch und Englisch gibt es nicht, entsprechend liegen keine Daten vor. Eine Aufgabekultur wie in der Mathematik konnte sich noch nicht entwickeln. Auf der anderen Seite steht das Engagement vieler Lehrpersonen mit Expertise im Fach Informatik und der Informatik-Didaktiker an den Universitäten, die eine Vielzahl von Unterrichtsbeispielen veröffentlichen und Aufgaben zum Lernen, Üben, Vertiefen und Prüfen entwickeln. Mit dem Ziel, das Schulfach Informatik in der Sekundarstufe zu stärken und zu konsolidieren, wurden im Jahr 2008 in der Gesellschaft für Informatik Bildungsstandards verabschiedet [GI08].

1.2 Taxonomien, Kompetenzmodelle, Ordnungsschemata

Puhlmann und Friedrich stellen dar, wie die Bildungsstandards der Informatik zur Beurteilung von Aufgaben für Unterricht und Test im Hinblick auf die Prozess- und Inhaltskompetenzen anwendbar sind [PF07]. Kompetenzmodelle, die sich auf einzelne Inhaltsbereiche beziehen, werden in jüngeren Arbeiten vorgestellt, z. B. von Kohl und Fothe für Algorithmen [KF07] und von Freischlad für Internetworking [Fr08]. Von Starr, Manaris und Stalvey wird ein Ansatz beschrieben, Blooms originale Taxonomie zur Curriculumsauslegung und Aufgabengestaltung heranzuziehen [St08]; Fuller et al. wenden eine zweidimensionale Version entlang der Dimensionen Interpreting und Producing an [Fu08]; Lahtinen beschreibt eine Typisierung von Programmieranfängern nach Blooms Lernzielstufen [La07]. Biggs und Collins stellen die SOLO Taxonomie vor (Prestructural, Unistructural, Multistructural, Relational) [BC82], die von Sheard et al. zur Klassifizierung von Prüfungsantworten eingesetzt wird [Sh08]. Bower schlägt eine zehnstufige Taxonomie (Declarative, Comprehension, Debugging, Prediction, Provide-an-example, Provide-a-model, Evaluate, Meet-a-design-specification, Solve-a-problem und Self-reflect tasks) zur Kategorisierung von Aufgaben vor [Bo08]. Curriculumsbedingt liegt der Fokus der internationalen Ansätze auf dem Lerninhalt Programmierung. In Deutschland hat Brinda für die strukturierte Gestaltung von Aufgaben zur objektorientierten Modellierung eine Klassifizierung nach Fachkern, Gegenstand und Aufgabentyp mit einer Einordnung der Aufgabentypen in Blooms Taxonomie entwickelt [Br04]. Steinert präsentiert ein curriculares Ordnungsschema zur Lernerfolgsanalyse, die Konstruktion von Lernzielgraphen aus Aufgaben und Musterlösungen [St07].

1.3 Spezifische Schwierigkeitsmerkmale von Informatikaufgaben

Allen Schemata ist gemeinsam, dass sie Curriculumsinhalte, Aufgaben oder Antworten klassifizieren oder Lerner kategorisieren. Sie sind nicht dazu geeignet, die Schwierigkeit von Aufgaben herauszuarbeiten. Das Augenmerk dieser Arbeit liegt auf den Kriterien, die die Schwierigkeit von Aufgaben und damit ihren Kompetenzanspruch bestimmen. Bisher wurden die charakterisierenden Merkmale Erfahrungsweltnähe, Abstraktionsgrad,

Komplexität, Formalisierungsgrad, Redundanz, Anforderungsbereich, Prozessbereich, kognitive Lernzielstufe und Art des Wissens identifiziert [Sc08]. Ausgehend von der Annahme, dass das komplexe Gefüge informatischer Kompetenz sich durch einfache Merkmale modellieren lässt, soll langfristig ein Klassifizierungsschema entwickelt werden, das im Abgleich mit dem daraus erwachsenden Kompetenzmodell dazu dient, Aufgaben anhand einfacher Kriterien nach ihrem Kompetenzanspruch zu klassifizieren.

2 Aufgabenklassifizierung: Durchführung

Zur empirischen Überprüfung der Merkmale werden die Aufgaben des Schülerwettbewerbs Informatik-Biber [Po09] verwendet – bewusst „echte“ statt eigens auf die Merkmale zugeschnittene Aufgaben und bewusst „alle“ statt ausgewählte Aufgaben. Der Informatik-Biber, eine Kooperation des Bundeswettbewerbs Informatik und der Arbeitsgruppe Didaktik der Informatik der Universität Münster, ist ein jährlich ausgerichteter Online-Wettbewerb, der sich an Schüler der Jahrgangsstufen 5 bis 7, 8 bis 10 und ab 11 richtet und keine Vorkenntnisse in Informatik voraussetzt. Die Aufgaben sind kurz und schnell zu erfassen. Sie entstammen den Themenbereichen, die für die Informatik als Schulfach relevant sind, und sollen informatische Denkweisen und Problembearbeitung erfordern, aber kein spezifisches Fachwissen. Die folgende Beispielaufgabe „Biber und Bisons“ wurde im Wettbewerb 2007 in den Jahrgangsstufen 8 bis 10 und ab 11 gestellt.

Biber sagen immer die Wahrheit, und Bisons lügen immer. Im Biber-und-Bison-Zeltlager wohnen insgesamt zehn Tiere. Ein blinder Maulwurf kommt vorbei und möchte wissen, wie viele Biber und wie viele Bisons anwesend sind. Darum fragt er jedes Tier: "Wie viele Biber gibt es hier?" Die zehn Antworten sind: 3, 4, 1, 4, 1, 1, 3, 4, 3, 2. Jetzt weiß der blinde Maulwurf genau Bescheid! Du auch?

Wie viele Biber sind im Biber-und-Bison-Zeltlager? A) 1 B) 2 C) 3 D) 4

2.1 Vorgehensweise

Gestützt durch eine umfangreiche Expertenumfrage werden im ersten Teil der Studie die Wettbewerbsaufgaben klassifiziert, über die Ergebnisse wird hier berichtet. (Die Klassifizierung wird allerdings noch nicht zur Kategorisierung der Aufgaben herangezogen, vielmehr soll sie im zweiten Schritt anhand der Bearbeitungsergebnisse der Wettbewerbsteilnehmer überprüft werden.) Um eine möglichst objektive Einordnung in das Kriterienschema zu erreichen, lag jede Aufgabe mehreren Personen mit Expertise in der Informatik-Didaktik zur Einschätzung vor. Weil die Einarbeitung in die Kriterien und die sorgfältige Einschätzung je Aufgabe einen beträchtlichen Zeitaufwand erfordern, wurden die 29 Aufgaben in Arbeitseinheiten zu je fünf Aufgaben aufgeteilt. Den Befragten lagen zur Orientierung außerdem drei Beispielaufgaben vor, deren Einordnung in einer Vorstudie innerhalb der Arbeitsgruppe Didaktik der Informatik der Universität Erlangen-Nürnberg abgestimmt wurde, sowie die folgende Spezifikation der Kriterien.¹

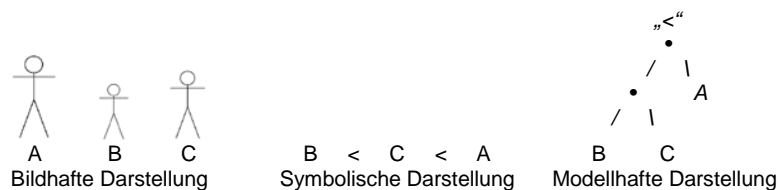
¹ Unter <http://ddi.informatik.uni-erlangen.de/research/kompetenzmodell/> stehen die versendeten Unterlagen (alle Aufgabenprofile, die Beispielprofile, Hinweise zu den Kriterien) zur Verfügung.

Unter der **Erfahrungsweltnähe (EN)** einer Aufgabe soll verstanden werden, wie nah sie dem persönlichen Erleben der Schüler ist. Dabei ist der Aufgabenkontext zu betrachten, nicht der Fachinhalt. Zum Beispiel könnte der Fachinhalt „Scheduling“ in unterschiedlich lebensweltnahe Kontexte eingebettet werden: Der Aufgabenkontext „Stundenplan“ ist dem Schüleralltag zuzuordnen, das heißt *nah* (=1) zur Erfahrungswelt. Der Aufgabenkontext „Fahrplan“ ist nicht unbedingt täglich erlebt, aber jederzeit erlebbar durch die Schüler, das heißt, er weist *mittlere* (=2) Entfernung zur Erfahrungswelt auf. Der Aufgabenkontext „Produktionsplanung“ ist im Allgemeinen nicht erlebbar durch die Schüler, jedoch vorstellbar, das heißt *entfernt* (=3) von der Erfahrungswelt.

Unter dem **Abstraktionsgrad (AG)** einer Aufgabe soll der Grad der Loslösung vom konkreten Beispiel verstanden werden: Lautet die Aufgabe, einen Weg in einem gegebenen Netzwerk zu suchen, so handelt es sich um einen konkreten Fall und die Aufgabe ist vom Abstraktionsgrad *konkret* (=1). Lautet die Aufgabe, mögliche Knotenausfälle in einem Netzwerk durchzuspielen, so handelt es sich um die Betrachtung gedachter Fälle. Die Aufgabe ist von *mittlerem* (=2) Abstraktionsgrad. Gilt es, einen Algorithmus zur Wegsuche in Netzwerken zu finden, so handelt es sich um ein allgemeines Verfahren, losgelöst vom Einzelfall. Die Aufgabe ist *abstrakt* (=3).

Die **Komplexität (KG)** ergibt sich aus der Anzahl der Fälle, Regeln oder Umformungen, die die Größe des Lösungsraums bestimmt. Zum Beispiel könnte eine Aufgabe im Kontext Schachspiel unterschiedlich komplex gestaltet werden: Ist ein einzelner Spielzug einer Figur zu betrachten, gilt die Aufgabe als *einfach* (=1). Sind es die Züge mehrerer Figuren, ist die Komplexität *mittel* (=2). Ist eine Folge von Zügen mehrerer Figuren zu betrachten, gilt die Aufgabe als *komplex* (=3).

Der **Formalisierungsgrad (FG)** ist ein Merkmal der Informationsdarstellung. Ein Text – „Von drei Personen ist A die größte, B die kleinste und C die mittlere.“ – oder ein Bild ist *informell* (=1). Die symbolische Repräsentation weist einen *mittleren* (=2) Formalisierungsgrad auf. Die modellhafte Darstellung, etwa als Baum oder Struktogramm ist *formal* (=3).



Die **Redundanz (RG)** ist ebenfalls ein Merkmal der Informationsdarstellung. Wird die gleiche Information mehrfach dargestellt, etwa als Text – „Von drei Personen ist A die größte, B die kleinste und C die mittlere.“ – und zusätzlich als Bild, so ist die Informationsdarstellung in der Aufgabe *redundant* (=1), sonst *nicht redundant* (=2).

Der **Anforderungsbereich (AB)** bezieht sich auf die geforderte Aktion. Drei Anforderungsbereiche werden in den Einheitlichen Prüfungsanforderungen in der Abiturprüfung Informatik [Ku04] unterschieden: *Wiedergabe* (=1), z. B. das Dualsystem beschreiben, *Anwendung* (=2), z. B. eine Zahl aus dem Dezimalsystem in das Dualsystem übertragen, und *Problemlösung* (=3), z. B. ein Verfahren zur Addition dualer Zahlen entwickeln.

Auch der **Prozessbereich (PB)** bezieht sich auf die geforderte Aktion. Fünf Prozessbereiche werden in den Bildungsstandards Informatik [GI08] unterschieden: *Modellieren und Implementieren* (=1), *Begründen und Bewerten* (=2), *Strukturieren und Vernetzen* (=3), *Kommunizieren und Kooperieren* (=4), *Darstellen und Interpretieren* (=5).

Sechs **kognitive Lernzielstufen (LS)** werden nach der zweidimensionalen Taxonomie von Anderson, Krathwohl et al. [AK01] in der ersten Dimension unterschieden: *Erinnern* (=1), *Verstehen* (=2), *Anwenden* (=3), *Analysieren* (=4), *Bewerten* (=5) und *Erschaffen* (=6).

In der zweiten Dimension der Taxonomie werden vier **Arten des Wissens (AW)** unterschieden: *Fakten* (=1), *Konzepte* (=2), *Prozeduren* (=3) und *Metakognition* (=4).

Die Kommunikation mit den Experten wurde ausschließlich schriftlich geführt. Die erste Kontaktaufnahme fand per E-Mail statt. Die Fragebögen lagen auf Papier vor und sowohl Versand als auch Rücklauf erfolgten per Post. Von den 45 Befragten sind neun weiblich. 21 unterrichten in der Sekundarstufe, davon 14 auch an der Universität. 22 sind nur an der Universität, zwei sind außerhalb der Lehre tätig (Abb. 1).

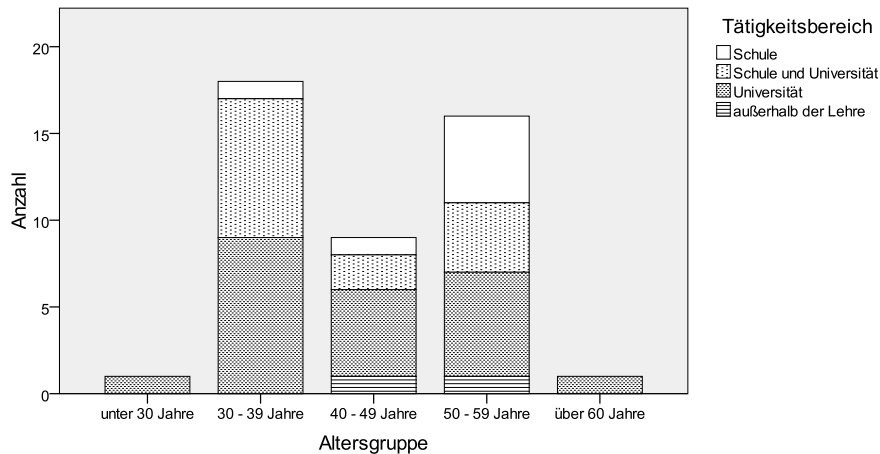


Abbildung 1: Altersgruppe und Tätigkeitsbereich der 45 Befragten

2.2 Datengrundlage

Der Studie liegen die anonymisierten Teilnahmedaten des Informatik-Bibers 2007 zu Grunde. Das Datenmaterial umfasst Jahrgangsstufe, Geschlecht, Bundesland und die Bearbeitungsergebnisse je Aufgabe für 21.802 Teilnehmer. Insgesamt wurden 29 Aufgaben aus allen Inhaltsbereichen laut Bildungsstandards Informatik gestellt (Abb. 2).

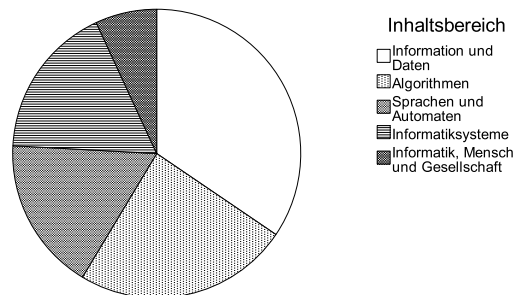


Abbildung 2: Inhaltsbereiche der Aufgaben des Informatik-Bibers 2007

Stärkster Inhaltsbereich ist *Information und Daten* mit zehn Aufgaben, davon zwei zum Thema Datenstrukturen, vier zu Graphen, vier zu Codierung. Auf *Algorithmen* entfallen sieben Aufgaben, auf *Sprachen und Automaten* fünf Aufgaben, eine zu Sprachen, eine zu Automaten, drei zum Logischen Schließen. (Die Bildungsstandards beziehen das Thema Logisches Schließen nicht explizit ein. Als formales Kalkül wurde es hier den *Sprachen*

zugeordnet.) Von den fünf Aufgaben des Inhaltsbereichs *Informatiksysteme* betreffen vier das Thema Anwenderwissen, eine das Thema Datenbanken. *Informatik, Mensch und Gesellschaft* ist mit nur zwei Aufgaben vertreten, beide zum Benutzerverhalten. Die Aufgaben wurden teilweise in mehreren Jahrgangsstufen gestellt. Aus Gründen der inhaltlichen Ausgewogenheit kam es auch vor, dass Aufgaben zum Beispiel in den Stufen 5 bis 7 und ab 11, nicht aber in 8 bis 10 gestellt wurden.

3 Aufgabenklassifizierung: Auswertungsmodalitäten

Die aufgabenweise Sichtung der 225 Aufgabenprofile ergab, dass je Aufgabe zwischen drei und siebzehn Einschätzungen vorliegen. Diese Spanne ist breit, nicht etwa weil sich die Aufgaben so unterschiedlich einschätzbar zeigten, sondern weil Aufgaben, die im Wettbewerb in mehreren Jahrgangsstufen gestellt wurden, entsprechend auch mehreren Experten zur Einschätzung vorlagen, und weil der Rücklauf nicht vollständig war, nur 45 von 54 versendeten Fragebogensätzen wurden zurückgesendet. Zunächst werden die Häufigkeitsverteilungen der Bewertungen je Merkmal betrachtet (Abb. 3).

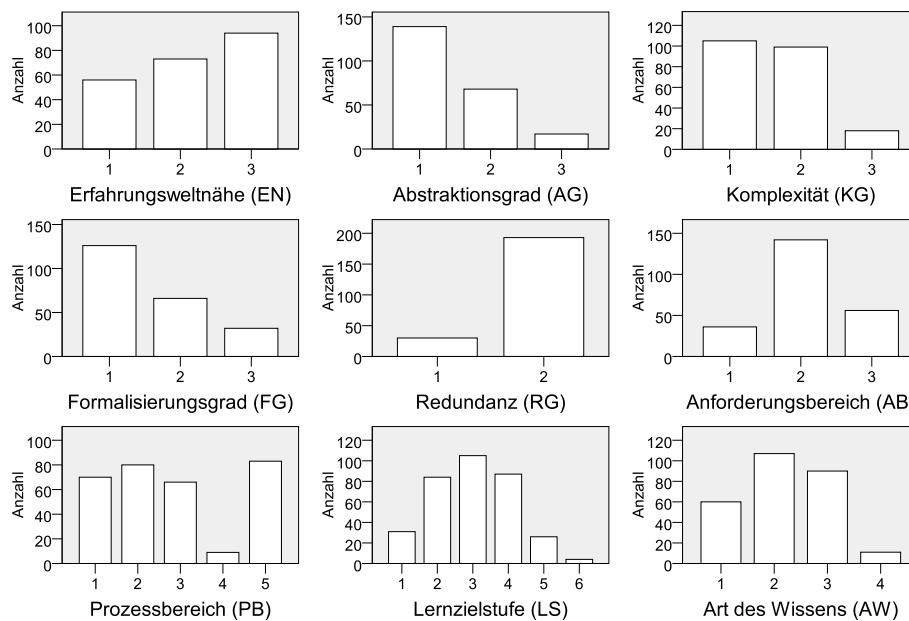


Abbildung 3: Häufigkeitsverteilungen der Merkmale EN bis AW auf der Basis von 225 Experten-Einschätzungen. Bei den Merkmalen AB bis AW waren Mehrfachantworten möglich.

Die Verteilung der Erfahrungsweltnähe ist steigend, die meisten Einschätzungen lauten *entfernt* von der Erfahrungswelt. Die Verteilungen der Merkmale Abstraktionsgrad, Komplexität und Formalisierungsgrad sind fallend, besonders *abstrakt* und *komplex* werden selten genannt. Das trifft auch auf *redundant* zu. Beim Anforderungsbereich fällt auf, dass der mit Abstand größte Anteil auf *Anwendung* entfällt. Die Prozessbereiche

werden gleichermaßen genannt, nur *Kommunizieren und Kooperieren* wird im Quiz-Format kaum angesprochen. Die Einschätzungen der Lernzielstufe sind glockenförmig um den Schwerpunkt *Anwenden* verteilt. Besonders *Erschaffen* ist dem Wettbewerbsformat Multiple-Choice entsprechend selten, wie auch die Wissensart *Metakognition*.

Um nun aus den Einzelbewertungen eine möglichst objektive Gesamtbewertung zu gewinnen, werden die Häufigkeitsverteilungen der Merkmale je Aufgabe ausgewertet. Tabelle 1 zeigt das Gesamtergebnis der Aufgabenklassifizierung.

Tabelle 1: Aufgabenklassifizierung nach den charakterisierenden Merkmalen, schattiert nach der Ausprägung. Durchschnittliche Übereinstimmung und Standardabweichung. (***) 1, 2, 5)

	EN 1-3	AG 1-3	KG 1-3	FG 1-3	RG 1-2	AB 1-3	PB 1-5	LS 1-6	AW 1-4
Bibers Geheimcode	2	1	2	1	2	2	3	3	⊗
Biberzahlen	3	2	2	2	⊗	2	5	3	3
Verwandlung	3	2	2	1	2	3	2 5	4	3
Zahlenreihe	3	1	1	1	2	2	1	3	3
Falschgeld	2	1	2	2	2	2	5	3	3
Dateisuche	2	⊗	2	⊗	2	2	3	2 3	3
Private E-Mail	1	1	1	1	⊗	1	2 4	1	1
Wetter	1	1	1	1	2	2	2 3	2 3	2
Schnitzeljagd	3	1	2	3	2	2	⊗	3 4	3
Endlos-Schleife	3	2	2	2	2	2	1 5	3 4	2 3
Verschlüsselung	3	2	2	2	2	2	***	3 4	2
Wertetausch	2	3	2	3	2	2	⊗	3	2 3
Ungeschützter Computer	1	1	1	1	2	1	2	2	1
Computervirus	1	1	1	1	2	1	2	2	1
Platzwechsel	3	2	2	2	⊗	3	1	3 4	3
Link	1	1	1	1	2	1	4	1	1
Primärschlüssel	3	1	1	2	2	2	2	2	1 2
Sicheres Passwort	1	1	1	1	2	2	2	5	1
Fenster schließen	1	1	1	1	2	1	5	1	1
Labyrinth	2	⊗	2	1	1	2 3	3 5	3 4	2 3
Binärbaum	3	1	2	3	2	2	5	3 4	2
Umparken	2	1	1	2	2	2	1 5	3	2 3
Netzwerkkabel	3	2	2	2	2	3	3 5	4	2
Morse-Code	3	1	2	2	2	2 3	3	4	3
Biber am Fluss	2	2	1	1	2	2	5	4	2 3
Dino-Ordnung	2	1	2	1	2	2	2 3	2 4	2
Links um!	2	1	1	1	2	2	1	3	⊗
POP und PUSH	2	1	2	1	2	2	1 5	3	2 3
Biber und Bisons	3	2	2	1	2	3	⊗	2 4	2
Durchschnittliche Übereinstimmung (%)	73,7	68,4	69,1	72,3	85,8	81,8	68,0	77,6	69,6
Standardabweichung (%)	13,4	20,1	16,9	19,3	18,4	14,2	15,8	17,4	16,7

Für die Merkmale EN bis RG gelten andere Auswertungsmodalitäten als für AB bis AW, bei denen Mehrfachantworten möglich waren. Als Aufgabenbewertung bezüglich EN bis RG wird der Modalwert der Bewertungsverteilung gewählt, also der meistgenannte Wert. Bei zwei benachbarten Modi wird die stärkere Ausprägung gewählt, das heißt *abstrakt*, falls von zehn Befragten je 0/5/5 *konkret/mittel/abstrakt* nennen (a). In allen anderen Fällen wird keine Gesamtbewertung ausgewiesen. Bezüglich AB bis AW ergibt sich die Bewertung aus den übereinstimmenden Nennungen von mindestens 50 % der Befragten, das heißt *Konzepte* und *Prozeduren*, falls von zehn Befragten je 2/8/5/1 die Wissensarten *Fakten/Konzepte/Prozeduren/Metakognition* nennen (b).

Als Maß für die Übereinstimmung der Einschätzungen wurde die prozentuale Nennungshäufigkeit des Modalwertes festgelegt. Das obige Beispiel (a) weist 50 % Übereinstimmung auf (0/5/5 Nennungen der Abstraktionsgrade 1-3 entsprechen bei zehn Befragten 0/50/50 %), und Beispiel (b) 80 % (2/8/5/1 Nennungen der Wissensarten 1-4 entsprechen bei zehn Befragten 20/80/50/10 %). Die durchschnittliche Übereinstimmung für ein Merkmal (Tab. 1) kann einen Hinweis auf dessen Anwendbarkeit und Handhabbarkeit geben. Die Klassifizierung hängt vom Übereinstimmungsmaß ab und wird im ersten Gang mit der Maßgabe durchgeführt, möglichst jeder Aufgabe bezüglich jedes Merkmals eine Bewertung zuzuweisen. Falls sich in der Analyse der empirischen Daten keine deutlichen Ergebnisse abzeichnen, kann zum Beispiel der Parameter Mindestübereinstimmung verändert werden, um eine strengere Klassifizierung zu gewinnen.

Die Aufgaben gehen gemäß der Klassifizierung in den zweiten Teil der Studie ein: die Beispielaufgabe „Biber und Bisons“ (Kap. 2) dient so als Testaufgabe für die Attribute *entfernt* von der Erfahrungswelt, *mittlerer* Abstraktionsgrad, *mittlere* Komplexität, *informell*, *nicht redundant*, Anforderung *Problemlösung*, Lernzielstufen *Verstehen*, *Analyisieren* und *Konzeptwissen*. (Für den Prozessbereich wurde keine übereinstimmende Einordnung erzielt. Von neun Befragten wurden die Bereiche 1-5 mit den Häufigkeiten 3/3/4/0/1 genannt.²) Im Folgenden wird reflektiert, inwieweit die Aufgaben sich der Einordnung nach den Kriterien erschließen, die Kriterien auf Akzeptanz seitens der Befragten stoßen und eine übereinstimmende Aufgabenbewertung möglich ist.

3.1 Erfahrungsweltnähe

Allein nach dem Grad der Übereinstimmung beurteilt, den die Befragten erreichen, scheint das Kriterium Erfahrungsweltnähe (EN) mit einer durchschnittlichen Übereinstimmung von 73,7 % bei einer Standardabweichung von 13,4 % gut anwendbar und handhabbar zu sein – im Durchschnitt gelangen drei von vier Experten zur gleichen Bewertung. Allerdings erfüllt das Merkmal nur bedingt den Anspruch, ausschließlich aufgabenbezogen einschätzbar zu sein, zumindest in der Sekundarstufe, denn die Erfahrungswelt z. B. von Zehn- und Sechzehnjährigen oder Gymnasiasten und Berufsschülern ist unterschiedlich. Dadurch ist die Anwendbarkeit begrenzt auf Aufgaben für bekannte Adressatengruppen. Auffallend ist weiter, dass die Befragten je nach Geschlecht unter-

² Unter <http://ddi.informatik.uni-erlangen.de/research/kompetenzmodell/> stehen die Häufigkeitstabellen für alle Aufgaben und Kriterien zur Verfügung.

schiedlich bewerten (Abb.4). Aus der Sicht der männlichen Befragten ist der größte Anteil, aus der Sicht der weiblichen Befragten der geringste Anteil der Aufgaben *entfernt* (=3) von der Erfahrungswelt. Das schränkt die Handhabbarkeit des Kriteriums ein.

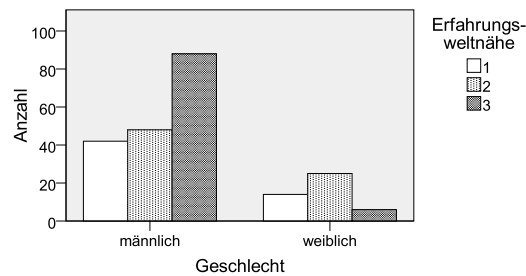


Abbildung 4: Geschlecht der Befragten und Einschätzung der Erfahrungsweltnähe

3.2 Abstraktionsgrad

Der Abstraktionsgrad (AG) hat im Merkmalsvergleich mit 68,4 % die geringste, etwa eine Zweidrittel-Übereinstimmung, bei einer relativ großen Streuung von 20,1 %. Zwar wählen die Experten bei „Zahlenreihe“, „Private E-Mail“, „Link“, „Sicheres Passwort“ zu 100 % übereinstimmend den Abstraktionsgrad *konkret*, aber bei „Biberzahlen“, „Wertetausch“, „Dateisuche“ verteilen sich die Einschätzungen zu 40/40/20 %, 20/40/40 % oder gar 40/20/40 % auf die Stufen *konkret/mittel/abstrakt*. Ein Erklärungsansatz ist, dass auf den Kontext statt auf die eigentliche Arbeitsanweisung Bezug genommen wird.

3.3 Komplexität

Die Komplexität (KG) weist mit 69,1 % eine ähnlich geringe Übereinstimmung auf wie der Abstraktionsgrad, es können aber alle Aufgaben bezüglich der Komplexität eindeutig klassifiziert werden. Zu bedenken bleibt, ob das spezifizierte Kriterium allen Facetten des Kompetenzanspruchs gerecht wird, der aus der Aufgabenkomplexität erwächst, beansprucht doch ein größerer Lösungsraum vielfältige Kompetenzen, zum Beispiel Konzentrations- und Durchhaltevermögen oder strategisches Denken.

3.4 Formalisierungsgrad

Die Bewertung des Formalisierungsgrades (FG) erfolgte mit zufriedenstellender Übereinstimmung (72,3 %), lediglich „Dateisuche“ fand keine eindeutige Bewertung. Die relativ starke Streuung (19,3 %) weist jedoch auf eventuelle Schwächen in der Spezifikation von *informell, mittel und formal* hin. Aus den Anmerkungen der Befragten geht hervor, dass die Stufen umfassend und trennscharf sind, aber ihre Anordnung wird hinterfragt. Eine symbolische Darstellung, beispielsweise eines Automaten als Tupel, weise gegebenenfalls einen höheren Formalisierungsgrad auf als eine modellhafte Repräsentation, etwa als Graph. Erst von der empirischen Überprüfung ist Aufschluss darüber zu erwarten, wie sich der Schwierigkeitsgrad zum Formalisierungsgrad verhält.

3.5 Redundanz

Die Bewertungsverteilung der Redundanz (RG) zeigt bereits, dass die Einschätzung *nicht redundant* deutlich überwiegt (Abb. 3). Unter den gegebenen Auswertungsmodalitäten würden mit 85,8 % durchschnittlicher Übereinstimmung alle Aufgaben als *nicht redundant* klassifiziert. Bei benachbarten Modi die stärkere Ausprägung zu wählen ist hier jedoch unangemessen, weil *redundant* und *nicht redundant* nicht als benachbart, sondern als gegenteilig zu betrachten sind. Deshalb bleibt bei „Biberzahlen“, „Private E-Mail“ und „Platzwechsel“ die Frage der Redundanz unentschieden. Der vorliegende Aufgabensatz, in dem „Labyrinth“ als einzige Aufgabe *redundant* klassifiziert wird, ist nicht geeignet, den Zusammenhang zwischen Redundanz und Aufgabenschwierigkeit zu überprüfen.

3.6 Anforderungsbereich

Der Anforderungsbereich (AB) wurde von den Befragten mit großer Übereinstimmung (81,8 %) bei geringer Standardabweichung (14,2 %) eingeschätzt. Da die Wettbewerbsteilnahme keine Vorkenntnisse voraussetzt, kann das Kriterium aufgabenbezogen, das heißt unabhängig vom Adressaten, beurteilt werden. Für jede Aufgabe werden in der Vorgabe alle erforderlichen Informationen zur Verfügung gestellt, die mit Sachverstand wiederzugeben, anzuwenden oder auf deren Basis das Problem zu lösen ist, um die richtige Antwortmöglichkeit identifizieren zu können. Es besteht ein gewisser Zusammenhang zwischen Anforderungsbereich und Lernzielstufe (s. dazu Abb. 5 und Kap. 3.8).

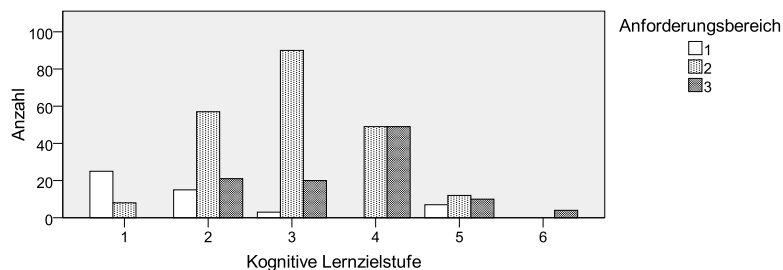


Abbildung 5: Kognitive Lernzielstufe und Anforderungsbereich

3.7 Prozessbereich

Der Prozessbereich (PB) zeigt mit 68 % die geringste Übereinstimmung. Elf der 45 Befragten vermerken, die Einschätzung dieses Merkmals sei mit der größten Unsicherheit erfolgt. Drei Experten geben an, sie hätten regelmäßig rein nach dem Ausschlussprinzip *Modellieren und Implementieren* gewählt. Nichtsdestoweniger zeigt die Bewertungsverteilung, dass die Prozessbereiche, bis auf *Kommunizieren und Kooperieren*, gleichermaßen in Anspruch genommen werden. Ob die resultierende Klassifizierung geeignet ist, einen Einfluss des Unterscheidungsmerkmals Prozessbereich auf die Aufgabenschwierigkeit festzustellen, wird im zweiten Teil der Studie untersucht.

3.8 Lernzielstufe

Die Einordnung der Aufgaben in die Lernzieltaxonomie (LS) erfolgt mit 77,6 % Übereinstimmung. Da den kognitiven Lernzielstufen so wie den Anforderungsbereichen die Taxonomie von Bloom zugrunde liegt, gibt es Überschneidungen in der Bewertung. Die gruppierte Häufigkeitsverteilung (Abb. 5) lässt erkennen, dass beispielsweise die Lernzielstufe *Anwenden* (=3) häufig mit dem Anforderungsbereich *Anwendung* (=2) zusammenfällt, die Bewertungen jedoch nicht zur Deckung gebracht werden können, da die Lernzielstufe *Analysieren* (=4) genau so häufig mit der Anforderung *Anwendung* (=2) zusammenfällt wie mit *Problemlösung* (=3). Welches der beiden Merkmale besser geeignet ist, die Aufgabenschwierigkeit abzubilden, ist eine Fragestellung der geplanten empirischen Untersuchung.

3.9 Art des Wissens

Die Wissensart (AW) wird mit relativ geringer Übereinstimmung von 69,6 % eingeschätzt. Die Aufgaben „Bibers Geheimcode“ und „Links um!“ können nicht eingeordnet werden. Der Wissensart *Metakognition* kann im Gesamtergebnis keine Aufgabe zugeordnet werden (die höchste Bewertung erreicht die Aufgabe „Biber und Bisons“, die den Wissensarten *Fakten/Konzepte/Prozeduren/Metakognition* von neun Befragten mit 2/6/4/3 Nennungen, das entspricht 22/67/44/33 %, zugeordnet wurde).

4 Zusammenfassung und Ausblick

Das Kriterienschema erweist sich als gut handhabbar und wird von den Befragten akzeptiert, wesentliche Kritikpunkte werden in Kapitel 3 diskutiert. Die Aufgaben erschließen sich der Bewertung: lediglich einzelne Aufgaben können bezüglich eines der Kriterien nicht eingeordnet werden, und nur die Lernzielstufe *Erschaffen* sowie die Wissensart *Metakognition* werden im Gesamtergebnis gar nicht erreicht. Bis auf die Erfahrungsweltnähe sind die Einschätzungen unabhängig von Geschlecht, Altersgruppe und Tätigkeitsbereich der Befragten. Einzig die Vertrautheit mit den Aufgaben wirkt sich aus, indem in der Gruppe der Befragten, die mit den Aufgaben des Informatik-Bibers noch nicht vertraut waren, sämtliche Merkmale eine größere Varianz bei gleicher Zentraltendenz aufweisen als in der Gruppe derer, die mit den Aufgaben schon vertraut waren.

Im ersten Teil der Studie wurde eine Klassifizierung der Aufgaben des Schülerwettbewerbs Informatik-Biber nach den charakterisierenden Merkmalen der Schwierigkeit erarbeitet. Darüber wurde in diesem Artikel berichtet. Im zweiten Teil, der noch nicht abgeschlossen ist, werden die Teilnahmedaten des Wettbewerbs, insbesondere die Bearbeitungserfolge, nach der Aufgabenklassifizierung aufgeschlüsselt. Das Ziel ist, zu überprüfen, ob und in welchem Maße die klassifizierenden Merkmale relevante Kriterien der Schwierigkeit von Aufgaben darstellen. Im abschließenden dritten Teil der Studie soll mittels kategorisierender statistischer Analyseverfahren das Schwierigkeitsgefüge als Modell des Kompetenzanspruchs sichtbar gemacht werden.

Danksagung

Danke an die Expertinnen und Experten der Informatik-Didaktik für die sorgfältige Bearbeitung der Fragebögen und besonders auch für die ausführlichen und wertvollen Kommentare. Danke für die freundliche Kooperationsbereitschaft an Dr. Wolfgang Pohl als Geschäftsführer des Bundeswettbewerbs Informatik/Einstieg Informatik, der die anonymisierten Wettbewerbsdaten des Informatik-Bibers 2007 zur Verfügung gestellt hat, und an Eljakim Schrijvers, der den technischen Zugriff auf die Daten ermöglicht hat.

Literaturverzeichnis

- [AK01] Anderson, L.W.; Krathwohl, D.R. (Hrsg.) et al.: A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives. Addison Wesley Longman, New York, 2001.
- [BC82] Biggs, J.B.; Collins, K.F.: Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome). Academic Press, New York, 1982.
- [Bo08] Bower, M.: A Taxonomy of Task Types in Computing. In (ACM Hrsg.): Proceedings of the 13th ITiCSE, Madrid, Spanien, 2008; S. 281-285.
- [Br04] Brinda, T.: Integration of new exercise classes into the Informatics education in object-oriented modelling. In: Education and Information Technologies 9 (2004) 2; S. 117-130.
- [Fr08] Freischlad, S.: Design of Exercises and Test Items for Internetworking Based on a Framework of Exercise Classes. In (Kendall, M.; Samways, B. Hrsg.): Learning to Live in the Knowledge Society. IFIP, Volume 281, Springer, Boston, 2008; S. 261-268.
- [Fu08] Fuller, U. et al.: Developing a Computer Science-specific Learning Taxonomy. In: ACM SIGCSE Bulletin 39 (2007) 4, ACM Press, New York; S. 152-170.
- [GI08] Gesellschaft für Informatik e. V.: Grundsätze und Standards für die Informatik in der Schule. Bildungsstandards Informatik für die Sek. I. LOGIN 28 (2008) 150/151; Beilage.
- [KF07] Kohl, L.; Fothe, M.: Algorithmen aus einer anderen Perspektive – Ein Vorschlag für ein Kompetenzmodell zum Inhaltsbereich „Algorithmen“ der Bildungsstandards Informatik. LOGIN 27 (2007) 146/147; S. 20-22.
- [Ku04] Kultusministerkonferenz (Hrsg.): Einheitliche Prüfungsanforderungen in der Abiturprüfung Informatik. Luchterhand, München, 2004.
- [La07] Lahtinen, E.: A Categorization of Novice Programmers: A Cluster Analysis Study. In (Sajaniemi, J.; Tukiainen, M. Hrsg.): Proceedings of the 19th annual Workshop of the Psychology of Programming Interest Group, Joensuu, Finnland, 2007; S. 32-41.
- [PF07] Puhlmann, H.; Friedrich, S.: Die Standards – und wie weiter? Zur Beurteilung von Aufgaben für Unterricht und Test. LOGIN 27 (2007) 146/147; S. 16-19.
- [Po09] Pohl, W. et al.: Informatik-Biber: Informatik-Einstieg und mehr. Im Tagungsband zur 13. GI-Fachtagung „Informatik und Schule“, INFOS 2009 in Berlin. Bonn, Köllen, 2009.
- [Sc08] Schlüter, K.: Je schwieriger die Aufgabe, desto klüger der Kopf? In (Brinda, T. et al. Hrsg.): Didaktik der Informatik – Aktuelle Forschungsergebnisse, 5. Workshop der GI-Fachgruppe "Didaktik der Informatik", Erlangen 2008. Bonn, Köllen, 2008; S. 77-86.
- [Sh08] Sheard, J. et al.: Going SOLO to Assess Novice Programmers. In (ACM Hrsg.): Proceedings of the 13th ITiCSE, Madrid, Spanien, 2008; S. 209-213.
- [St07] Steinert, M.: Lernzielgraphen und Lernerfolgsanalyse. In (Schubert, S. Hrsg.): Didaktik der Informatik in Theorie und Praxis, INFOS 2007. Bonn, Köllen, 2007; S. 147-158.
- [St08] Starr, C.W. et al.: Bloom's Taxonomy Revisited: Specifying Assessable Learning Objectives in Computer Science. In: ACM SIGCSE Bulletin 40 (2008) 1, ACM Press, New York; S. 261-265.