Engel, Joachim; Sedlmeier, Peter

Zum Verständnis von Zufall und Variabilität in empirischen Daten bei Schülern

Unterrichtswissenschaft 32 (2004) 2, S. 169-191

Quellenangabe/ Reference:
Engel, Joachim; Sedlmeier, Peter: Zum Verständnis von Zufall und Variabilität in empirischen Daten bei Schülern - In: Unterrichtswissenschaft 32 (2004) 2, S. 169-191 - URN: urn:nbn:de:0111-opus-58121 - DOI: 10.25656/01:5812

https://nbn-resolving.org/urn:nbn:de:0111-opus-58121
https://doi.org/10.25656/01:5812

Unterrichtswissenschaft

Zeitschrift für Lernforschung
32. Jahrgang / 2004 / Heft 2

Thema
Neue Lehr-Lernkultur

Verantwortlicher Herausgeber
Frank Achtenhagen

Joachim Engel, Peter Sedlmeier

Zum Verständnis von Zufall und Variabilität in empirischen Daten bei Schülern

Students' Understanding of Chance and Variability in Empirical Data

Im Zentrum statistischen Denkens steht der Umgang mit Variabilität in empirischen Daten und die Fähigkeit, mit Hilfe des Zufallsbegriffs nicht erklärte Variation in Daten zu modellieren. Die vorliegende Studie berichtet von einer Querschnittsuntersuchung mit 222 Schülern der Klassenstufen 5, 7 und 9 zur Entwicklung des Verständnisses von Zufall und Variabilität. Dazu wurde anhand geeigneter Aufgaben untersucht, wie stark statistische Kompetenz bei Schülern ausgeprägt ist und ob sie sich mit zunehmenden Alter der Schüler verändert. Die Ergebnisse wurden zu vergleichbaren früheren Untersuchungen bei Kindern und bei Erwachsenen in Beziehung gesetzt. Die Studie replizierte in modifizierter Form frühere Untersuchungen in anderen Ländern und erbrachte vergleichbare Ergebnisse für deutsche Schüler. Insbesondere liegen keinerlei Anzeichen für eine Verbesserung des Verständnisses von Zufall und Variabilität mit zunehmendem Alter der Schüler vor. Die Ergebnisse sind auch konsistent mit Befunden aus der Urteilsforschung bei Erwachsenen.

Dealing properly with variability in empirical data lies at the core of statistical reasoning and thinking. A key idea is to use chance processes to model unexplained variation in data. The paper reports the results of a study with 222 middle school students attending grades 5, 7 and 9 about the development of their concepts of chance and variability. Based on items similar to those used in previous studies we investigated how students' statistical literacy evolves with increasing age. Our results are discussed and related to earlier studies with children and with adults. Our study confirms earlier results found with students from other countries. In particular, there are no indications of an improvement with increasing age. Our findings are consistent with common findings in the area of judgment under uncertainty.

1. Zufall, statistische Kompetenz und Intuition

„Im Leben ist nichts gewiss außer dem Tod und den Steuern."

Dieses Zitat von Benjamin Franklin (1706-1790, US-Präsident) weist darauf hin, dass Unsicherheit und Ungewissheit zu den Wesensbedingungen der menschlichen Existenz gehören. Im Alltag sind wir ständig mit Situationen konfrontiert, in denen wir handeln und Entscheidungen treffen müssen, ohne dass uns alle Informationen und Einsichten verfügbar sind, um mit Bestimmtheit die Konsequenzen unseres Tuns absehen zu können. Wir können uns kundig machen, Erfahrungen berücksichtigen und Daten über die von uns zu bewältigende Situation mit einbeziehen, aber absolute Sicherheit haben wir kaum. In Kulturen und Epochen, in denen der Wahrscheinlichkeitsbegriff wenig entwickelt ist, werden häufig unangemessene Erklärungsversuche und Sinndeutungen wie Aberglaube, Magie oder moralische Sinngebungen (Lohn und Strafe für begangene Taten) für die Erklärung von Geschehnissen herangezogen, wenn keine unmittelbaren und zwingenden kausalen Ursachen identifiziert werden können. Zufälligkeit ist das Gegenstück zur Kausalität. Bei der Aufnahme und Verarbeitung von empirischer Information erweist sich der Zufallsbegriff als sehr nützlich. Wir lesen einen Trend oder ein Muster aus vorliegenden Daten und erklären alles was übrigbleibt, d.h. die Differenz zwischen Daten und Trend - die Residuen -, als Zufall. Dadurch ist das Zufallsverständnis stark mit der Kompetenz zum statistischen Denken verbunden.

Die praktische Relevanz dieser Kompetenz zum statistischen Denken lässt wünschen, dass ihr in der schulischen Ausbildung (und nicht nur im Rahmen des Mathematikunterrichts) besondere Aufmerksamkeit geschenkt wird. In der folgenden Studie untersuchen wir anhand geeigneter Aufgaben, wie stark diese statistische Kompetenz bei Schülern[1] ausgeprägt ist und wie sie sich im Verlauf der Schulausbildung verändert. Einige Ergebnisse aus der Entwicklungspsychologie geben dabei Anlass zu Optimismus. Verschiedene Untersuchungen weisen darauf hin, dass Kinder schon früh über valide Intuitionen zu Wahrscheinlichkeiten verfügen: Schon im Vorschulalter sind Kinder sensitiv für das Wirken von Zufall (siehe z.B. Kuzmak & Gelman, 1986) und schon ab 8 Jahren wissen die meisten Kinder intuitiv, dass konjunktive Mengen (z.B. Blumen, die gelb sind und Primeln sind) nicht größer sein können als Komponenten (im Beispiel: Primeln) (Inhelder & Piaget, 1959/ 1964). Bereits ab 11 Jahren zeigen Kinder ein Verständnis des empirischen Gesetzes der großen Zahlen (Piaget & Inhelder, 1951/ 1975). Diese Ergebnisse deuten auf eine fortschreitende Verbesserung im statistischen Urteilen hin. In der Literatur zum statistischen Urteilen bei Erwachsenen verändert sich jedoch das Bild dramatisch: Anhand zahlreicher Beispiele wird demonstriert, dass Erwachsene große Schwierigkeiten in der adäquaten Wahrnehmung von Wahrscheinlichkeiten, der Konjunktion von Wahrscheinlichkeiten oder der Rolle des Stichprobenumfangs beim empirischen Gesetz der großen Zahl haben (Kahneman, Slovic & Tversky, 1982; Pittelli-Palmarini, 1994). Dieses pessimistische Bild wurde mittlerweile zwar etwas relativiert (z.B. Gigerenzer, 1996), aber bei bestimmten

1 Im Folgenden sind damit Schülerinnen und Schüler gemeint.

Arten von Aufgaben können fehlerhafte Urteile immer wieder repliziert werden. Wie kommt es zu diesen Problemen bei Erwachsenen? Könnte die schulische Ausbildung mit dieser Verschlechterung der statistischen Kompetenz zu tun haben?

Dafür gibt es tatsächlich einige Indizien. Wenn Kinder in die Schule kommen, dann haben sie valide primäre Intuitionen über Zufall - die Schule scheint sie ihnen aber wieder auszutreiben, weil die Lehrerinnen und Lehrer ihnen die Welt deterministisch erklären: „*The child is taught [in school] that explanation consists in specifying a cause; that a scientific prediction must be a certainty; that ambiguity and uncertainty are not acceptable in scientific reasoning and so on. Even if all this is not explicitly stated it is implied in all that is taught in school*" (Fischbein, 1975, p. 71). Fischbein basierte seine Schlussfolgerungen auf Studien zur Intuition über den Wahrscheinlichkeitsbegriff bei Schülern im Alter zwischen 5 und 14 Jahren. Weitere Evidenz für allgemeine Missverständnisse beim statistischen Urteilen, die auf den ungünstigen Einfluss der Schulausbildung hinweisen, stammen von Shaugnessy (1992), Batanero, et al. (1994) und Green (1982, 1986, 1990).

Trägt die Schule zur Entwicklung von angemessenen Intuitionen über Zufall bei oder fördert sie bei den Schülern eher unangemessenes deterministisches Kausaldenken, wie Fischbeins Zitat suggeriert? Wenn Letzteres der Fall ist, dann sollten sich die Leistungen der Schüler in geeigneten Aufgaben über die Schuljahre hinweg nicht verbessern, sondern in der Tendenz eher schlechter werden. Wir gehen der Hypothese nach, dass die schulische Ausbildung einen wenig fördernden Einfluss auf das statistische Denken hat. Mit der Konzentration auf Schüler der Sekundarstufe I wählen wir eine vergleichbare Altersgruppe wie Green (1982). Dazu setzen wir unterschiedliche Aufgaben ein. Unsere Studie repliziert in modifizierter Form und bestätigt für deutsche Schüler schon in Untersuchungen in anderen Ländern gefundene Resultate.

2. Auswahl der Aufgaben

Bei der Auswahl von Aufgaben zu unserer Fragestellung war Vorsicht geboten, damit es zu keiner Interferenz zwischen im Verlaufe der Schuljahre allgemein zunehmender mathematischer Fähigkeit und der uns speziell interessierenden Kompetenz im Umgang mit Zufallsphänomenen kommt. Daher wurden Aufgaben ausgeschlossen, deren Lösungswege mit allgemeiner im Verlauf der Schulzeit erworbener mathematischer Kompetenz (z.B. Umgang mit Verhältnissen oder Brüchen) korrelieren. Außerdem wählten wir aus Gründen der Vergleichbarkeit und Generalisierbarkeit Aufgaben, die in der Literatur schon bei anderen Untersuchungen verwendet wurden und die unterschiedliche Aspekte des Urteilens mit Wahrscheinlichkeiten abdecken.

Die ausgewählten Aufgaben beziehen sich auf die Erstellung einer typischen Zufallsfolge, Vorstellungen über die Zufallsverteilung in der Ebene, auf das empirische Gesetz der großen Zahl und auf die Wahrscheinlichkeit von konjunktiven Ereignissen.

2.1 Zufallsfolgen

Eine für viele Menschen konzeptionell schwer zu fassende Idee ist die Unabhängigkeit von Ereignissen. Wenn auch mathematisch auf der Kalkülebene einfach zu definieren durch die Eigenschaft, dass die Wahrscheinlichkeit eines konjunktiven Ereignisses das Produkt der Wahrscheinlichkeiten der beiden Einzelereignisse ist, so zeigt die kognitionspsychologische Forschung (siehe etwa Nickerson, 2002) hier massive Verstehensprobleme auf. Steinbring (1986) weist darauf hin, dass die Idee der Unabhängigkeit einen theoretischen Charakter hat, dessen Anwendung in einem konkreten Kontext Schwierigkeiten bereitet. Ein viel zitiertes Beispiel ist die „Gamblers fallacy“: Viele Menschen glauben, dass beim Werfen von Münzen nach einer Anzahl von vorangegangenen Ereignissen „Zahl“ die Wahrscheinlichkeit für „Wappen“ steigt. Um das Verständnis der Unabhängigkeit einer Zufallsfolge zu erfassen, forderten wir Schüler auf, eine ihnen typisch erscheinende Abfolge von Münzwürfen zu notieren.

Aufgabe 1: Stell Dir vor, Du wirfst zwanzig mal eine Münze. Bei jedem Wurf liegt entweder Wappen (W) oder Zahl (Z) oben. Wie könnte das Ergebnis der 20 Würfe aussehen? Welche Folge von W und Z könnte sich ergeben? Trage eine Dir typisch erscheinende Folge ein!

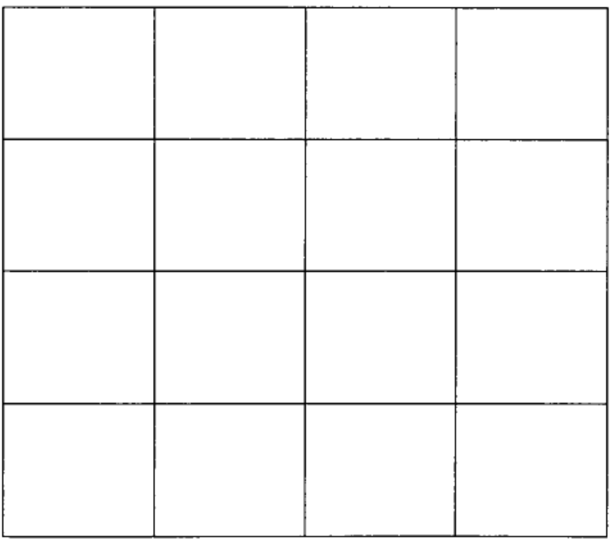
2.2 Zufallsverteilungen

Spielt in der vorangegangen Frage die zeitliche Abfolge von eintreffenden Ereignissen eine besondere Rolle, so geht es im nächsten Item um die ebene Verteilung von Zufallsresultaten. Trotz globalen Regelmäßigkeiten lassen sich Zufallsereignisse einzeln nicht voraussagen. Wie entscheiden Schüler im Konflikt zwischen den Vorgaben einer globalen Gleichverteilung und dem unvorhersehbaren Verhalten einzelner Elemente bzw. einer kleinen Stichprobe?

Die folgende Aufgabe geht auf Piaget & Inhelder (1975) zurück und wurde in verschiedenen Varianten von Green (1982, 1986, 1990) an Schülern getestet. Ausgangspunkt ist ein Alltagsphänomen, das auch schon von Kindern beobachtet wird. Wenn es anfängt zu schneien, wie verteilen sich die ersten Schneeflocken innerhalb eines überschaubaren Gebietes? Der kognitive Konflikt ist provoziert durch die Unvorhersehbarkeit des Verhaltens jeder einzelnen Schneeflocke einerseits mit der Erfahrungstatsache andererseits, dass nach einer gewissen Zeit des Schneefalls der Schnee doch im Wesentlichen gleich hoch auf einem Flachdach liegt. Das erstaunliche an den Ergebnissen von Piaget & Inhelder und Green ist nicht nur die Tatsache, dass sich sehr viele Schüler für eine streng regelmäßige Verteilung der Schneeflocken entschieden, die nichts mit Zufall zu tun hat, sondern gerade auch die Feststellung, dass die Tendenz zu strenger Regelmäßigkeit mit dem Alter der Schüler steigt. Green (1982) leitet hieraus die Hypothese von zwei gegensätzlich gerichteten Tendenzen im Schülerdenken ab: Reifung und zunehmende Erfahrungen versus der Dominanz mechanistisch-deterministischen Denkens, das eine angemessene Respektierung von Zufall behindert, indem es alles kausal zu erklären versucht.

Im Gegensatz zu Green, der die Aufgabe im multiple-choice Format mit jeweils verschiedenen Mustern zur Auswahl vorlegte, präsentierten wir den Schülern die Aufgabe in einer offenen Version, um durch spezielle Vorgaben die Perspektive nicht einzuengen. Ziel dabei ist es, herauszufinden, welche Vorstellungen Kinder von Zufallseffekten haben und wie sie sich zwischen zufälliger Variation und gleichmäßiger Verteilung der Schneeflocken entscheiden. Die Aufforderung zu einer verbalen Erklärung sollte dabei noch zusätzlich Einblick in die Gedankenwelt der Schüler geben.

Aufgabe 2: Das quadratische Flachdach einer Gartenhütte besteht aus 16 gleich großen quadratischen Platten.



Es beginnt zu schneien. Nach einer kurzen Weile sind 16 Schneeflocken auf dem Dach gelandet. Bitte trage ein, wo die 16 Schneeflocken hingefallen sein könnten (ein x für eine Schneeflocke).

Erkläre Deine Antwort:

2.3 Zum empirischen Gesetz der großen Zahl

In der Literatur dazu, wie sensitiv Kinder und Erwachsene für den Einfluss der Stichprobengröße bei statistischen Urteilen sind, trifft man auf eine erstaunliche Inkonsistenz: Kinder zeigen laut Piaget und Inhelder (1975) schon mit 11 oder 12 Jahre eine deutliche Sensitivität für den Einfluss der Stichprobengröße, während bis vor kurzem die vorherrschende Meinung in der Urteilsforschung war, dass Erwachsene die Größe einer Stichprobe in relevanten Urteilen nicht beachten (z.B. Kahneman & Tversky, 1972; Piattelli-Palmarini, 1994). Eine nähere Analyse der Ergebnisse bei Erwachsenen zeigt jedoch, dass in der Urteilsforschung zwei unterschiedliche Arten von Aufgaben verwendet wurden (Sedlmeier, 1998; Sedlmeier & Gigerenzer, 1997): Zum einen handelt es sich um *Häufigkeitsverteilungs*-Aufgaben, in denen Vertrauensurteile zu Anteilen oder Mittelwerten aus einer Stichprobe (einer Häufigkeitsverteilung) verlangt wurden und zum andern um *Stichprobenverteilungs*-Aufgaben, in denen die Variation von empirischen Stichprobenverteilungen beurteilt werden musste. Häufigkeitsverteilungs-Aufgaben (zu denen auch die Aufgaben von Piaget und Inhelder gerechnet werden können) wurden generell richtig gelöst, Stichprobenverteilungs-Aufgaben jedoch nicht. Sedlmeier & Gigerenzer (1997) erklären das damit, dass Menschen über eine Intuition verfügen, die dem empirischen Gesetz der Großen Zahlen entspricht, die auf Häufigkeits- aber nicht auf Stichprobenverteilungs-Aufgaben anwendbar ist (siehe Sedlmeier, 1999; 2002, für eine lerntheoretische Erklärung). In der gegenwärtigen Studie sollte nun untersucht werden, ob sich der Unterschied in den Lösungsraten zwischen Häufigkeits- und Stichprobenverteilungs-Aufgaben auch schon bei Kindern und Jugendlichen zeigt und ob sich die Sensibilität für den Einfluss der Stichprobengröße im Verlauf der Schulzeit verändert. Die verwendeten Aufgaben waren Variationen des „Maternity-Ward" Problems von Kahneman & Tversky (1972). Die Stichprobenverteilungs-Version des Problems war:

Aufgabe 3: Ungefähr die Hälfte aller Neugeborenen sind Mädchen, die andere Hälfte Jungen.

Krankenhaus A: im Durchschnitt werden 3 Kinder pro Tag geboren.

Krankenhaus B: im Durchschnitt werden 5 Kinder pro Tag geboren.

In welchem Krankenhaus wird es im Laufe eines Jahres mehr Tage geben, an denen alle Geburten Mädchen sind? Kreuze an!

☐ Krankenhaus A (mit 3 Geburten pro Tag)

☐ Krankenhaus B (mit 5 Geburten pro Tag)

☐ In beiden Krankenhäusern sind die Chancen dafür ungefähr gleich

In der Häufigkeitsverteilungsversion erhielten die Schüler stattdessen die Frage:

In welchem Krankenhaus wird es an einem bestimmten Tag eher vorkommen, dass alle Geburten Mädchen sind? Kreuze an!

Die Häufigkeitsverteilung bezieht sich in diesem Problem auf die Verteilung von Jungen und Mädchen in einer Stichprobe, d.h. den Geburten an einem Tag, während die Stichprobenverteilung aus den 365 Geburtenanteilen der Mädchen über ein Jahr hinweg besteht.

2.4 Wahrscheinlichkeit von Konjunktionen

Auch in Bezug darauf, ob die Wahrscheinlichkeit von Konjunktionen richtig eingeschätzt wird, gibt es unterschiedliche Ansätze in der Literatur. So fanden Inhelder und Piaget (1959/1964), wie schon erwähnt, dass Kinder im Alter von 8 Jahren durchaus in der Lage waren, zu erkennen, dass konjunktive Mengen nicht größer sein können als Komponentenmengen. Wieder war jedoch, initiiert durch die Ergebnisse einer Studie von Tversky und Kahneman (1983), die vielfach repliziert wurde, die vorherrschende Meinung in der Urteilsforschung, dass Erwachsene generell nicht in der Lage sind, die Wahrscheinlichkeiten von Konjunktionen richtig einzuschätzen. Hertwig & Gigerenzer (1999) konnten jedoch zeigen, dass die fehlerhafte Einschätzung der Wahrscheinlichkeit von Konjunktionen durch semantische Ambiguitäten bei der Aufgabenformulierung (z.B. wie das Wort „und" zu verstehen sei) zustande kommen kann. Eine Präsentation der Aufgabeninformation in Form von Häufigkeiten, bei der die Ambiguität in der Formulierung ausgeschaltet war (wie bei Inhelder & Piaget) führte dagegen zu Lösungsraten von bis zu 100%. Es besteht allerdings nach wie vor kein Konsens über den genauen Wirkmechanismus einer Häufigkeitsformulierung (Mellers, Hertwig & Kahneman, 2001). Wir verwendeten in unserer Studie eine Aufgabe, die dieselbe Struktur hat, wie die in der Urteilsforschung verwendeten. Zusätzlich zum Darbietungsformat (Wahrscheinlichkeiten vs. Häufigkeiten) variierten wir auch die zeitliche Perspektive (s.u.). In der (üblichen) Wahrscheinlichkeitsversion lautete die Aufgabe (Version A)

Aufgabe 4:
Ein Schüler hat im Abschlusszeugnis in Mathematik die Note Vier erhalten. Welche der folgenden Aussagen trifft für ihn eher zu?

Er hatte eine Sechs in Mathe im Halbjahreszeugnis
Er hatte im zweiten Halbjahr Nachhilfe in Mathe erhalten und hatte eine Sechs in Mathe im Halbjahreszeugnis.

In der korrespondierenden Häufigkeitsversion sah die Aufgabe wie folgt aus (Version B)

30 Schüler eines Jahrgangs haben im Abschlusszeugnis in Mathematik die Note Vier erhalten. Welche der folgenden Aussagen trifft für mehr dieser Schüler zu?

Sie hatten eine Sechs in Mathe im Halbjahreszeugnis.
Sie hatten im zweiten Halbjahr Nachhilfe in Mathe erhalten und hatten eine Sechs in Mathe im Halbjahreszeugnis.

Der Grund für die Variation der zeitlichen Perspektive war, dass die semantische Ambiguität der Problemformulierung größer sein könnte, wenn die zu beurteilenden Aussagen zeitlich *vor* der Problembeschreibung liegen (wie in Version A und B), als wenn sich diese Aussagen auf (von der Problembeschreibung aus) zukünftige Ereignisse beziehen. Die entsprechenden „zukunftsgerichteten" Versionen waren

in Version C:

Ein Schüler hatte im Halbjahreszeugnis in Mathematik die Note Sechs erhalten. Welche der folgenden Aussagen trifft für ihn eher zu?

Er hat eine Vier in Mathe im Abschlusszeugnis.
Er hat im zweiten Halbjahr Nachhilfe in Mathe erhalten und hat eine Vier in Mathe im Abschlusszeugnis.

in Version D:

30 Schüler eines Jahrgangs hatten im Halbjahreszeugnis in Mathematik die Note Sechs erhalten. Welche der folgenden Aussagen trifft für mehr dieser Schüler zu?

Sie haben eine Vier in Mathe im Abschlusszeugnis.
Sie haben im zweiten Halbjahr Nachhilfe in Mathe erhalten und haben eine Vier in Mathe im Abschlusszeugnis.

3. Methode

3.1 Versuchsteilnehmer:
Um unsere These zu prüfen, wurden im September 2002 insgesamt 222 Schülerinnen und Schüler der Klassenstufen 5, 7 und 9 aus Hauptschule, Realschule und Gymnasium untersucht. Dazu wurde jeweils einer ganzen Klasse der jeweiligen Stufe und des jeweiligen Schultyps im Großraum Stuttgart vier Aufgaben vorgelegt. Die genaue Zusammensetzung der Stichprobe ergibt sich aus Tabelle 1. Hinsichtlich der Variablen Alter und Geschlecht unterschieden sich die Klassen pro Stufe nicht wesentlich. Die jeweiligen Durchschnittsalter betrugen 10,6 Jahre für die Fünftklässler, 12,7 für die Siebtklässler und 14,9 für die Neuntklässler. Insgesamt wurden 106 Jungen und 116 Mädchen getestet. Hinsichtlich des Leistungsvermögens der Klassen (bezogen auf das Fach Mathematik) wurde bei der Auswahl der Klassen darauf geachtet, dass sie nach Lehrerurteil einem mittleren Niveau des jeweiligen Schultyps entsprechen.

Die drei Schulgattungen wurden ausgewählt, um in etwa eine repräsentative Stichprobe aller Schüler zu erhalten. In Baden-Württemberg besuchen die Schüler in den entsprechenden Altersgruppen zu etwa gleichen Teilen Hauptschule, Realschule und Gymnasium. Die Schüler wurden nach den landesüblichen Lehrplänen für öffentliche Schulen unterrichtet. Da Stochastik in Baden-Württemberg nur in Gymnasien (und dort erst in Klasse 10) unterrichtet wird, verfügten die Schüler über keine unterrichtlich-relevanten Vorerfahrungen im Bereich Zufallsexperimente und Simulationen.

Tab. 1: Darstellung der Stichprobe

	Hauptschule	Realschule	Gymnasium	Summe
Klasse 5	25	24	27	76
Klasse 7	23	26	28	77
Klasse 9	25	25	19	69
Summe	73	75	74	222

Zur differentiellen Analyse wurden bei Aufgabe 3 und 4 (für die Schüler nicht erkennbar) verschiedene Versionen des Fragebogens eingesetzt über deren Verwendung die Schüler nicht informiert wurden.

4. Ergebnisse & Diskussion

Unsere Hauptfragestellung bezieht sich auf die Veränderung der statistischen Kompetenz über die Schulzeit hinweg. Das Hauptergebnis pro Aufgabe war deswegen der Trend über die Schuljahre 5, 7 und 9 hinweg. Dabei wurde meist nicht zwischen den einzelnen Schultypen differenziert, da sowohl in der Hauptschule als auch auf dem Gymnasium Selektionseffekte auftreten, die sich aber gegenseitig kompensieren sollten. In der Hauptschule wird mit steigender Klassenstufe ein immer größerer Anteil von Schülern sein, die die Kriterien für einen Übertritt in weiterführende Schulen nicht erfüllen, während die Selektion im Gymnasium (und teilweise wohl auch in der Realschule) umgekehrt sein wird: Schüler, die die Anforderungen nicht erfüllen, wechseln wieder zurück in andere Schulen. Insofern wäre eine isolierte Betrachtung des Verlaufs bei einzelnen Schulgattungen schwierig zu interpretieren. Trotzdem analysierten wir die Daten auch hinsichtlich systematischer Unterschiede bezüglich der einzelnen Schultypen und berichten diese Ergebnisse in den Fällen, in denen tatsächlich solche Unterschiede auftraten.

Für die Auswertung der beiden ersten Aufgaben, der Produktion von Zufallsreihen und Zufallsverteilungen bieten sich mehrere Methoden an (e.g., Nickerson, 2002; Green, 1990). Deswegen wird diesen Aufgaben mehr Platz gewidmet als den Aufgaben zum empirischen Gesetz der großen Zahlen und zur Wahrscheinlichkeit von konjunktiven Ereignissen.

4.1 Produktion von Zufallsreihen

Die Antworten der Schüler zu Aufgabe 1 wurden unter vier Gesichtspunkten analysiert, der relativen Häufigkeit von „Wappen", der relativen Übergangshäufigkeiten, der Anzahl der Runs und der Länge des längsten Runs.

Verteilung der Anzahl der Wappen innerhalb der Münzwurfserie von 20 Würfen

Ein Vergleich der Mittelwerte der Anzahl der Wappen für die jeweiligen Klassenstufen und Schultypen sollte prüfen, ob die Schüler eine der beiden Resultate Wappen oder Zahl bevorzugt wählen oder ob sie sich gleich oft für Wappen und Zahl entscheiden.

Tab. 2: Durchschnittlich gewählte Anzahl von Wappen. Die Werte in Klammern sind die Mittelwerte unter Ausschluss von zwei Ausreißern.

	Klasse 5	Klasse 7	Klasse 9	Gesamt
Hauptschule	9,96	9,87	10,04	9,96
Realschule	9,75	10,8	10,2	10,01
Gymnasium	9,26 (9,62)	10,0 (10,26)	9,73	9,66 (9,89)
Gesamt	9,64 (9,77)	9,98 (10,08)	10,01	9,88 (9,95)

Aus Tabelle 2 lässt sich schließen, dass die Schüler über alle Klassenstufen und Schultypen hinweg in etwa in gleicher Zahl Wappen und Zahl produzierten, und dabei sehr nahe am theoretischen Wert 10 liegen. Wird der Fünftklässler des Gymnasiums, dessen Münzwurf aus 20 Z besteht und der Siebtklässler (Gymnasium) mit 17 Z aus dem Datensatz entfernt, so liegen die Mittelwerte noch näher am Wert 10. Als Median ergab sich der Wert 10 sogar in allen Klassen über alle Schultypen hinweg. Es gibt also keine bedeutsame Bevorzugung von Zahl über Wappen. Green (1990) berechnet hier bei seinen ganz ähnlichen Untersuchungen auch die jeweilige Standardabweichung und stellt dann fest, dass sie konsistent unter dem theoretisch korrekten Wert von

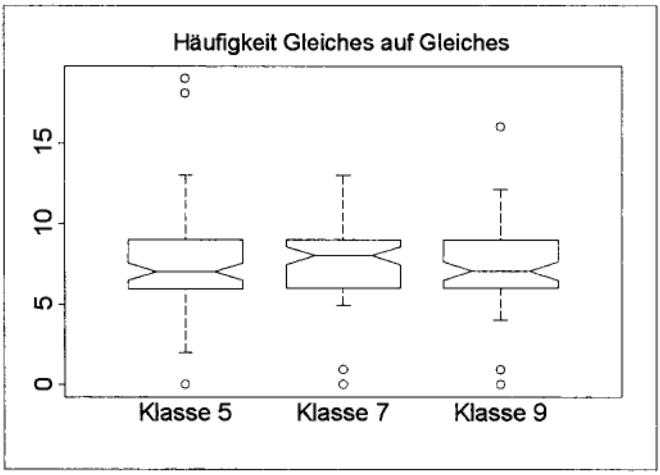
$$\sqrt{20 \cdot \frac{1}{2} \cdot \frac{1}{2}} \approx 2{,}23$$

liegt. Daraus wird dann der Schluss gezogen, dass die zufällige Variabilität unterschätzt wird, d.h. Schüler wählen zu häufig 9, 10 oder 11 Wappen, also Werte, die sich zu sehr um den Erwartungswert 10 häufen, um zufällige Variabilität zu repräsentieren. Wir halten diesen Schluss hier für anfechtbar, da jeder einzelne der ca. 25 Schüler pro Klasse aufgefordert war, *eine einzige* ihm typisch erscheinende Münzwurfserie zu notieren, was konzeptionell zu unterscheiden ist von der Aufforderung, 25 verschiedene Münzwurfserien zu notieren. Während letzteres Design geeignet ist, die individuell beigemessene Variabilität zu messen, ist es bei unserer (und bei Green's) Frage nur allzu logisch, dass die einzige von einem Schüler notierte Münzwurfserie die wahrscheinlichste Anzahl von Wappen, nämlich 10, hat.

Analyse der relativen Übergangshäufigkeiten:
Wie oft folgt Gleiches auf Gleiches?

Bezog sich das obige Kriterium auf die Gesamtzahl von Wappen, so stehen in den folgenden Kriterien die Aufeinanderfolge von Wappen und Zahl im Zentrum der Analyse. Ein wichtiger Gesichtspunkt dabei ist die Unabhängigkeit der einzelnen Münzwürfe. Da die Münze „kein Gedächtnis" hat, ist die Wahrscheinlichkeit für Wappen bei jedem Wurf immer 0,5, egal welche Vorgeschichte an Ergebnissen sich realisiert hat. Hier versagt die Intuition bei vielen Menschen. Hat sich z.B. in einer Münzwurfserie fünf Mal hintereinander Wappen ergeben, so erwartet - wie in vielen Studien dokumentiert - die Mehrzahl der Erwachsenen beim nächsten Münzwurf mit höherer Wahrscheinlichkeit das Ereignis Zahl („*Jetzt ist aber Zahl dran*"). Die menschliche Intuition tut sich offensichtlich sehr schwer mit dem Begriff der stochastischen Unabhängigkeit, hier repräsentiert durch die Irregularität der Zufallsfolgen. Um das Verständnis der Irregularität von Zufallsfolgen zu beurteilen, berechneten wir für jeden Schüler die jeweilige Übergangshäufigkeiten, mit der gleiche Symbole (W bzw. Z) aufeinander folgen. Beispielsweise kommen in der aus acht Symbolen bestehenden Reihe WWZWWZZZ unter den sieben Übergängen vier Übergänge von Gleichem zu Gleichen vor, nämlich zunächst W auf W, dann wiederum W auf W und schließlich zwei Übergänge von Z auf Z. Somit ergibt sich eine relative Häufigkeit für „Gleiches folgt auf Gleiches" von 4/7. Echte Unabhängigkeit wird hier repräsentiert durch einen Wert der relativen Häufigkeit nahe 0,5. Eine aus 20 Würfen bestehende Münzwurfreihe hat 19 Übergänge. Unter der Unabhängigkeitsannahme errechnet sich für die Zahl der Übergänge von Gleichem zu Gleichem ein Erwartungswert von $19 \cdot 0{,}5 = 9{,}5$. Abbildung 1 zeigt Boxplots der Übergangshäufigkeiten von Gleichem zu Gleichem, getrennt für die jeweiligen Klassenstufen. Es gibt kaum Unterschiede über die Klassenstufen hinweg. Zur gleichen Schlussfolgerung kommt man beim Vergleich der Mittelwerte der Übergangswahrscheinlichkeiten: während die Münzwurfreihen der Siebtklässler im Durchschnitt 8,01 Gleiches auf Gleiches folgen ließen, lagen die entsprechenden Werte bei den Fünftklässlern bei 7,58 und bei den Neuntklässlern bei 7,48. Wie viel Prozent der Schüler liegen hier unter dem theoretischen Erwartungswert? In Klasse 5 sind es 80%, in Klasse 7 sind es 83% und in Klasse 9 sogar 84% aller Schüler, die zu selten Gleiches auf Gleiches folgen lassen. Trotz der (geringfügigen) Verschlechterung über die Jahre hinweg, halten die beobachteten Unterschiede einer inferenzstatistischen Untersuchung nicht stand.

Abb. 1: Boxplots der Häufigkeiten, mit denen beim 20-fachen Münzwurf Gleiches auf Gleiches folgt.



Anzahl der Runs

Während alle einzelnen Realisierungen einer Münzwurfreihe mit gleicher Wahrscheinlichkeit hier mit

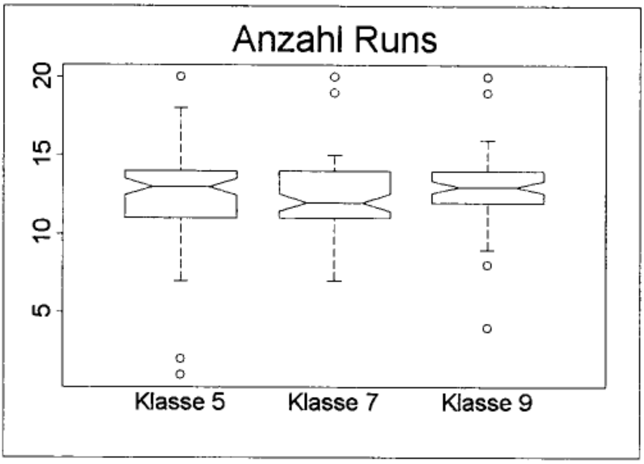
$$\boldsymbol{P} = \left(\frac{1}{2}\right)^{20}$$

auftreten, kann zur Beurteilung der Zufälligkeit konkreter Reihen das Auftreten bestimmter Muster herangezogen werden. Man bezeichnet die Hintereinanderfolge von gleichen Ergebnissen als einen Run (oder Lauf). Z.B. besteht die Abfolge WWZWWZZZ aus 4 Runs (zuerst ein W Run bestehend aus 2 W's gefolgt von einem Z Run, dann wiederum ein W Run bestehend aus 2 W's und zuletzt ein Z Run aus 3 Z's). Bei zufällig entstandenen Reihen ist das Vorkommen von übermäßig vielen, wie auch von extrem wenig Runs sehr unwahrscheinlich. Wahrscheinlichkeitstheoretische Betrachtungen lassen beim zwanzigfachen Münzwurf unter der Unabhängigkeitsannahme im Mittel 10,5 Runs erwarten. Abbildung 2 zeigt Boxplots für die einzelnen Klassenstufen und Tabelle 3 die empirischen Mittelwerte der Anzahl der Runs.

Tab. 3: Mittelwerte der Anzahl Runs

	Klasse 5	Klasse 7	Klasse 9	Gesamt
Hauptschule	13,44	12,13	12,96	12,86
Realschule	13,75	12,50	12,28	12,83
Gymnasium	10,78	12,14	12,42	11,72
Gesamt	12,59	12,26	12,55	12,47

Abb. 2: Boxplots für die Anzahl der Runs



Auch hier zeigt sich im wesentlichen dasselbe Bild wie im vorangegangenen Abschnitt. Egal ob man die Analyse eher auf den Mittelwerten oder den gegenüber Ausreißern weniger empfindlichen Medianen und Quantilen (siehe Boxplots in Abbildung 2) basiert, so gilt: Während bei den Schultypen Gymnasiasten am besten abschneiden, sind signifikante Unterschiede in den Klassenstufen nicht zu identifizieren. Die Münzwurfreihen weisen zu viele Runs auf, d.h. mehr als die theoretische zu erwartende Zahl. Bei den Fünftklässlern geben 80,26% der Schüler zu viele Runs an, in Klasse sind es 81,82% und in Klasse 9 sogar 84,81 % der Schüler.

Länge des längsten Runs

Ein weiteres verbreitetes Kriterium zur Beurteilung der Zufälligkeit einer Münzwurfserie ist die Länge des längsten Runs. Aufgrund eines mangelnden Verständnisses von Unabhängigkeit wird die Länge des längsten Runs von vielen Erwachsenen als zu kurz angegeben. Theoretische Überlegungen (siehe etwa Eichelsbacher, 2002) weisen darauf hin, dass die zu erwartende Länge des längsten Runs proportional zum logarithmus dualis der Serienlänge ist:

$$\boldsymbol{L}_n \approx \frac{\log(\boldsymbol{n})}{\log(2)} = \mathrm{ld}(\boldsymbol{n}) .$$

Im Falle von $n = 20$ errechnet sich somit

$$\boldsymbol{L}_n \approx 4{,}322 ,$$

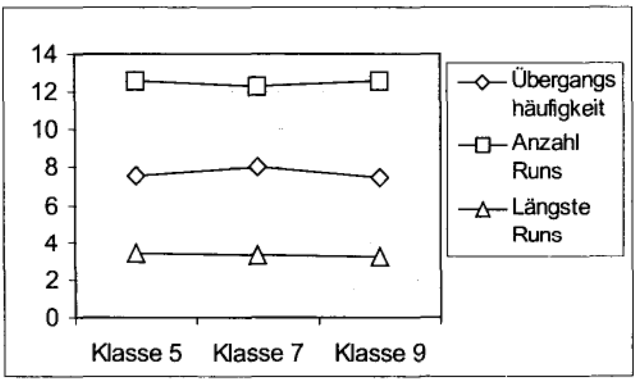
d.h. ein Run der Länge 4 oder 5 liegt voll und ganz im Bereich dessen, was in einer Reihe der Länge 20 zu erwarten ist. Tabelle 4 stellt die empirischen Mittelwerte dar. Ein Boxplot ist hier wenig sinnvoll, da *alle* Mediane und *zugleich alle* unteren Viertelswerte exakt 3 betragen. Auch hier zeigt sich, dass die Gymnasiasten dem theoretischen Wert am nächsten kommen, während Realschüler sogar geringfügig schlechter als die Hauptschüler abschneiden. Hingegen sind Unterschiede auf dem Klassenstufenniveau wiederum sehr gering.

Tab. 4: Mittelwert der Länge des längsten Runs

	Klasse 5	Klasse 7	Klasse 9	Gesamt
Hauptschule	2,80	3,39	3,24	3,14
Realschule	2,88	3,15	3,16	3,06
Gymnasium	4,41	3,36	3,21	3,70
Gesamt	3,39	3,29	3,20	3,29

Zusammenfassend lässt sich somit feststellen: Bei der Simulation von Münzwürfen sind die Schüler über alle Klassenstufen hinweg sehr genau, wenn es darum geht, gleiche Wahrscheinlichkeiten für Wappen und Zahl zu produzieren. Schüler haben jedoch deutliche Schwierigkeiten damit, die Unabhängigkeit der Ereignisse angemessen zu repräsentieren. Dies gilt über alle Klassenstufen hinweg. Die längsten Runs sind nicht lange genug, während die Anzahl der Runs zu hoch ist, verglichen mit den theoretisch zu erwartenden Ergebnissen. Die Schwierigkeit, den Begriff der Unabhängigkeit aufeinanderfolgender Ereignisse zu erfassen, spiegelt sich auch bei den Übergangshäufigkeiten wieder, mit denen Gleiches auf Gleiches folgt. Die relativen Übergangshäufigkeiten sind konsistent kleiner als 0,5. Die genannten Schwierigkeiten treten bei Schülern des Gymnasiums sichtbar weniger deutlich auf als bei Realschülern oder Hauptschülern. Von einer Verbesserung mit zunehmenden Jahren des Schulbesuchs kann auf gar keinen Fall die Rede sein (Abbildung 3).

Abb. 3: Darstellung der Ergebnisse zu Aufgabe 2 im Überblick des zeitlichen Ablaufs der Klassenstufen. Die jeweiligen theoretischen Richtzahlen sind: Übergangshäufigkeit=9,5; Anzahl Runs=10,5; längster Run= 4, 322.



4.2 Vorstellungen über die Gleichverteilung in der Ebene

Piaget & Inhelder (1975) untersuchten als erste das Verständnis für das Muster in Zufallsfolgen bei Kindern. Sie entwarfen ein Szenario, bei dem Regentropfen auf die Kacheln eines Pflasters fielen. Das Verlangen nach Regularität dominierte die Vorhersagen der Kinder. Danach gefragt, wo die nächsten Regentropfen hinfallen würden, verteilten 6 bis 9 jährige Kinder die Tropfen in ungefähr gleicher Zahl auf jedes Pflasterquadrat. Wenn auf allen bis auf einem Quadrat schon ein Tropfen gefallen war, platzierten die Kinder zwangsläufig den nächsten Tropfen auf das einzige noch leere Quadrat. Mit zunehmendem Alter verschwand dieses Proportionalitätsdenken und Irregularitäten in der Verteilung wurden eher akzeptiert. Piaget & Inhelder zogen daraus die Schlussfolgerung, dass in dieser Phase Kinder begannen, das Gesetz der großen Zahl zu verstehen, das gleichzeitig globale Regularität und lokale Variabilität des Experiments erklärt. Diese Theorie wurde jedoch von Green (1986) diskutiert, dessen Untersuchungen bei 11 bis 16 jährigen Schülern darauf hinweisen, das der Prozentsatz von Kindern mit zunehmenden Alter abnahm, die zufällige und halbzufällige Verteilungen erkannten. Green (1982) stellte 2930 englischen Schülern im Alter zwischen 11 und 16 die Regentropfenaufgabe im multiple-choice Format (vier mögliche Muster vorgegeben), ebenso eine Variante der Münzwurfserie im Multiple-choice Format und entdeckte keinerlei Verbesserung mit zunehmenden Alter. In Green (1986) berichtet der Autor über die Schneeflockenaufgabe mit 1600 Kindern im Alter von 7 bis 11 (eingeteilt in vier Altersgruppen) und kommt zu dem Ergebnis, dass beim Vergleich dieser Altersgruppen keinerlei Verbesserungen über die Jahre stattfindet mit Ausnahme der jüngsten Altersgruppe der 7-8 jährigen, die deutlich schlechter abschneiden. Borovcnik und Bentz (1991) diskutieren diese Aufgabe und kritisieren die Eignung der vorgegebenen Antwort-Kategorien zur Feststellung des Erfassens des Wirkens des Zufalls.

Wir haben in Frage 2 diesen Aufgabentyp aufgegriffen, aber anders als bei Green durften die Schüler ohne irgendwelche Vorgaben die Verteilung der Schneeflocken auf dem Garagendach frei wählen. Zusätzlich wurden sie aufgefordert, ihre Lösung in Worten zu beschreiben. Dadurch trifft die Kritik von Borovcnik und Bentz auf unsere Aufgabenformulierung nicht mehr zu. Bei der Auswertung der Antworten identifizierten wir vier verschiedene Muster:

1. „Strenger Determinist“: Feste, starre Muster dafür, wie die Flocken fallen, sind klar erkennbar,: Z.B. fällt jede Flocke genau in die Mitte jeder Kachel und keine Kachel bleibt frei.
2. „Gemäßigter Determinist“: Muster sind zweifellos feststellbar, wenngleich die Verteilung der Flocken nicht bis ins letzte Detail vorbestimmt ist. Z.B. erhält jede Kachel genau eine Schneeflocke, die aber irgendwo ohne erkennbares Muster auf der Kachel liegt, oder es schneit nur auf

Kacheln, die in einer bestimmten Anordnung liegen (z.B. nur auf die inneren oder nur auf die äußeren Kacheln).

3. „Novizen“: Es ist offensichtlich, dass die Schneeflocken irgendwo auf dem Dach zu liegen kommen. Muster sind nicht erkennbar, insbesondere gibt es mindestens eine leere Kachel. Jedoch ist die Variabilität der Belegungszahlen (Anzahl der Flocken pro Kachel) recht klein. Es gibt zwar Kacheln, die mehr als eine Flocke erhalten und andere Kacheln, die frei bleiben. Maximal drei Kacheln sind leer.
4. „Experte“: Offensichtlich zufällige Verteilung. Muster sind nicht erkennbar, zwischen vier bis acht Kacheln bleiben frei. Es sind keinerlei Symmetrien und keinerlei Muster und Regelmäßigkeiten erkennbar. Es ist offensichtlich, dass die Schneeflocken irgendwo auf dem Dach zu liegen kommen.

Die Klassifizierung der Schülerantworten in eine dieser vier Kategorien erfolgte per Urteil von drei unabhängig arbeitenden Experten. In Zweifelsfällen wurde die schriftliche Begründung zur Klassifizierung mit herangezogen. Insgesamt gelang nur bei einem Schüler, der insgesamt nur 5 Flocken einzeichnete, eine Klassifizierung nicht. Die quantitative Auswertung erfolgt mit Hilfe dreier unterschiedlicher Kriterien (Score 1, Score 2 und Score 3), die wie folgt definiert wurden:

Score 1 = relative Zahl derjenigen, die keine strengen Deterministen sind
Score 2 = relative Zahl der Experten oder Novizen
Score 3 = relative Zahl der Experten

Aufschlussreich bei der offensichtlichen Schwierigkeit, Zufälligkeit in der Verteilung der Schneeflocken angemessen zu respektieren, sind hierbei auch die im Fragebogen erbetenen Begründungen durch die Schüler. Hier ein paar Beispiele für die Position der Deterministen:

- *Weil es 16 Platten sind und 16 Schneeflocken fallen, deshalb fällt auf jede Platte [genau eine Flocke]*(Hauptschule, Klasse 5)
- *Weil auf jede Platte eine Schneeflocke hingefallen ist* (Hauptschule, Klasse 7)
- *Die Schneeflocken fallen gleichmäßig* (Realschule, Klasse 7)
- *Weil überall gleichviel Schneeflocken hinfallen* (Gymnasium, Klasse 7)
- *Theoretisch müssen sich die 16 Schneeflocken gleichmäßig auf die 16 Platten verteilen* (Realschule, Klasse 9)
- *Weil es gleichmäßig schneit* (Gymnasium, Klasse 9)

Zur Rolle des Unterrichts bei der Dominanz deterministischen Denkens passt auch folgende Episode: In einer siebten Klasse identifizierte der Lehrer für den Untersuchungsleiter seine beiden „Musterschülerinnen“. Diese fielen dadurch auf, dass sie ihre Kreuze besonders akkurat in die Mitte der 16 jeweiligen Felder setzten.

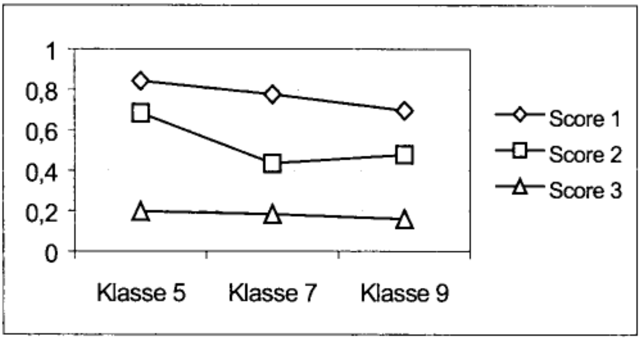
Bei Schülerantworten, die ein angemessenes Zufallsverständnis ausdrückten, kamen sowohl die Unvorhersehbarkeit des Zufalls zum Ausdruck wie auch eine Vorstellung von Zufall als Einfluss einer unkontrollierten Drittvariable wie Wind oder Wärme.

- *Weil durch den Wind die Schneeflocken da hingefallen sind* (Realschule, Klasse 5)
- *Weil in der Naturwissenschaft nicht immer alles perfekt ist. Ich hab deswegen 3 Kästchen freigelassen* (Gymnasium, Klasse 5)
- *Weil die Schneeflocken einmal da und einmal da hinfliegen* (Hauptschule, Klasse 7)
- *Es kommt darauf an, wie der Wind weht.* (Hauptschule, Klasse 7)
- *Schneeflocken fallen ohne System* (Gymnasium, Klasse 7)
- *Der Schnee fällt ja nicht einfach so. Es ist ja nicht kontrollierbar oder vorhersehbar* (Hauptschule, Klasse 9)
- *Man kann nicht wissen, wo die Schneeflocken landen. Und die Chance, dass alle 16 Schneeflocken auf alle 16 Platten landen ist sehr gering* (Hauptschule, Klasse 9)
- *Es gibt keine Begründung, sie fallen einfach zufällig. Hab die Augen zugemacht und per Zufall mit dem Stift diese Punkte getroffen* (Gymnasium, Klasse 9)

Eine Betrachtung der Schülerzitate gibt einen interessanten Einblick in die Vorstellungswelt der Schüler, wie sie ihr Denken über Zufall bzw. Bestimmtheit begründen. Die Zitate sind aber wenig geeignet, um Trends über die zeitliche Entwicklung zu identifizieren. Dazu dient eine Analyse der drei Scores. Sie lassen keine systematischen Unterschiede zwischen den Schultypen erkennen, während jedoch eine Verschlechterung über die Schuljahre hinweg offenkundig wird (siehe Abbildung 4). Zu genau der gleichen Schlussfolgerung ist auch schon Green (1982) gekommen, der die drei unterschiedliche Muster des Schneefalls vorgibt, die im wesentlichen unseren Kategorien von „strenger Determinist“, „gemäßigter Determinist“ und „Experte“ entspricht, und dazu mehrere Fragen im Multiple-Choice Format stellt. Zu dem Ergebnis, dass sich die Zahl derer, die sich für die Expertenlösung entscheiden, von 26 % bei den 11-Jährigen auf 18 % bei den 16-Jährigen reduziert, bemerkt Green (1982):

> „The astonishing thing about this item is that performance declines with age. ... It can be hypothesized that we see at work here two oppsoing tendencies - maturation/ experience on the one hand and dominance of mathematical/ scientific deductivism on the other which stifles the appreciation of randomness by seeking to codify and explain everything.“

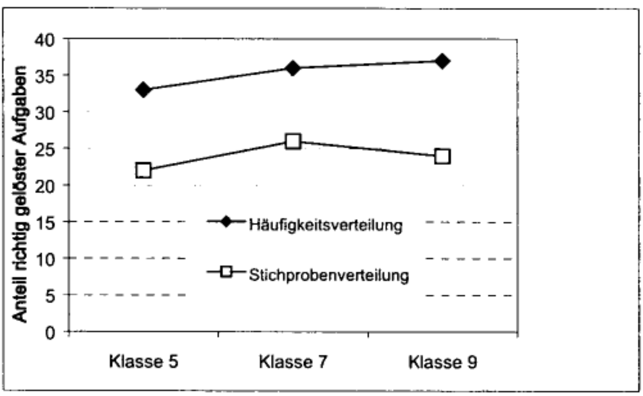
Abb. 4: Darstellung der Ergebnisse zu Aufgabe 2



4.3 Einfluss der Stichprobengröße

Abbildung 5 zeigt, wie sich die Sensitivität für die Auswirkung der Stichprobengröße bei einer Variation der „maternity ward" Aufgabe über die Schulklassen hinweg verändert. Für die aggregierten Daten zeigt sich der auch in früheren Studien gefundene Vorteil von Häufigkeitsverteilungs-Aufgaben. Allerdings ist die Lösungsrate für beide Arten von Aufgaben geringer als der Durchschnitt aus den Studien mit erwachsenen Versuchsteilnehmern (Sedlmeier & Gigerenzer, 1997). Die vergleichsweise niedrigen Lösungsraten könnten auf Probleme beim Verständnis der Aufgabe hindeuten (Sedlmeier, 1998).

Abb. 5: Darstellung der Ergebnisse zu Aufgabe 3



Wird die Sensitivität der Schüler für die Auswirkung der Stichprobengröße größer, je länger sie die Schule besuchen? Die Ergebnisse in Abbildung 5 deuten darauf hin, dass dies, wenn überhaupt, nur in unwesentlichem Ausmaß der Fall ist. Während bei der Häufigkeitsverteilungs-Aufgabe zwar ein leichter positiver Trend vorhanden ist, kehrt sich dieser anfänglich auch bei der Stichprobenverteilungs-Aufgabe vorhandene Trend wieder um. Man sollte auch beachten, dass die Lösungsrate selbst bei der Häufigkeitsverteilungs-Aufgabe fast auf Zufallsniveau liegt (drei Lösungsalternativen standen zur Auswahl). Insgesamt gesehen scheint sich also die Sensitivität für die Auswirkung der Größe einer Stichprobe auf die Genauigkeit der Schätzung eines Populationsparameters über die Schulzeit hinweg nicht zu verbessern.

4.4 Wahrscheinlichkeit von Konjunktionen

In Studien mit Erwachsenen liegen die Lösungsraten bei Konjunktionsaufgaben von der Art wie sie auch in der gegenwärtigen Studie verwendet wurden oft nicht höher als etwa 20 %. Ein ähnlich niedriges Niveau zeigt Abbildung 6 für die Schüler der Klasse 9. Darüber hinaus ist ein deutlicher Abwärtstrend zu sehen. Mit fortschreitender Klassenzahl sinkt der Anteil der richtig gelösten Aufgaben. Die zeitliche Perspektive, das heißt, ob die zu beurteilenden Ereignisse vor oder nach dem Eingangsereignis (eine Vier im Abschlusszeugnis vs. eine Sechs im Zwischenzeugnis) auftraten, hatte keinen Einfluss auf das Ergebnis. Erstaunlicherweise hatte auch das Darbietungsformat, Häufigkeiten vs. Wahrscheinlichkeiten, keinen systematischen Einfluss auf die Lösungsraten: während die Häufigkeitsversion tendenziell in der Klassen 7 etwas häufiger richtig gelöst wurde als die Wahrscheinlichkeitsversion, war es in der Klasse 5 umgekehrt. Die Ergebnisse für Klasse 9 unterschieden sich nur um einen Prozentpunkt zugunsten der Häufigkeitsversion.

Abb. 6: Darstellung der Ergebnisse zu Aufgabe 4.

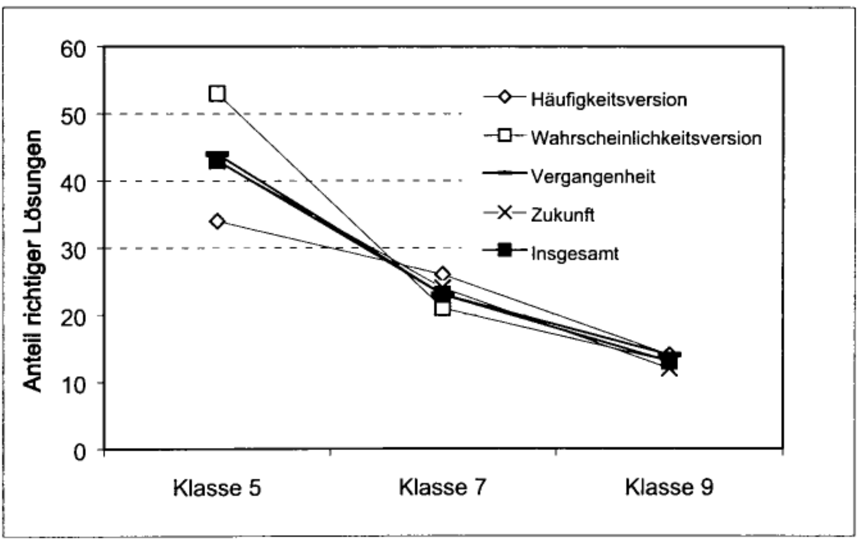


Abbildung 6 zeigt einen eindeutigen Trend: Die Schüler waren mit fortschreitender Klassenzahl immer weniger in der Lage eine (relativ einfache) Aufgabe zur Wahrscheinlichkeit von konjunktiven Aufgaben richtig zu lösen. Der fehlende Effekt in der zeitlichen Perspektive könnte entweder darauf zurückzuführen sein, dass dieser Faktor keine Rolle bei der Lösung dieser Art von Aufgaben spielt oder darauf, dass eine mögliche Vereinfachung der Aufgabenstellung nicht gelungen ist. Dies wiederum könnte daran liegen, dass zwar die Perspektive ausgehend von der einleitenden Problembeschreibung variiert wurde, aber die zu beurteilenden Ereignisse in allen Version schon eingetreten waren (alles spielt sich in der Vergangenheit ab). Am erstaunlichsten ist der fehlende Effekt des Darbietungsformats, wozu wir im Moment keine Erklärung anbieten können. Insgesamt ist das Ergebnis jedoch sehr deutlich konsistent mit der Annahme, dass statistische Intuitionen sich über die Schulzeit hinweg eher verschlechtern.

5. Diskussion und Implikationen für die Schule

Unsere Untersuchungen belegen, dass sich das Verständnis von Zufall und zufallsbedingter Variabilität in empirischen Daten im Verlaufe der Schuljahre bei 10 bis 15 jährigen Schülern nicht bessert. Im Gegenteil: wenn Trends überhaupt vorhanden sind, dann verschlechtert sich die Fähigkeit, Zufall und Variabilität angemessen einschätzen zu können. Dieses Resultat ist konsistent mit früheren Ergebnissen bei vergleichbaren Altersgruppen von Green (1982) und generalisiert den früheren Befund durch eine erweiterte Auswahl von Aufgaben und eine repräsentative Auswahl von Schülern. Es ist unseres Wissens auch die erste derartige Studie mit deutschen Schülern. Welche Konsequenzen sind daraus für die Currculumsentwicklung und Schulplanung zu ziehen?

Erfahrungen mit Zufall betreffen ebenso wie Erfahrungen mit kausalem Denken interdisziplinär viele Unterrichtsfächer. Zufall sollte Gegenstand von fächerverbindenden und fächerübergreifenden Themen sein, bei denen probabilistische mit deterministischen Erklärungen konfrontiert werden. Exemplarisch seien genannt: Verbindungen der Vererbungslehre in der Biologie mit Statistik, das Wetter, die Börse, Sportergebnisse oder Wahlprognosen. Experimente mit Zufallsstichproben können aber auch in sozialwissenschaftlichen Fächern durchgeführt werden. In besonderer Weise ist die Stochastik als „Mathematik des Zufalls“ angesprochen. Es kann allerdings im Unterricht nicht darum gehen, die Begriffs- und Methodenapparate der Stochastik rein formal als mathematische Theorien aufzubauen, in der Hoffnung, dass Fähigkeiten im Umgehen mit diesen formalen Apparaten die Fähigkeiten für eine konzeptionelles Verstehen von Zufall und zufallsbedingter Variabilität in empirischen Daten schon begründen. Ein Stochastikunterricht, der den Kalkül der Mathematik des Zufalls stark betont ohne Intuitionen statistischen Denkens viel Beachtung zu schenken, wird da nicht helfen (siehe Rasfeld, 2002). Kritisiert wird an einer sich zu sehr am axiomatischen Ansatz orientierten Didaktik, dass sie sich viel zu schnell auf ein präzises Kalkül und eine klare Sprache fixiert. Intuitive Vorstellungen als die wirklichen Quellen kreativen Denkens (Borovcnik, 1992) werden dabei zu schnell ausgeblendet. Primäre Intuitionen, d.h. Vorstellungen, die sich ohne systematische Behandlung eines Begriffs oder Konzepts entwickelt haben, sind manchmal in der Schule direkt verwendbar, manchmal müssen sie aber auch systematisch hinterfragt werden. Viele Fehlvorstellungen setzen sich fest - wie auch die Fülle an Paradoxa in der Stochastik zeigt - wenn nicht frühzeitig Vorstellungen (Sekundärintuitionen) aufgebaut werden, die sich erst aufgrund einer systematischen Behandlung und der Verbindung von Intuitionen mit Konzepten der Theorie herausbilden (siehe Tietze et al., 2002). Dabei scheinen drei Faktoren eine entscheidende Rolle zu spielen: das Darbietungsformat, das heißt, die Art und Weise, in der Information dargeboten wird, „learning by doing", und die Alltagsnähe der verwendeten Aufgaben (Sedlmeier, 1999). Eine überwiegend strukturmathematisch und weniger anwendungsorientierte Behandlung der Stochastik wäre also auf alle Fälle kontraproduktiv.

Der Erwerb von validen statistischen (Sekundär-) Intuitionen ist nur als gradueller Prozess zu verwirklichen, z.B. indem Lernende durch Experimente in Spiel- und Simulationssituationen schrittweise zu einer kompletten Formalisierung hingeführt werden (siehe z.B. Sedlmeier, 2001; Sedlmeier & Köhlers, 2000). Daher ist ein Unterricht zu fordern, der durch einen hohen Anteil experimenteller Arbeiten und durch selbständige Datenerhebungen charakterisiert ist und sich dabei neue Technologien (Computer, grafikfähige Taschenrechner, Internet) zu Nutze macht, um Simulationen und Demonstrationen zu planen und durchzuführen. Vom Zufallsgenerator erzeugte Daten können maßgeblich dazu beitragen, bei Schülern eine Intuition für zufallsbedingte Variabilität in empirischen Daten zu entwickeln. Indem Schüler Experimente mit Zufallsgeneratoren bzw. Zufallszahlen planen, durchführen und auswerten, sammeln sie Erfahrungen in stochastischen Situationen und können ihre Intuitionen über zufällige Erscheinungen und Wahrscheinlichkeiten überprüfen und gegebenenfalls korrigieren.

Der Aufbau angemessener Intuitionen muss an den Erfahrungen der Schüler ansetzen und praxis- und alltagsorientiert sein. Am Anfang des Stochastikunterrichts sollten daher eigene Aktivitäten mit Zufallsexperimenten stehen. Das schließt die Planung und Durchführung von Experimenten, Simulationen und Demonstrationen sowie die Konstruktion von Modellen und deren Interpretation ein. Anhand von Simulationen kann die Überprüfung von Intuitionen und die Bildung stochastischer Modelle sehr ertragreich geübt werden. Simulationen lassen sich mit Computerunterstützung besonders effektiv durchführen. Solche Experimente resultieren immer in direkt beobachtbaren Ergebnissen und nicht direkt in Wahrscheinlichkeiten, deren Verständnis häufig sehr schwierig ist. Diese Art und Weise der Informationsdarbietung (Häufigkeiten statt Wahrscheinlichkeiten) erleichtert letztendlich auch das Benutzen von formaleren Darstellungsarten (Wassner, Martignon & Sedlmeier, 2002).

Zufall bedeutet, dass es in einer gegebenen Situation mehr als nur ein mögliches Ergebnis gibt, dass das tatsächlich eintretende Ergebnis nicht vorhersagbar ist, dass - zumindest in der Vorstellung - die Möglichkeit besteht, das Experiment mehrfach zu wiederholen, und dass die Folge der durch Wiederholung erhaltenen Resultate keine Muster besitzt, die sich kontrollieren oder vorhersagen lassen. In dieser offensichtlichen Unordnung können eine Vielzahl von globalen Regelmäßigkeiten entdeckt werden, deren offensichtlichste die Stabilisierung der relativen Häufigkeit eines möglichen Ergebnisses ist. Diese globale Regelmäßigkeit ist die Grundlage, die uns erlaubt, Zufallsphänomene systematisch mit Hilfe der Wahrscheinlichkeitsrechnung zu studieren. Ein grundlegendes Verständnis von Zufall und Wahrscheinlichkeit ist notwendig für das Verstehen der Ungewissheiten des Lebens und erleichtert den Umgang damit. Es wäre schade, wenn unseren Schülern dieses grundlegende Verständnis und das darauf beruhende Hilfsmittel der Wahrscheinlichkeitsrechnung durch eine inadäquate Schulausbildung, nicht nur im Rahmen des Mathematikunterrichts, vorenthalten würde.

Literatur

Batanero, C., Godino, J., Vallecillos, A., Green, D. & Holmes, P.(1994): Errors and difficulties in understanding elementary statistical concepts. *Int. J. Math. Educ. Sci. Technol.*, 25 (4), 527 -47.

Borovcnik, M. (1992): *Stochastik im Wechselspiel von Intuitionen und Mathematik*. Bibliographisches Institut, Mannheim.

Borovcnik, M. & Bentz, H. J. (1991): Empirical Research in Understanding Probability. In: Chance Encounters: Probability in Education, eds. R. Kapadia and M. Borovcnik.

Eichelsbacher, P. (2002): Mit Runs den Zufall besser verstehen. *Stochastik in der Schule* (22), Heft 1, 2-7

Fischbein, E. (1975): *The Intuitive Sources of Probabilistic Thinking in Children.*. Reidel: Dordrecht-Holland.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review*, *103*, 592–596.

Green, D. R. (1982): A Survey of Probability Concepts in 3000 Students aged 11-16 Years. In: D. R. Grey (Ed.): *Proceedings of the First International Conference on Teaching Statistics*, Teaching Statistics Trust, University of Sheffield, 766-783.

Green, D. R. (1986): Children's understanding of randomness. In: R. Davidson & J. Swift (eds.) *Proceedings of the Second International Conference on Teaching Statistics*. Victoria, British Columbia.

Green, D. R. (1990): *A Longitudinal Study of Pupil's Probability Concepts*. Loughborough: Loughborough University

Hertwig, R. & Gigerenzer, G. (1999). The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning error. *Journal of Behavioral Decision Making*, *12*, 275-306.

Inhelder, B. & Piaget, J. (1964). *The early growth of logic in the child.* (E. A. Lunzer & D. Papert, Trans.). London: Routledge & Kegan Paul. (Original work published 1959)

Kahneman, D. & Tversky, A. (1972): Subjective probability: a judgement of representativeness. *Cognitive Psychology*, 3, 430-454

Kahneman, D., Slovic, P. & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Kuzmak, S. D. & Gelman, R. (1986). Young children's understanding of random phenomena. *Child Development*, *57*, 559–566.

Mellers, B., Hertwig, R. & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An Exercise in adversarial collaboration. *Psychological Science*, *12*, 269-275.
Nickerson, R. (2002): The Production and Perception of Randomness. *Psychological Review*, 109 (2), 330-357
Piaget, J. & Inhelder, B. (1975). *The origin of the idea of chance in children* (L. Leake, Jr., P. Burrel & H. D. Fishbein, Trans.). New York: Norton. (Original work published 1951)
Piattelli-Palmarini, M. (1994). *Inevitable illusions: How mistakes of reason rule our minds*. New York: Wiley.
Rasfeld, P. (2002): Verbessert der Stochastikunterricht intuitives stochastisches Denken? Erste Ergebnisse einer empirischen Studie. In: W. Peschek (Hrsg.) *Beiträge zum Mathematikunterricht*, Franzbecker: Hildesheim
Piaget, J. & Inhelder, B. (1975): T*he Origin of the Idea of Chance in Children.* Routledge
Sedlmeier, P.(1998): The distribution matters: two types of sample-size tasks. *Journal of Behavioral Decision Making*, 11, 281-301
Sedlmeier, P. (1999) *Improving statistical reasoning: Theoretical models and practical implications*. Mahwah: Lawrence Erlbaum Associates.
Sedlmeier, P. (2001). Statistik ohne Formeln. In M. Borovcnik, J. Engel & D. Wickmann (Hrsg). Anregungen zum Stochastikunterricht. Hildesheim: Franzbecker. (S. 83-95)
Sedlmeier, P. (2002) Associative learning and frequency judgments: The PASS model. In P. Sedlmeier & T. Betsch (Eds.). *Etc: Frequency processing and cognition*. Oxford: Oxford University Press (pp. 137-152)..
Sedlmeier, P. & Gigerenzer, G. (1997): Intuitions about sample size: the empirical law of large numbers. *Journal of Behavioral Decision Making*, 10, 33- 51
Sedlmeier, P. & Köhlers, D. (2001): *Wahrscheinlichkeiten im Alltag: Statistik ohne Formeln*. Braunschweig: Westermann.
Shaugnessy, J. M. (1992): Research in probability and statistics: reflections and directions. In D. A. Grouws (ed.) *Handbook of Research on Mathematics Teaching and Learning*. Edited by D. Grouws, New York: Macmillan, 465-494.
Steinbring, H. (1986): L'independence stochastique. *Recherches en Didactique des Mathematiques*, 7 (3), 99-118.
Tietze, U., Klika, M. & Wolpers, H. (2002): *Mathematikunterricht in der Sekundarstufe II. Band 3: Didaktik der Stochastik*. Wiesbaden: Vieweg.
Tversky, A. & Kahneman, D. (1983): Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review*, 90, 293-315.
Wassner, C., Martignon, L. & Sedlmeier, P. (2002). Entscheidungsfindung unter Unsicherheit als fächerübergreifende Kompetenz. *Zeitschrift für Pädagogik*, 45. (Beiheft), 35-50.

Anschrift der Autoren:

PD Dr. Joachim Engel
Pädagogische Hochschule Ludwigsburg
Postfach 220
71602 Ludwigsburg
E-Mail: engel_joachim@ph-ludwigsburg.de

Prof. Dr. Peter Sedlmeier
TU Chemnitz, Institut für Psychologie
Straße der Nationen 62
0911 Chemnitz
E-Mail: peter.sedlmeier@phil.tu-chemnitz.de