

Wahrscheinlichkeitsvergleich und inferenzstatistisches Schließen

Fähigkeiten von Kindern des 4. und 6. Schuljahrs bei Basisproblemen aus dem Bereich Daten und Zufall

von

Anke Lindmeier, Kiel & Kristina Reiss, München

Kurzfassung: Mathematikunterricht soll zu einem kumulativen Aufbau der Fähigkeiten im Bereich der Leitidee *Daten, Häufigkeit und Wahrscheinlichkeit* bereits von der Grundschule an beitragen. Allerdings sind Befunde zu Fähigkeiten von Schülerinnen und Schülern in der Primarstufe immer noch rar. In diesem Artikel werden die Fähigkeiten bei den zwei Basisproblemen des Wahrscheinlichkeitsvergleichs und inferenzstatistischen Schließens in den Fokus genommen. Die Ergebnisse zeigen, dass für Kinder in den Jahrgangsstufen 4 und 6 mit Hilfe von speziellen, altersadäquaten Problemstellungen durchaus Kompetenzen sichtbar werden, aber auch typische Strategien und (Fehl-)Vorstellungen zum Tragen kommen. Die verwendeten Aufgabenformate werden in dem Artikel detailliert beschrieben und unterrichtliche Implikationen der Befunde werden diskutiert.

Abstract: Mathematical instruction should foster the cumulative acquisition of abilities in the field of *data and chance* from the early years in elementary school on. However, the knowledge about abilities of students in elementary grades is still deficient. This article focusses on the abilities to handle two basic problems dealing with the comparison of probabilities and inferential reasoning. The results show that children in grade 4 and 6 are already able to solve these problems if they are posed as special, age-adequate problems. However, typical strategies and (mis-)conceptions can also be identified. This article includes a detailed description of the problems used and discusses instructional implications of the study.

1 Theoretischer Hintergrund

Stochastik¹ hat in der Vergangenheit zwar eine eher kleine Rolle im Curriculum der Primarstufe gespielt, ist aber mit der Einführung der Bildungsstandards (KMK, 2004a) als Leitidee *Daten, Häufigkeit und Wahrscheinlichkeit* prominent in den Mathematikunterricht gekommen. Dabei geht es im Unterricht bis zur Jahrgangs-

¹ In diesem Artikel wird der deutschsprachigen Tradition entsprechend Stochastik als Gebiet verstanden, das sowohl Statistik als auch Wahrscheinlichkeitsrechnung umfasst (vgl. Burrill & Biehler, 2012).

stufe 4 vor allem um Datengewinnung und Datendarstellung, um Grundbegriffe der Wahrscheinlichkeitsrechnung und um einfache Zufallsexperimente. Die Leitidee „Daten und Zufall“ der Bildungsstandards für den Mittleren Schulabschluss bzw. den Hauptschulabschluss führt diesen Themenstrang fort (KMK, 2003; 2004b). Die Bildungsstandards für das Abitur (KMK, 2012) beschreiben, wie der kumulative Kompetenzaufbau durch die Beherrschung mehrstufiger Zufallsexperimente und durch Einblicke in die beurteilende Statistik fortgesetzt wird und zum gymnasialen Allgemeinbildungsauftrag beitragen soll. Dabei sehen die Bildungsstandards einen zweigleisigen Zugang zur Leitidee vor, der bereits früh Methoden der Statistik mit dem klassischen Wahrscheinlichkeitsbegriff vernetzt.

Dieser Artikel fokussiert auf den Elementarbereich. Die hier genannten stochastischen Inhalte und Basisprobleme sind wichtig für den systematischen Aufbau des Fachs Mathematik, sie können aber auch allgemeiner als Grundlage für den Umgang mit wissenschaftlicher Evidenz gesehen werden (vgl. Bullock, Sodian & Koerber, 2009; Kuhn & Pearsall, 2000). Im Sinne des *statistical literacy* Konzepts (Gal, 2004) sind die Fähigkeiten zur Bewältigung der adressierten Basisprobleme notwendige Voraussetzung für den Aufbau von Kompetenzen zum reflektiert-kritischen Umgang mit unsicheren Situationen im Alltag. Wir betrachten an dieser Stelle – neben Fähigkeiten zur Unterscheidung von sicheren, möglichen (aber nicht sicheren) und unmöglichen Ereignissen – Fähigkeiten zum Vergleich von Wahrscheinlichkeiten aus einfachen Zufallsexperimenten und auch das elementare inferenzstatistische Schließen vom Merkmal einer Stichprobe auf das Merkmal der Grundgesamtheit. Dies wird im Folgenden begründet, dabei wird die aktuelle Forschungslage für die genannten Basisprobleme aufgearbeitet.

1.1 Basisproblem Wahrscheinlichkeitseinschätzung

Die korrekte Beurteilung von sicheren, möglichen (aber nicht sicheren) und unmöglichen Ereignissen kann man als grundlegende Voraussetzung für den Umgang mit Zufall und Wahrscheinlichkeit sehen. Empirische Forschung mit psychologischem, aber auch mathematikdidaktischem Hintergrund zeigt auf, welche Schwierigkeiten junge Kinder damit haben können. So baten Shtulman und Carey (2007) Kinder um die Beurteilung alltäglicher und bekannter möglicher Ereignisse (z. B. einen Apfel essen) sowie unmöglicher (z. B. durch eine Mauer aus Ziegelsteinen laufen) und unwahrscheinlicher Ereignisse (z. B. einen Alligator unter dem Bett finden). Kinder im Vor- und Grundschulalter hatten Probleme, zwischen unmöglichen und unwahrscheinlichen Ereignissen zu unterscheiden, wobei dies nicht auf sprachliche Probleme zurückgeführt werden konnte. Auch in stärker mathematisch-orientierten Kontexten fiel diese Unterscheidung Grundschulkindern noch schwer (Green, 1982; Fischbein, Nello & Marino, 1991; Lindmeier et al., 2011). In unseren Arbeiten zeigte sich allerdings, dass bereits Kinder der 2. Jahrgangsstufe zwischen sicheren und möglichen (aber nicht sicheren) Ereignissen gut differenzie-

ren konnten und sich die Fähigkeiten im Basisproblem Wahrscheinlichkeits-einschätzung bei einfachen Zufallsexperimenten bis Jahrgangsstufe 6 weiter ausbildeten (Lindmeier et al., 2011). Offensichtlich wird also bereits im Verlauf der Grundschulzeit ein Verständnis für die Einschätzung von Wahrscheinlichkeiten in einfachen Situationen erworben, so dass der Umgang mit Unsicherheit angebahnt wird.

1.1 Basisproblem Wahrscheinlichkeitsvergleich

Darauf aufbauend wird die Fähigkeit, Wahrscheinlichkeiten in einfachen Zufalls-situationen zu vergleichen, als grundlegend für die Ausbildung stochastischer Kompetenz verstanden. Die Forschung beschreibt relativ stabile Schwierigkeiten von Schülerinnen und Schülern in diesem Bereich. Dabei beziehen sich die meisten mathematikdidaktischen Untersuchungen auf eine bestimmte Form des Wahr-scheinlichkeitsvergleichs: Bekannt sind die Aufgaben zum Urnenvergleich (gege-ben sind Urnen verschiedener Zusammensetzung und feste Gewinnelemente; ge-fordert wird die Wahl der günstigeren Urne, z. B. Yost, Siegel & Andrews, 1962) oder zur Anpassung von Urnen (gegeben sind Urnen verschiedener Zusammenset-zung und feste Gewinnelemente; gefordert wird die Veränderung einer Urne, so dass es egal ist, aus welcher Urne gezogen wird, z. B. Falk & Wilkening, 1998). Notwendig zur Lösung dieser Aufgaben ist ein Vergleich der Grundpopulationen bei gleichbleibendem Gewinnereignis. Die meisten dieser Aufgaben sind nicht ohne Rückgriff auf proportionales Denken lösbar, weswegen diese Art von Wahr-scheinlichkeitsvergleich in der Grundschule nur bedingt geleistet werden kann (In-helder & Piaget, 1958). Dies zeigt sich unter anderem auch in den o. g. Studien (für einen Überblick siehe Langrall & Mooney, 2005).

Für wissenschaftliches Denken zentral ist allerdings eine weitere Art des Wahr-scheinlichkeitsvergleichs, bei dem bei stabiler Grundpopulation das Problem darin besteht, die Wahrscheinlichkeiten unterschiedlicher Ereignisse zu vergleichen. Dies tritt zum Beispiel auf, wenn Vermutungen über die Eintrittswahrscheinlich-keit unterschiedlicher Ereignisse gebildet werden müssen (Beispielproblem: „In ei-nem Fluss leben deutlich weniger Hechte als Karpfen. Angelt Max eher zwei Karp-fen oder einen Hecht und einen Karpfen?“). Bei dieser speziellen Art des Wahr-scheinlichkeitsvergleichs über derselben Grundpopulation können – je nach gege-benen Ereignissen – strukturelle Eigenheiten des Ergebnisraums genutzt werden und Vergleiche ohne Rückgriff auf die exakten Wahrscheinlichkeiten durchgeführt werden. Insbesondere eignen sich diese Vergleiche um bei Lernenden, die noch nicht in Proportionen denken können, ein Grundverständnis für die Struktur von Wahrscheinlichkeitsräumen zu überprüfen. Während im fortgeschrittenen Sto-chastikunterricht selbstverständlich Aufgaben vorkommen, die diesem Typ zuzu-ordnen sind (Beispielproblem: „Welches Ereignis ist beim zweimaligen Würfeln wahrscheinlicher? Ein 6er-Pasch oder Augensumme 3?“), wurde diese Art des

Wahrscheinlichkeitsvergleichs in der Forschung und bei jüngeren Kindern bisher vernachlässigt.

Bei Aufgaben zum Wahrscheinlichkeitsvergleich zeigen Kinder – und auch Erwachsene – stabile Schwierigkeiten. Diese können teilweise mit Hilfe der gut beschriebenen Heuristiken erklärt werden, die Menschen beim Problemlösen im Bereich *Daten und Zufall* verwenden (Überblicke bei Kahneman, 2011; Weixler, 2009). Solche Heuristiken können – je nach Problem – zielführend sein oder aber zu systematischen Fehlern² führen (Kahneman & Tversky, 1973). Beim dem hier betrachteten Basisproblem des Wahrscheinlichkeitsvergleichs gilt dies vor allem für drei Phänomene, die als *conjunction fallacy*, *misconception of representativeness* und *difficulty to compare compound and simple events* bekannt sind (Kahneman, Slovic & Tversky, 1982; Tversky & Kahneman, 1971; 1973; Kahneman & Tversky, 1973; Fischbein & Schnarch, 1997).

So werden Unterereignisse nicht als solche erkannt und z. B. das Ereignis „ein Kind hat einen rosa Pullover an“ als weniger wahrscheinlich interpretiert als „ein Kind hat einen rosa Pullover an und ist ein Mädchen“ (*conjunction fallacy*). Ganz eindeutig ist die Forschungslage allerdings nicht. So zeigen Untersuchungen von Fiedler (1988), dass es eine Rolle spielen kann, ob das (unwahrscheinlichere) Unterereignis besser vorstellbar ist (*availability heuristics*). Fiedler (1988) konnte zeigen, dass mit einer veränderten Aufgabenstellung, in der nach der Häufigkeit der Ereignisse gefragt wird, die beschriebene Fehlvorstellung bei Studierenden weniger stark zu Tage trat. Offen ist, inwiefern eine besonders günstige Darstellung der Ereignisse helfen kann, die Struktur des Ergebnisraums zur Lösung von Vergleichsproblemen zu nutzen.

Außerdem werden Ereignisse, die Aspekte der Grundpopulation repräsentieren, als wahrscheinlicher angesehen als weniger repräsentativ erscheinende Ereignisse (*representativeness heuristics*). So würde z. B. die Würfelsequenz „2, 4, 1, 3“ als wahrscheinlicher eingeschätzt werden als die Würfelsequenz „1, 2, 3, 4“. Fischbein und Schnarch (1997) konnten zeigen, dass Repräsentativitätsvorstellungen in Jahrgangsstufe 5 noch stark ausgeprägt sind. Bei Aufgaben zum Wahrscheinlichkeitsvergleich kann sich eine Repräsentativitätsheuristik je nach Vergleichspaar erschwerend oder erleichternd auswirken, wie unten im Detail ausgeführt wird.

Als weiteres Phänomen wird die Schwierigkeit im Umgang mit einfachen und zusammengesetzten Ereignissen beschrieben (*compound and simple events*). Auf Basis einer Laplace-Verteilung von Farben würden z. B. die beiden Ereignisse „rot-rot“ und „rot-blau“ aus Ziehungen mit Zurücklegen und ohne Berücksichtigung der

² Sind aus den Heuristiken resultierende Fehlleistungen (oder Verzerrungen, engl. *biases*) Gegenstand der Forschung, so werden die Heuristiken auch gerne als Fehlvorstellungen bezeichnet, siehe auch Weixler (2009).

Reihenfolge als gleich wahrscheinlich eingeschätzt werden, obwohl die Kombination „rot-blau“ auf zwei Arten gezogen werden kann und damit wahrscheinlicher ist. Diese Schwierigkeit erwies sich bis ins junge Erwachsenenalter hinein als stabil (Fischbein & Schnarch, 1997). Es ist demnach zu vermuten, dass diese Schwierigkeiten auch bei einfachen Aufgaben zum Wahrscheinlichkeitsvergleich bei Schülerinnen und Schülern zu Tage treten.

1.2 Basisproblem inferenzstatistisches Schließen

Als weiteres Basisproblem im Bereich Stochastik soll in diesem Artikel das elementare Schließen von einer Stichprobe auf ein Merkmal der Grundgesamtheit behandelt werden. Nisbett et al. (1983) weisen darauf hin, dass die Generalisierung über induktive Klassenbildung eine der wichtigsten Problemlöseaktivitäten des Menschen ist, die z. B. zentral für den Konzepterwerb im Allgemeinen ist. Maker und Rubin (2009) beschreiben als Charakteristikum des inferenzstatistischen Schließens das Generalisieren über vorliegende Daten hinaus, wobei die vorliegenden Daten als Evidenz in dem Sinne genutzt werden, dass die Generalisierung durch die Daten begründet wird.

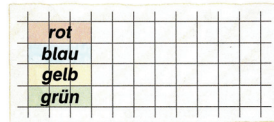
Auch wenn die beurteilende oder schließende Statistik explizit als Lerngegenstand erst in den Bildungsstandards für die Allgemeine Hochschulreife (KMK, 2012) genannt wird, ist das dahinterstehende Prinzip ein grundlegender Baustein stochastischen Verständnisses. Vorbereitet wird inferenzstatistisches Schließen im Elementarbereich durch die Gewinnung und Aufbereitung von Daten. Dabei sollen Daten strukturiert und – je nach Fragestellung und Datencharakteristika – geeignet repräsentiert werden sowie im entgegengesetzten Prozess Informationen aus Repräsentationen entnommen werden. Darauf aufbauend werden in der Sekundarstufe einfache Interpretationen von Daten, z. B. mit Hilfe von Mittelwerts- und Streumaßen gefordert (KMK, 2003; 2004b). Für diese Fähigkeiten ist es allerdings notwendig zu erkennen, dass man – obwohl es sich um zufallsbehaftete Situationen handelt – aus einer Stichprobe zumindest in einem gewissen Maß Rückschlüsse auf die zugrundeliegende Population machen kann.

An einem Aufgabenbeispiel aus der Grundschule sei dies verdeutlicht (Abbildung 1). Während die erste Teilaufgabe einen rein deskriptiven Charakter hat und die Lernenden eine strukturierte Repräsentation der Ergebnisse des wiederholten Zufallsexperiments erstellen, erfordert die Fragestellung der Teilaufgaben b und c den Schluss von den Ergebnissen auf die Grundgesamtheit, in diesem Fall den Beutelinhalt. Für diese Teilaufgaben ist also ein Grundverständnis inferenzstatistischen Schließens erforderlich.

Im Bereich der Mathematikdidaktik wurde der Begriff des *informal inferential reasoning* vorgeschlagen, um solche, für die beurteilende Statistik propädeutischen, Arbeitsweisen zu beschreiben (für einen Überblick siehe Harradine, Batanero &

Rossman, 2011). Zieffler et al. (2008) fassen unter diesen Begriff drei Problemtypen: (a) Das Generalisieren von Stichprobeneigenschaften auf Populationseigenschaften, (b) den Vergleich von zwei Grundpopulationen auf Basis zweier Stichproben und (c) die Einschätzung, ob gegebene Daten gewisse Erwartungen oder Modelle stützen oder nicht.

- a) Jede Gruppe bekommt ein Säckchen mit den gleichen sechs Steckwürfeln.
Ihr dürft 10-mal ziehen. Legt den Würfel nach jedem Ziehen zurück ins Säckchen.
Erstellt ein Balkendiagramm:



- b) Überlegt jetzt: Was könnte im Säckchen sein?
c) Könnte es eines von diesen Säckchen sein?
Könnt ihr ein Säckchen ausschließen?

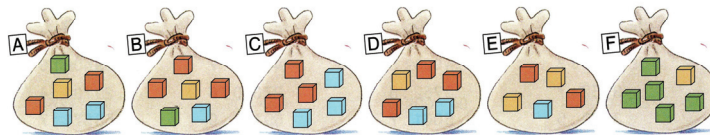


Abbildung 1: Aufgabenbeispiel aus Zahlenzauber 4 (Betz et al., 2012)

Es liegen kaum mathematikdidaktische Forschungsbefunde zu Fähigkeiten von Kindern im Elementar- und beginnenden Sekundarstufenalter in diesem Bereich vor. Zu erwähnen ist die Untersuchung von Eichler und Vogel (2012), in der sie Aufgaben aus allen drei Problemtypen mit dem speziellen Fokus auf die Komplexität der benötigten mentalen Modellierungen einsetzten. Für eine theoretische Bestimmung der Aufgabenkomplexität beschreiben sie vier schwierigkeitsbeeinflussende Faktoren. Sind z. B. für die betrachtete statistische Situation sowohl Informationen zu den (1) Zufallsobjekten als auch ein (2) Datensatz gegeben, ist die (3) Datengenerierung zudem transparent und keine (4) mentale Simulation notwendig, so ist eine Aufgabe als wenig komplex einzustufen. Entsprechend werden Aufgaben mit lückenhafter Information zu den Zufallsobjekten oder Aufgaben, für die mentale Simulationen – z. B. durch eine Vorhersage von zukünftigen Ereignissen – benötigt werden, als komplexer eingestuft. Es zeigt sich, dass die theoretisch angenommenen Komplexitätsstufen bei Lernenden der Jahrgangsstufen 4 und 6 abgebildet werden können. Weiter untersuchten Watson und Moritz (1998) und Watson (2001) Fähigkeiten bei Aufgaben des Problemtyps (b). Dabei zeigt sich, dass bereits in Jahrgangsstufe 3 der Vergleich von Populationen auf Grundlage von Stichproben unter der stark vereinfachenden Einschränkung der gleichen Stichprobengröße durch grafische Verfahren geleistet werden kann. Allerdings sind Ver-

gleichsaufgaben mit unterschiedlicher Stichprobengröße auch für manche Lernende am Ende der Sekundarstufe noch schwierig. Wollring (1994c) führt in der 1. Jahrgangsstufe eine Versuchsserie mit Zylinderwürfeln durch, wobei in der Aufgabenstellung alle drei Problemtypen des informal inferential reasoning verwendet werden. Hier zeigt sich, wie frequentistische Argumentationen präformal geleistet werden können, allerdings durch Formulierungsschwierigkeiten und fehlende Erfahrung erschwert werden. Es treten bei Vorhersagen zudem starke Reihenfolgeeffekte auf. Ferner berichtet Wollring (1994c) Untersuchungen zum Problemtyp (a) in Form von Rekonstruktionsaufgaben für Urnen und Glücksräder. Dabei kann er frühe Fähigkeiten zum inferenzstatistischen Schließen in einer experimentellen Umgebung mit Kindern ab dem 1. Schuljahr sichtbar machen.

Weiter lohnt sich ein Blick in die psychologische Literatur zum wissenschaftlichen Denken, um den Forschungsstand zum inferenzstatistischen Schließen aufzuarbeiten. Als ein Problem werden induktive Schlüsse von (Kovarianz-)Daten auf Wirkzusammenhänge untersucht, die Aufgaben bilden also meist Problemtyp (b) oder Problemtyp (c) mit einem Fokus auf Wirkmodelle ab. Dabei hat die Koordinierung von Evidenz einerseits und erklärenden Theorien sowie persönlichen Überzeugungen andererseits besondere Aufmerksamkeit erfahren. Alltagssituationen und -erfahrungen der Schülerinnen und Schüler, deren erkenntnistheoretisches Verständnis sowie deren metakognitive Fähigkeiten spielen hier eine bedeutende Rolle (für einen Überblick siehe Zimmerman, 2007). Es finden sich Hinweise darauf, dass schon Schülerinnen und Schüler zu Schulbeginn in gewissen Kontexten Daten als Erkenntnisquelle anerkennen: Bereits in Alter von 6 Jahren können Kinder zwischen ihrem eigenen Wissen und dem (möglicherweise dazu widersprechenden) evidenzbasierten Erkenntnissen dritter Personen sehr gut unterscheiden (Ruffman et al., 1993; Koerber et al., 2005).

Zum Problemtyp (a), der Generalisierung von Merkmalen einer Stichprobe auf Populationseigenschaften, existieren weiter die Studien zum induktiven Schließen von Gutheil und Gelman (1997), wobei hier auf Wissen über alltägliche Kategorisierungen (z. B. die Kategorie Schmetterlinge oder allgemeinere Tierklassen) zurückgegriffen wird. Dabei soll von Eigenschaften einer kleinen Gruppe einer Tierart auf die Eigenschaften eines anderen Vertreters dieser Art geschlossen werden. Für Erwachsene zeigte sich, dass die Stichprobengröße und die Variabilität (oder besser Repräsentativität der Stichprobe bezüglich der Stammkategorie) einen Einfluss auf den Schluss hat, in dem Sinne, dass größeren und variableren Stichproben eine bessere Vorhersagekraft zugebilligt wird (Gutheil & Gelman, 1997; vgl. auch Nisbett et al., 1983). Kinder in der 2. und 3. Jahrgangsstufe zeigten bei diesen Aufgaben noch erhebliche Schwierigkeiten. Insbesondere gab es keine Hinweise darauf, dass die Kinder in diesem Alter die Merkmale Stichprobengröße und Stichprobenvariabilität getrennt voneinander nutzten, um induktive Schlüsse zu ziehen. Bei diesen Aufgaben muss allerdings zum einen alltägliches Wissen über Katego-

rien und deren Bandbreite verfügbar sein, um Ähnlichkeiten zwischen Stichproben und der Kategorie betrachten zu können. Zum anderen muss von einer Stichprobe auf die Grundgesamtheit geschlossen werden, damit dann wiederum eine Vorhersage für spezifische Vertreter der Grundgesamtheit getroffen werden kann. Somit handelt es sich bei den hier eingesetzten Problemen um mehrschrittige Probleme, die zudem vom Vorwissen im jeweiligen Bereich abhängig sind. Einsichten in basale Fähigkeiten zum Schluss von einer Stichprobe auf Merkmale der Grundgesamtheit können damit kaum erlangt werden.

Auch im Bereich des inferenzstatistischen Schließens spielen die im Abschnitt zum Wahrscheinlichkeitsvergleich beschriebenen (Fehl-)Vorstellungen und Schwierigkeiten, die aus den bekannten Heuristiken erwachsen, eine Rolle. Insbesondere sei hier erneut auf die Repräsentativitätsvorstellung verwiesen, die allerdings im Kontext der schließenden Statistik kaum als *Fehl*vorstellung bezeichnet werden kann: Eine Heuristik, die auf Ähnlichkeiten zwischen Stichproben- und Populationsmerkmalen fokussiert, ist in Anwendungskontexten oft zielführend (Kahneman & Tversky, 1973). Werden dabei Verhältnisse von Merkmalsausprägungen beachtet, so führt eine Repräsentativitätsheuristik eher zum Erfolg als die Beachtung der absoluten Anzahlen von einzelnen Merkmalsausprägungen³. Zudem hängt der Erfolg einer Repräsentativitätsheuristik davon ab, wie variabel oder homogen das betrachtete Merkmal in einer Grundgesamtheit und wie groß und variabel die Stichprobe im Verhältnis dazu ist (Nisbett et al., 1983; Pfannkuch, 2005).

Allerdings ist bekannt, dass selbst Erwachsene teils eine gewisse Insensibilität in Bezug auf die Größe von Stichproben zeigen. Sie beachten dann nicht, dass im Allgemeinen eine größere Stichprobe zu einer besseren Vorhersagegenauigkeit führt, so dass sie kleinen und großen Stichproben dieselbe Qualität zusprechen (Tversky & Kahneman, 1971). Dieses auch als *effect of the sample size* beschriebene Phänomen ist mit der o. g. Repräsentativitätsheuristik verwandt, indem die Repräsentativitätsaussage des Gesetzes der großen Zahlen auf kleine Anzahlen übergeneralisiert wird. Fischbein und Schnarch (1997) konnten zwischen den Jahrgangsstufen 5 und 11 sogar eine Verfestigung dieser Insensibilität nachweisen. Masnick und Morris (2002) fanden für Kinder in Jahrgangsstufe 3 gar einen inversen Effekt, so dass größere Stichproben für die Kinder zu weniger sicheren Aussagen führten. Die Befundlage ist allerdings uneindeutig, da Kinder ab 11 Jahren in anderen Untersuchungen wiederum kompetent mit unterschiedlichen Stichprobengrößen umgehen (z. B. Inhelder & Piaget, 1975). Auch Wollring (1994c) findet in

³ Unter Verwendung eines frequentistischen Wahrscheinlichkeitsbegriffs wird die Wahrscheinlichkeit eines Ereignisses ja sogar als seine (theoretische) relative Häufigkeit bei unendlich vielen Versuchen angenommen. Das Bindeglied zwischen frequentistischem und axiomatischem Wahrscheinlichkeitsbegriff ist demzufolge das Gesetz der großen Zahlen, das die (statistische) Konvergenz der relativen Häufigkeit beschreibt.

frequentistischen Situationen Kompetenzen von Grundschulkindern, die auf ein intuitives Verständnis des Gesetzes der großen Zahlen hinweisen. Sedlmeier und Gigerenzer (1997; 2000) zeigen, dass die Berücksichtigung von Stichprobengrößen – zumindest für Erwachsene – wesentlich davon abhängt, welche Art von Aufgabe sie bearbeiten. Dabei erweisen sich Erwachsene als weniger anfällig für die Stichprobenvernachlässigung, wenn sie Parameter der Grundgesamtheit (und nicht der Stichprobenverteilung) schätzen sollen.

Insgesamt bleibt offen, inwiefern junge Schülerinnen und Schülern bei basalen Aufgaben des inferenzstatistischen Schließens von einer Stichprobe auf ein Merkmal der Grundgesamtheit von Repräsentativitätsheuristiken Gebrauch machen können und ob ihre Urteile von der Größe der Stichprobe beeinflusst werden.

1.3 Forschungsfragen

Zusammenfassend lässt sich feststellen, dass der Umgang mit zufallsbehafteten Situationen als zunehmend bedeutsamer in der Grundschule angesehen wird und präpädagogisch für den Aufbau von stochastischen Fähigkeiten im Laufe der Sekundarstufe gilt. Zudem sind stochastische Fähigkeiten notwendig, um Kompetenzen im wissenschaftlichen Denken zu erwerben. Andererseits handelt es sich aber um einen Inhaltsbereich, der selbst älteren Schülerinnen und Schülern und Erwachsenen nicht unbedingt leicht zugänglich ist, wobei die Befundlage im Grundschulbereich wenig aussagekräftig ist. Die vorliegende Studie hatte entsprechend das Ziel, hier detailliertes Wissen über die Fähigkeiten am Ende der Grundschule und zu Beginn der Sekundarstufe zu bekommen. Wir untersuchten entsprechend querschnittlich Kinder im 4. und 6. Schuljahr. Da der Übergang zur Sekundarstufe zudem in einem Alter stattfindet, in dem Kinder nach Piaget in die Phase der formalen Operationen eintreten, ist zu erwarten dass die beschriebenen typischen Schwierigkeiten – falls nachzuweisen – bei den älteren Kindern weniger deutlich ausgeprägt sind. Dabei interessierten besonders die bisher wenig untersuchten Fähigkeiten beim Umgang mit den Basisproblemen Wahrscheinlichkeitsvergleich und inferenzstatistisches Schließen wie oben beschrieben. Im Fokus standen die folgenden Forschungsfragen, die auf eine Beschreibung des Status Quo für die einzelnen Basisprobleme abzielten:

- Welche Fähigkeiten haben die Schülerinnen und Schüler in Bezug auf die genannten Basisprobleme?
- Können bekannte (Fehl-)Vorstellungen oder Schwierigkeiten nachgewiesen werden?
- Welche Entscheidungsheuristiken können dabei beobachtet werden?
- Können Fähigkeitsunterschiede für Kinder der Jahrgangsstufen 4 und 6 festgestellt werden?

2 Design der explorativen Studie und Methoden

Die Beantwortung dieser Fragen setzt geeignete Messinstrumente voraus. Dabei sind im Primärbereich besondere Bedingungen zu beachten. Sie sollen hier anhand von vier Problemfeldern diskutiert werden.

2.1 Allgemeine Gestaltungsprinzipien

2.1.1 Erhebungsmodus

In den niedrigen Jahrgangsstufen eignen sich schriftliche Erhebungen nur bedingt. Insbesondere können komplexere Aufgaben nicht in Textform präsentiert werden und längere Begründungen nicht schriftlich eingefordert werden. Auch wenn die Lese- und Schreibfähigkeiten der Schülerinnen und Schüler bereits fortgeschritten sind kann eine rein sprachliche Präsentation der Aufgabenstellungen ungeeignet sein, da zumindest die Fehlvorstellung der *conjunction fallacy* auch auf einem linguistischen Problem beruhen kann (Fiedler, 1988).

Zur vergleichenden Untersuchung von Fähigkeiten bietet sich entsprechend die Interviewform an. In altersadäquaten, spielähnlichen Kontexten können dabei die Fähigkeiten der einzelnen Kinder aktiviert werden und unabhängig von Lese- oder Rechtschreibkenntnissen erfasst werden, wobei allerdings dennoch verbale Fähigkeiten einen Einfluss auf die Untersuchungsergebnisse haben können (s. a. Spielinterviews bei Wollring, 1994a; 1994b; 1994c). Durch eine strukturierte Interviewführung bleibt gewährleistet, dass die Äußerungen der Kinder vergleichbar sind und möglichst wenig durch den jeweiligen Interviewer beeinflusst werden. Mit Hilfe von speziellen Kontrollaufgaben kann zudem überprüft werden, ob die Kinder das Aufgabenformat verstanden haben. Für das hier geschilderte Projekt wurden die Interviews zudem videografiert, um im Anschluss eine Analyse der Begründungen durchführen zu können.

2.1.2 Sprachliche Gestaltung

Stochastische Inhalte nehmen im realen Unterricht der Grundschule eher wenig Raum ein, sodass man nicht von einer standardisierten Sprache und im Unterricht erworbenen tragfähigen Grundkonzepten ausgehen kann. Auch wenn die Bildungsstandards für die Grundschule (KMK, 2004a) und dementsprechend viele Lehrpläne und Rahmenrichtlinien in Deutschland den Inhaltsbereich nennen, dürften bei manchen Kindern stochastische Fähigkeiten eher im Alltag als in der Schule erworben sein (vgl. Fischbein, 1975, Begriff der Primärintuition). Beim Blick in aktuelle Schulbücher finden sich wenige Aufgaben aus diesem Bereich, was diese Vermutung unterstützt. Sollen also frühe Fähigkeiten erhoben werden, so muss dies entsprechend auch in der sprachlichen Gestaltung berücksichtigt werden. Es wurde deshalb in dieser Studie gänzlich darauf verzichtet, fachsprachliche Begriffe – wie *wahrscheinlich*, *sicher* und *unmöglich*, *Stichprobe*, *Ereignis*, *Grundgesamt-*

heit etc. – im Interview zu verwenden. Wie die Operationalisierung der Fähigkeiten zu den genannten Basisproblemen im Detail realisiert wurde, ist unten beschrieben.

2.1.3 Darstellungsformen

Darstellungen nehmen in der Kommunikation mit den Schülerinnen und Schülern eine zentrale Rolle ein. Sie können aber auch die Schwierigkeit der Aufgabenstellung beeinflussen, z. B. wenn sie so gestaltet sind, dass sie eine hohe (überflüssige) Arbeitsgedächtnisbelastung darstellen. Im Idealfall können Darstellungen sogar dazu beitragen, das Arbeitsgedächtnis zu entlasten und so Kindern helfen, sich auf Problemlöse- und Argumentationsprozesse zu konzentrieren (vgl. *cognitive load theory*: Sweller, van Merriënboer & Paas, 1998). Bei der Präsentation von Anzahlen kann das Gedächtnis im Speziellen weiter entlastet werden, wenn durch strukturierte und bekannte Darstellung die (Quasi-)Simultanerfassung von Anzahlen unterstützt wird. Im Primarbereich werden hier üblicherweise 5-er und 10-er Strukturierungen genutzt und sollten Schülerinnen und Schülern vertraut sein (vgl. „Rechenschiffe“, Padberg & Benz, 2011).

Im Zusammenhang mit zufallsbehafteten Situationen geht der Nutzen spezieller Darstellungsformen für die Grundpopulation, die Ergebnisse einer Stichprobenziehung oder die Ereignisse allerdings deutlich darüber hinaus, da die Darstellungen gewisse Charakteristika des zugrundeliegenden mathematischen Modells, des Zufallsraums, abbilden können. Die Darstellungsform kann damit unterstützend wirken, indem sie die Aktivierung der zur stochastischen Situation passenden mentalen Modelle bei den Schülerinnen und Schülern erleichtert (Schnotz & Bannert, 1999). So könnte ein Urneninhalt z. B. fotorealistisch präsentiert werden oder aber bereits – in Abstraktion der Eigenschaften – die Kugeln farblich sortiert und strukturiert darstellen (Abbildung 3 links). Im letzten Fall kann die Darstellung stochastische Denkprozesse unterstützen, da die für das Urnenmodell relevante Sichtweise auf die Farbverhältnisse leichter eingenommen werden kann. Natürlich muss vor allem bei Kindern der Primarstufe dabei sichergestellt werden, dass keine Fehlvorstellung einer „geordneten“ Urne entsteht, was jedoch deutlich gemacht werden kann, wenn im Interview sowohl die reale Urne als auch die dazugehörige Verteilung in strukturierter Darstellung zur Verfügung steht.

Ebenso kann die Wahl eines Protokolls für Ergebnisse von Stichprobenziehungen unterschiedlich gut die strukturellen Eigenschaften des zugrundeliegenden stochastischen Problems abbilden. Wollring (2007) prägte dafür den Begriff des Dokumentationsraums, der eine Artikulationsunterstützung darstellen kann. Je nach Aufgabenstellung können bei evidenzbasierten Aufgaben z. B. Formate, die die Zugreihenfolge wiedergeben („Urliste“), oder Formate die eine proportionale Evaluation der Gewinnanteile erleichtern (z. B. „Strichlisten“, „Streifendiagramme“, „Steckwürfelrekorder“, „Ringrekorder“, Wollring, 1994b; 1994c) geeigneter sein.

Beim Vergleich von Wahrscheinlichkeiten kann die Darstellung z. B. eine Unterereignis-Ereignis Beziehung unterstützen oder aber verschleiern. In dieser Studie wurde darauf geachtet, strukturell unterstützende Darstellungen und Protokolle zu verwenden, wie unten im Detail beschrieben wird.

2.1.4 Antwortformate

Da geschlossene Formate sich in unterschiedlichen Studien als besonders geeignet erwiesen, frühe Fähigkeiten in statistischen Kontexten zu erfassen, wurden auch in dieser Studie ausschließlich geschlossene Antwortformate eingesetzt. Ein Vorteil dieser Antwortformate wird darin gesehen, dass sprachliche Fähigkeiten eine geringere Rolle spielen (Yost, Siegel & Andrews, 1962). Zudem wird durch geschlossene Antwortformate der Raum der möglichen Antworten einerseits eingeschränkt und andererseits strukturiert. Anderson und Schlottmann (1991) z. B. diskretisierten die (kontinuierliche) Wahrscheinlichkeitsskala äquidistant, indem sie 17 unterschiedlich lange, der Länge nach geordnete Holzstäbe verwendeten, so dass die Wahrscheinlichkeit also über die Stablänge repräsentiert wurde. Damit konnten sie bereits bei 4-Jährigen ein abgestuftes Verständnis für die Wahrscheinlichkeit von Ereignissen in Urnenkontexten nachweisen. Auch in den Studien von Fischbein (1975) zeigte sich, dass frühe Fähigkeiten beim Umgang mit Unsicherheit sichtbar gemacht werden können, wenn mit Hilfe von geschlossenen Antwortformaten Entscheidung getroffen werden sollen. Allerdings lassen geschlossene Antwortformate nur bei sehr speziellen Aufgabenstellungen einen Rückschluss auf den Lösungsprozess zu. In dieser Studie wurde deswegen bei den Aufgaben zum Wahrscheinlichkeitsvergleich und zum inferenzstatistischen Schließen zusätzlich eine Begründung für die getroffenen Entscheidungen eingefordert, um Hinweise auf die verwendeten Heuristiken zu erhalten. Wie die hier diskutierten Gestaltungsprinzipien in den Aufgaben zu den drei Basisproblemen im Detail umgesetzt wurden, ist aus den folgenden Beschreibungen ersichtlich.

2.2 Erhebung der Fähigkeiten zur Wahrscheinlichkeitseinschätzung

In der Untersuchung wurden sechs Aufgaben zum Verständnis der stochastischen Konzepte „sicheres Ereignis“, „wahrscheinliches, aber nicht sicheres Ereignis“ und „unmögliches Ereignis“ genutzt. Erhoben werden sollten die Fähigkeiten, vorgegebene Ereignisse einzuschätzen, wobei allerdings basale Fähigkeiten überprüft werden sollten und damit keine feinere Unterscheidung als die oben genannte überprüft werden sollte. Die Erhebungsinstrumente wurden wie beschrieben für den Altersbereich entwickelt und erwiesen sich als geeignet, ohne Rückgriff auf eine spezielle Terminologie ein intuitives Verständnis für die Wahrscheinlichkeit einzelner Ereignisse zu erfassen (Lindmeier et al., 2011; Reiss et al., 2013).

Dazu mussten die Schülerinnen und Schüler in einem simulierten Glücksspiel die Wahrscheinlichkeit einschätzen, aus mehreren Beuteln mit jeweils einem Zug ei-

nen blauen Würfel zu ziehen. Die Beutel enthielten Würfel in zwei verschiedenen Farben mit unterschiedlicher Zusammensetzung. So gab es z. B. einen Beutel mit sechs blauen und vier roten Würfeln. Ein anderer Beutel beinhaltete 99 rote und einen blauen Würfel. Der Inhalt der Beutel wurde zudem durch strukturierte Schaufeln grafisch repräsentiert. Die Aufgabe bestand darin zu entscheiden, ob es bei einmaligem verdeckten Ziehen aus diesem Beutel möglich ist, einen blauen Würfel zu ziehen (Impulsfrage: „Stell dir vor, ich ziehe einen Würfel aus dem Beutel. Überlege: Werde ich einen blauen Würfel ziehen?“). In Tabelle 1 sind drei Beispielaufgaben skizziert.

Die Kinder wurden dann aufgefordert, auf vorgegebene Felder einer Antwortskala zu deuten und somit anzugeben, wie wahrscheinlich es ist, dass ein blauer Würfel gezogen wird (Zielereignis). Die Antwortskala wurde diskretisiert, so dass neben den beiden extremen Feldern (entsprechend Wahrscheinlichkeiten von $p = 1$ und $p = 0$) zwei Wahrscheinlichkeitsbereiche für $0 < p < 1$ angeboten wurden. Somit konnten zwar zwei unterschiedliche Grade an Sicherheit kommuniziert werden, diese wurden in der Auswertung jedoch nicht differenziert. Gemäß dem o.g. sprachlichen Gestaltungsprinzip wurden Bezeichnungen gewählt, die kein formales Verständnis voraussetzen. Dazu wurde für die Randbereiche auf natürliche Häufigkeiten zurückgegriffen („es passiert auf jeden Fall“, „es passiert auf keinen Fall“). Im Mittelbereich erwiesen sich in Voruntersuchungen Umschreibungen unter Verwendung des Ausdrucks „es ist möglich“ als verständlich („es ist möglich, aber eher nicht“, „es ist gut möglich“).

In einer Aufgabe war es unmöglich, einen blauen Würfel zu ziehen. In zwei Fällen lag die Wahrscheinlichkeit bei über 50%, sodass es „gut möglich“ war, einen blauen Würfel zu ziehen. In den übrigen drei Fällen war es zwar möglich, einen blauen Würfel zu ziehen, aber die Wahrscheinlichkeit dafür war sehr klein ($p \leq 10\%$). Diese drei Aufgaben können als Indikatoren für Fähigkeiten im Umgang mit unwahrscheinlichen Ereignissen verwendet werden. Damit waren die Aufgaben insbesondere dazu geeignet, mögliche Schwierigkeiten von Kindern bei der Unterscheidung zwischen unmöglichen und wenig wahrscheinlichen Ereignissen zu erfassen (Lindmeier et al., 2011; Reiss et al., 2013).

2.3 Erhebung der Fähigkeiten zum Wahrscheinlichkeitsvergleich

Zur Erfassung der Fähigkeiten zum Wahrscheinlichkeitsvergleich wurde mit den Schülerinnen und Schülern ein Glücksspiel simuliert (s. a. Lindmeier et al., 2012). Grundlage war ein Beutel mit 15 blauen und zehn roten Würfeln, aus dem mit Zurücklegen gezogen wurde. Auf Karten wurden Gewinnereignisse präsentiert, so dass z. B. mit der Karte in Abbildung 2 links gewann, wer in zwei Zügen zuerst einen roten und dann einen blauen Würfel zog. Den Schülerinnen und Schülern wurden nun je Paare von Karten vorgelegt. Mit der Frage „Was glaubst du, welche Karte ist die bessere? Mit welcher gewinnst du eher?“ wurden sie aufgefordert, die

Gewinnwahrscheinlichkeiten der Karten im Vergleich zu evaluieren. Die Karten konnten dabei entweder zwei Züge (Abbildung 2 links) oder aber nur einen Zug (ohne Abbildung) erfordern. Es wurde außerdem ein grauer Symbolstein eingesetzt, der für das (sichere) Ereignis „ein roter oder ein blauer Würfel wird gezogen“ stand (Abbildung 2 rechts).

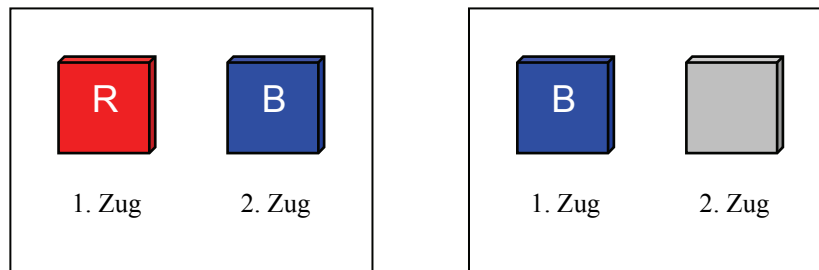


Abbildung 2: Gewinnkarte „Erst rot, dann blau“ (links) und Gewinnkarte „Erst blau, dann rot oder blau“ (rechts)

Acht verschiedene Kartenpaare (siehe Tabelle 2) wurden mit den Kindern bearbeitet. Dabei stand in drei Fällen eine Gewinnkarte für ein Unterereignis der anderen Gewinnkarte (Beispiel: rot gegen rot/blau). Diese Aufgaben können gelöst werden, ohne dass man die Farbzusammensetzung des Beutels beachtet, indem man die strukturellen Eigenschaften des Ergebnisraums nutzt und die Karten vergleicht. Vier weitere der geforderten Vergleiche können ebenfalls ohne Berechnung der Wahrscheinlichkeiten gelöst werden, allerdings muss die Beutelzusammensetzung berücksichtigt werden (Beispiel: rot/blau gegen blau/grau). Ein Vergleich (rot gegen blau/blau) kann ohne exakte Berechnung der Wahrscheinlichkeiten nicht gelöst werden, da einmal ein Zug, im anderen Fall zwei Züge erforderlich sind und damit die Wahrscheinlichkeiten sehr nahe beieinander liegen. Solch ein Vergleich von einem einfachen gegen ein zusammengesetztes Ereignis war bei drei Aufgaben zu leisten. Die in Bezug auf die Beutelzusammensetzung repräsentativer erscheinenden Ereignisse traten sowohl als günstigere als auch ungünstigere Wahl auf, so dass bei Anwendung einer Repräsentativitätsheuristik charakteristische Lösungen zu erwarten sind (siehe auch Tabelle 2, letzte Spalte zur Aufgabencharakteristik). Zusätzlich zur Identifizierung der „besseren Karte“ wurde eine Begründung für die Entscheidung des Kindes eingefordert (Impulsfrage: „Warum denkst du, dass diese Karte die bessere ist?“), so dass Entscheidungsheuristiken transparent wurden.

2.4 Erhebung der Fähigkeiten zum inferenzstatistischen Schließen

Zur Erfassung der Fähigkeiten zum inferenzstatistischen Schließen wurde mit den Schülerinnen und Schülern eine Knobelaufgabe simuliert. Die Aufgabe bestand

darin, Beutel mit 15 Würfeln, aber mit unbekannter Zusammensetzung roter und blauer Würfel einer von drei gegebenen möglichen Zusammensetzungen zuzuordnen. Dabei sollte als Entscheidungsgrundlage das Ergebnis einer Stichprobenziehung (Modus: Ziehen eines Würfels mit Zurücklegen) dienen. Insgesamt handelt es sich also um eine Aufgabe, die mit Sedlmeier und Gigerenzer (1997; 2000) günstig sein sollte, um Fähigkeiten von Schülerinnen und Schülern im Umgang mit unterschiedlichen Stichprobengrößen sichtbar zu machen, da sie zu dem erleichternden Aufgabentyp (Schätzung der Häufigkeitsverteilung) gehört.

Sollen Schülerinnen und Schüler in Interviews vom Merkmal einer Stichprobe auf das Merkmal in der Grundgesamtheit schließen, so können die entsprechenden Daten entweder im Interview generiert oder aber vom Interviewleiter zur Verfügung gestellt werden. Im ersten Fall entstehen zwei Schwierigkeiten: Zum einen tendieren Kinder dazu – vor allem, wenn sie keine Sensibilität für unterschiedliche Stichprobengrößen aufweisen – wenige Daten zu generieren (Wollring, 2007). Zudem würde die generierte Evidenz von Interview zu Interview erheblich variieren, so dass keine vergleichbaren Aufgaben zustande kommen und kaum generalisierende Aussagen über die Fähigkeiten der Kinder getroffen werden können. In dieser Studie wurde daher eine Präsentation der Ergebnisse durch den Interviewer vorgenommen. Durch diese Standardisierung musste allerdings in Kauf genommen werden, dass die untersuchten Kinder nicht eigenhändige Erfahrungen mit dem Zufallsgenerator machen konnten.

Allerdings ist bekannt, dass es leichter fällt, inferenzstatistische Probleme zu lösen, wenn der Prozess der Datengewinnung transparent ist (Nisbett et al., 1983, Wollring, 1994c). Dies zeigte sich in Pilotierungen unseres Zugangs vor allem für Schülerinnen und Schüler der vierten Jahrgangsstufe, weswegen die gesamte Datengewinnung per Video präsentiert wurde. Im Video zogen entsprechend Handpuppen aus Beuteln unbekannter Zusammensetzung und protokollierten die Ergebnisse. In Pilotierungen wurde die Akzeptanz der Videos überprüft und es konnten keine Hinweise darauf gefunden werden, dass Schülerinnen und Schüler der anvisierten Jahrgangsstufen den Handpuppen eine Manipulationsabsicht unterstellten oder andere Probleme mit dem Itemformat hatten. Insgesamt war also sowohl der Prozess der Datengenerierung als auch das Datenmaterial vollständig einsichtig. Es musste keine weitergehende mentale Simulation vorgenommen werden, so dass es sich auch mit der Taxonomie für die Aufgabenkomplexität frequentistischer Probleme von Eichler und Vogel (2012) um eine Aufgabenstellung niedriger Komplexität handelt.

Als Protokollart wurde in dieser Studie ein summatives und strukturiertes Protokoll nach gezogener Farbe gewählt, das die Reihenfolge der Züge vernachlässigt (Abbildung 3 rechts). Diese Protokollart sollte den Vergleich der Daten mit den möglichen Grundgesamtheiten und das Erkennen von einfachen multiplikativen Struktu-

ren erleichtern. Zudem sollten auf diese Weise mögliche Reihenfolgeeffekte (vgl. Fischbein & Schnarch, 1997 zu recency- and primacy-Effekten sowie Tversky & Kahneman, 1973 zu Urteilsverzerrungen durch Charakteristika der Zugreihenfolge) abgeschwächt werden.

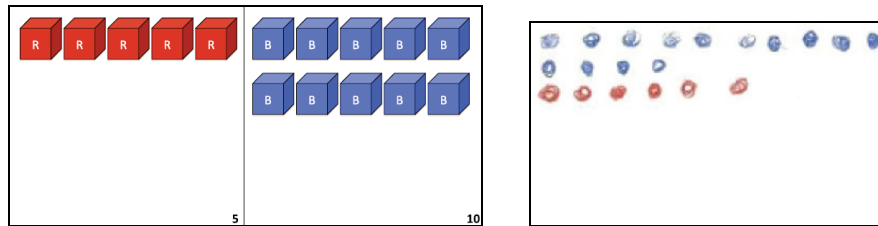


Abbildung 3: Beispiel für Antwortoption in Form einer Beutelzusammensetzung (links) und Datenprotokoll aus 20-facher Ziehung (rechts, Item WIB7)

Die Entscheidung der Schülerinnen und Schüler wurde wieder in einem geschlossenen Antwortformat eingefordert. Dazu wurden die drei möglichen Beutelzusammensetzungen 3 : 12, 5 : 10 und 14 : 1 verwendet (jeweils rot : blau, vgl. Beispiel Abbildung 3 links). Zusätzlich wurden Begründungen verlangt, so dass die Entscheidungsheuristiken der Kinder sichtbar wurden. Mit Zieffler et al. (2008) deckt die hier verwendete informelle inferenzstatistische Aufgabe demnach das Basisproblem der Generalisierung von Merkmalen einer Stichprobe auf die Grundgesamtheit ab (vgl. oben, Problemtyp (a)). Durch die gegebenen Antwortoptionen wird allerdings auch die Evaluation einer Datenbasis vor dem Hintergrund von drei möglichen Modellen (Grundgesamtheiten) ermöglicht, so dass auch Aspekte des Problemtyps (c) in einfacher Form vorliegen.

Insgesamt mussten die Kinder acht Aufgaben bearbeiten, also acht Datensätze einer möglichen Beutelzusammensetzung zuordnen (s. a. Lindmeier & Reiss, 2013). Dabei beruhten vier Datensätze auf einer zehnfachen und vier Datensätze auf einer 20-fachen Ziehung. Die Datensätze waren so gestaltet, dass je ein Experiment mit zehn und ein Experiment mit 20 Zügen das selbe Verhältnis von roten zu blauen Zügen aufwies, um die Sensibilität der Schülerinnen und Schüler für unterschiedliche Stichprobengrößen untersuchen zu können. Zudem wurde sichergestellt, dass die Ereignisse innerhalb des 5%-Annahmebereichs von genau einem der drei Beutel lag, wobei natürlich die Wahrscheinlichkeit der Fehler zweiter Art teilweise beträchtlich hoch lag (vor allem für Zuglänge zehn). Die Aufgaben sind in Tabelle 3 beschrieben.

3 Ergebnisse

3.1 Stichprobe

Insgesamt nahmen an der berichteten explorativen Untersuchung 19 Schülerinnen und 21 Schüler aus Bayern teil, davon acht Mädchen und neun Jungen aus der 4. Jahrgangsstufe sowie elf Mädchen und zwölf Jungen aus der 6. Jahrgangsstufe. Die Schülerinnen und Schüler der 6. Jahrgangsstufe besuchten eine Realschule. Die Teilnahme an der Studie war freiwillig. Im Mittel waren die Kinder der 4. Jahrgangsstufe $M = 10:0$ Jahre:Monate ($SD = 5$ Monate) und die Kinder der 6. Jahrgangsstufe $M = 12:7$ Jahre:Monate ($SD = 7$ Monate) alt. Über unterrichtliche Erfahrungen der Schülerinnen und Schüler im Bereich *Daten, Häufigkeit und Wahrscheinlichkeit* waren keine verlässlichen Informationen zugänglich⁴.

3.2 Fähigkeiten zur Wahrscheinlichkeitseinschätzung

Die jüngeren Kinder konnten im Mittel etwas mehr als vier der sechs Aufgaben zur Wahrscheinlichkeitseinschätzung lösen ($M = 4,12$; $SD = 1,41$). Die Sechstklässler schätzten die Wahrscheinlichkeit von deutlich mehr als vier Aufgaben ($M = 4,52$; $SD = 1,16$) korrekt ein. Der Unterschied zwischen den Jahrgangsstufen ist statistisch allerdings nicht bedeutsam ($t(38) = 0,99$, $p > 0,05$).

Item	Beutelinhalt		p(Blau) Abgebildetes Konzept	Lösungsrate (SD)	
	Würfel	davon blau		4. Schuljahr	6. Schuljahr
GK1	10	6	0,60 Mögliches Ereignis	0,94 (0,24)	0,96 (0,21)
GK6	100	1	0,01 Unwahrscheinliches Ereignis	0,41 (0,51)	0,39 (0,50)
GK4	10	0	0,00 Unmögliches Ereignis	1,00 (0,00)	1,00 (0,00)

Tabelle 1: Beispielitems zur Wahrscheinlichkeitseinschätzung mit Lösungsraten

⁴ Im bayerischen Lehrplan der Grundschule und der Realschule bis einschließlich Schuljahr 6 sind Anteile deskriptiver Statistik, der Umgang mit einfachen Laplace-Experimenten sowie die Beschreibung von Daten mit Hilfe von relativen Häufigkeiten und Lagemaßen vorgesehen. Insgesamt haben Inhalte der Leitidee *Daten, Häufigkeit und Wahrscheinlichkeit* bzw. *Daten und Zufall* in bayerischen Lehrplänen vor der Oberstufe allerdings eine junge Tradition, so dass angenommen werden kann, dass der implementierte Lehrplan vom intendierten abweicht und eher weniger Inhalte realisiert werden. Verlässliche Informationen und Untersuchungen dazu stehen allerdings aus.

In Tabelle 1 sind die Lösungsraten für drei der Aufgaben angegeben, wobei ein mögliches (GK1), ein unwahrscheinliches (im obigen Sinne, $p \leq 10\%$, GK6) und ein unmögliches Ereignis (GK4) ausgewählt sind. Die Lösungsraten lassen erkennen, dass die Schülerinnen und Schüler das unmögliche Ereignis und Ereignisse von mittlerer Wahrscheinlichkeit richtig einschätzen können. Ereignisse mit sehr niedriger Wahrscheinlichkeit werden als unmögliche Ereignisse falsch eingeschätzt, was zu einer niedrigen Lösungsrate führt. Insgesamt kann man jedoch schlussfolgern, dass die Schülerinnen und Schüler über ein altersgemäßes grundlegendes Verständnis von stochastischen Basiskonzepten verfügen. Die Aufgaben zur Überprüfung stochastischer Grundkonzepte wurden von den Schülerinnen und Schülern ähnlich gut wie in anderen Studien in dem Forschungsprogramm gelöst (Lindmeier, et al., 2011; Reiss et al., 2013).

3.3 Fähigkeiten zum Wahrscheinlichkeitsvergleich

Die Schülerinnen und Schüler mussten bei den Aufgaben zum Wahrscheinlichkeitsvergleich eine Karte als die bessere Karte identifizieren, so dass die Lösungsrate alleine auf Grund der Ratewahrscheinlichkeit schon 50% beträgt. Erst eine signifikante Abweichung der Lösungsrate von der Ratewahrscheinlichkeit nach oben kann also als Fähigkeit der Schülerinnen und Schüler gewertet werden. Bei zwei Aufgaben weicht die Lösungsrate tatsächlich nicht signifikant von der Ratewahrscheinlichkeit ab (Tabelle 2).

Der Vergleich der beiden einfachen Karten „blauer Würfel“ gegen „roter Würfel“ (WV1) wurde sehr gut gelöst, was darauf hinweist, dass das Aufgabenformat von den Kindern verstanden wurde. Die Aufgabe „ein roter Würfel“ gegen „zwei blaue Würfel“ (WV7) wurde nicht überzufällig gut gelöst, was den Erwartungen entspricht. Der Vergleich dieser beiden Karten kann mit elementaren Mitteln nicht geleistet werden, weswegen die Lösungsrate der Ratewahrscheinlichkeit entspricht.

Die Lösungsraten der beiden Aufgaben WV6 und WV3 können Aufschluss darüber geben, ob die Schülerinnen und Schüler einer Repräsentativitätsfehlvorstellung unterliegen. Während eine solche Fehlvorstellung für Aufgabe WV6 erleichternd wirken würde, wäre dies für Aufgabe WV3 ein erschwerender Faktor. Tatsächlich wurde die letztere Aufgabe schlechter gelöst als die meisten anderen, während die Lösungsraten für die Aufgabe WV6 als gut erachtet werden können. In der 6. Jahrgangsstufe wurde diese Aufgabe fast von allen Schülerinnen und Schülern gelöst.

Weiter fällt jedoch auf, dass eine Repräsentativitätsvorstellung auch bei den Aufgaben WV5 und WV8 erschwerend wirken könnte. Bei beiden Aufgaben wird allerdings die repräsentativ wirkende Karte (ein roter und ein blauer Würfel) gegen eine Karte mit einem grauen Würfel verglichen. Offensichtlich erkannten die Kinder den Joker-Charakter des grauen Würfels, obwohl dies nicht explizit thema-

tisiert wurde. Die Lösungsraten sind fast durchgängig 100%, so dass bei Einsatz eines sicheren Ereignisses (ein gezogener Würfel wird immer „rot oder blau“ sein) keine eventuell bestehende Repräsentativitätsfehlvorstellung zu Tage treten kann.

Item	p(Ereignis)		Ereignis		Lösungsrate (SD)		Anmerkung
	Karte 1	Karte 2	Karte 1	Karte 2	4. Schuljahr	6. Schuljahr	
WV5	0,24	0,60	RB	BG	1,00* (0,00)	1,00* (0,00)	Repräsentativität (erschwerend)
WV8	0,40	0,24	RG	RB	0,94* (0,24)	1,00* (0,00)	Untereignis vs. Repräsentativität
WV1	0,60	0,40	B	R	0,82* (0,39)	0,96* (0,21)	Elementar- ereignisse
WV6	0,24	0,16	RB	RR	0,88* (0,33)	0,96* (0,21)	Repräsentativität (erleichternd)
WV4	0,60	0,36	B	BB	0,88* (0,33)	0,91* (0,29)	Untereignis einfach vs. zusammengesetzt
WV3	0,36	0,24	BB	RB	0,76* (0,44)	0,87* (0,34)	Repräsentativität (erschwerend)
WV2	0,40	0,24	R	RB	0,47 (0,51)	0,52 (0,51)	Untereignis Repräsentativität einfach vs. zusammengesetzt
WV7	0,40	0,36	R	BB	0,47 (0,51)	0,48 (0,51)	einfach vs. zusammengesetzt

* Abweichung von der Ratewahrscheinlichkeit signifikant, t-Test, $p < 0,05$
 B = blauer Würfel, R = roter Würfel, G = grauer Würfel, d.h. roter oder blauer Würfel

Tabelle 2: Items zum Wahrscheinlichkeitsvergleich mit Lösungsraten
 (sortiert nach Lösungsrate 6. Schuljahr)

In den beiden übrigen Aufgaben mussten die Kinder jeweils eine Karte mit einem Zug gegen eine Karte mit zwei Zügen vergleichen. Dies fiel im einfacheren Fall (blau gegen blau/blau) leicht, während der Vergleich der Karten „ein roter Würfel“ gegen „ein roter und ein blauer Würfel“ nicht von der Ratewahrscheinlichkeit abweichend gelöst wurde. Dieser Befund ist insbesondere interessant, da diese Aufgabe (WV2) mathematisch äquivalent zum Vergleich rot/grau gegen rot/blau (WV8) ist, aber beide Aufgaben deutlich unterschiedliche Lösungsraten aufweisen. Dies kann als Indikator für Schwierigkeiten beim Vergleich von einfachen mit zusammengesetzten Ereignissen gewertet werden. Da jedoch in diesem Vergleich

auch die spezielle Konstellation vorliegt, dass ein Ereignis als Unterereignis des anderen verstanden werden kann, kann diese Aufgabe einen Hinweis darauf geben, wie stark die Fähigkeit der Kinder zum Umgang mit Unterereignissen von der Darstellung der Ereignisse abhängt: Die Modellierung als Ereignisse mit gleicher Zuglänge erleichtert den Vergleich der Ereignisse wesentlich.

Die mittleren Fähigkeiten der Kinder in Jahrgangsstufe 4 und 6 in dieser Studie unterscheiden sich nicht, es kann also auf Basis der Skalenmittelwerte kein Unterschied festgestellt werden (4. Jahrgangsstufe: $M = 6,24$, $SD = 1,03$; 6. Jahrgangsstufe: $M = 6,70$, $SD = 1,15$; $t(38) = 1,31$, $p > 0,05$). In der Studie wurden allerdings auch Begründungen für die Wahl der Schülerinnen und Schüler eingefordert. Diese lassen sich grob kategorisieren, je nachdem ob sie einen probabilistischen Zugang sichtbar werden lassen oder aber die Entscheidung ohne probabilistische Überlegung gefällt wurde. Erwägungen wie „hier [zeigt auf Ereigniskarte] ist nur ein blauer Würfel und den ziehe ich leichter [zeigt auf gegebene Grundgesamtheit mit $p(\text{blau}) = 0,6$]“ würden dabei als probabilistischer Zugang betrachtet werden, auch wenn die Überlegung nicht allgemein zielführend ist. Nicht-probabilistische Begründungen stützen sich z. B. auf eigenen Wunsch („weil ich einen blauen und einen roten ziehen möchte“), auf animistische Vorstellungen (Wollring, 1994a) oder auf physikalische Erklärungen („weil die roten Würfel bestimmt unter den blauen liegen“).

Betrachtet man nun das Begründungsniveau der Schülerinnen und Schüler, so zeigt sich, dass bereits in Jahrgangsstufe 4 ca. 50% der Begründungen probabilistische Überlegungen spiegeln. Der Anteil dieser höherstehenden Begründungen ist in Jahrgangsstufe 6 mit ca. 65% noch etwas größer, was allerdings ebenfalls keinen statistisch signifikanten Anstieg darstellt ($t(38) = -1,85$, $p > 0,05$, Abbildung 4).

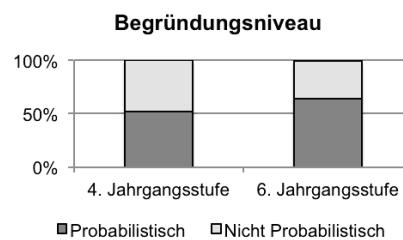


Abbildung 4: Niveau der Begründungen beim Wahrscheinlichkeitsvergleich

3.4 Fähigkeiten zum inferenzstatistischen Schließen

Trotz der sorgfältigen Aufgabenwahl hatten einige Kinder bei den Aufgaben zum inferenzstatistischen Schließen Verständnisschwierigkeiten. Der Zugmodus „mit Zurücklegen“ wurde bereits zu Beginn des Interviews handlungsorientiert einge-

führt und dabei das Verständnis der Schülerinnen und Schüler vom Interviewer überprüft. Im Video war deutlich zu sehen, wie die agierenden Handpuppen die einzelnen Würfel zogen, deren Farbe protokollierten, den Würfel gut sichtbar zurücklegten und den Beutel ausführlich schüttelten. Trotzdem zeigten je fünf Kinder in Jahrgangsstufe 4 und 6 deutliche Irritationen, als im Video 20 Züge aus einem Beutel mit 15 Würfeln getätigt wurden. Sie äußerten explizit, dass dies nicht möglich sei. Diese Kinder wurden in den folgenden Analysen ausgeschlossen, da auf Grund dieser Verständnisschwierigkeiten kein Rückschluss auf deren Fähigkeiten zum inferenzstatistischen Schließen getätigt werden kann. Die Stichprobengröße reduziert sich also im Folgenden auf $N = 12$ (Jahrgangsstufe 4) und $N = 18$ (Jahrgangsstufe 6).

Die übrigen Schülerinnen und Schüler konnten die Aufgaben zum inferenzstatistischen Schließen sehr gut lösen, so dass ein Deckeneffekt beobachtet wurde. Im Mittel wurden in Jahrgangsstufe 4 $M = 7,75$ ($SD = 0,45$) und in Jahrgangsstufe 6 $M = 7,61$ ($SD = 0,61$) der acht Aufgaben gelöst (Tabelle 3). In Bezug auf die Lösungsrate akkumuliert über Aufgaben konnten keine Unterschiede zwischen den Jahrgangsstufen gefunden werden (Mann-Whitney U-Test $U = 140,5$, $p > 0,05$).

Item	h(blau)	Zuglänge	p(Ergebnis aus Zielbeutel)	Lösungsraten		
				4. Schuljahr	6. Schuljahr	Gesamt
WB11	0,50	10	0,14	1,00	0,89	0,93
WB16		20	0,05	1,00	1,00	1,00
WB12	0,30	10	0,30	1,00	1,00	1,00
WB18		20	0,22	1,00	1,00	1,00
WB13	1,00	10	0,50	0,92	0,94	0,93
WB15		20	0,25	1,00	1,00	1,00
WB14	0,70	10	0,26	1,00	1,00	1,00
WB17		20	0,18	0,86	0,74	0,78

Tabelle 3: Items zum inferenzstatistischen Schließen mit Lösungsraten

Da je eine Aufgabe mit zehn und 20 Zügen die gleiche Verteilung roter und blauer Züge aufwies, konnte überprüft werden, ob die Kinder sensibel in Bezug auf die Stichprobengröße waren. Dazu wurden sie am Ende des Interviews gebeten, das Vorgehen der Puppe 1 (je zehn Züge) und der Puppe 2 (je 20 Züge) zu vergleichen und einzuschätzen, welche Puppe sicherer den Beutel vorhersagen kann. Hier zeigte sich, dass nur 58,3% der jüngeren und 77,8% der älteren Kinder erkannten, dass

die größere Datenbasis zu einer sichereren Erkenntnis führt, was jedoch keinen signifikanten Unterschied darstellt ($t(38) = -1,13$, $p > 0,05$).

Zur Auswertung der Entscheidungsheuristiken wurde ein Kodiermanual entwickelt, das wiederum zwischen Antworten mit und ohne Bezug zum probabilistischen Charakter der Aufgaben (s. o. Begründungen zum Wahrscheinlichkeitsvergleich) unterschied. Bei den probabilistischen Begründungen wurden einfache Verweise auf den Zufall oder Glück („wenn man Glück hat, kann man aus diesem Beutel sowas ziehen“) von elaborierteren Entscheidungsheuristiken unterschieden. Diese konnten auf einem Vergleich von (absoluten) Anzahlen („weil hier fünf blaue sind und hier auch“), auf dem Abgleich von Differenzen („weil es hier und hier weniger rote als blaue sind“) oder auf proportionalen Strategien beruhen („weil das doppelt so viele blaue wie rote Würfel sind und hier auch ungefähr doppelt so viele blaue sind“).

In Tabelle 4 ist angegeben, wie die Kategorien über alle Kinder und Aufgaben akkumuliert vergeben wurden. Während proportionale Begründungen so gut wie nicht herangezogen wurden, konnten die meisten Kinder eine Begründung, die sich auf Anzahlen oder Differenzen stützte, explizieren (84,2%). Die Schülerinnen und Schüler der 6. Jahrgangsstufe verwendeten deutlich häufiger Differenzen als Anzahlen zur Begründung, so dass die Besetzung der Kategorien über alle Aufgaben kumuliert für die beiden Jahrgangsstufen signifikant unterschiedlich ist ($\chi^2(5) = 54,98$, $p < 0,001$). Tabelle 4 illustriert diese Tendenz zu höherstehenden Strategien in Jahrgangsstufe 6 für die beiden parallelen Aufgaben, die für die Schülerinnen und Schüler der Studie am schwierigsten waren.

	Akkumuliert über alle 8 Items und 30 Kinder	WIB4		WIB7		
		4. Schuljahr	6. Schuljahr	4. Schuljahr	6. Schuljahr	
Keine Begründung	0,4%	–	–	1	–	
Nicht-probabilistische Begründung	7,9%	2	2	1	1	
Probabilis- tische Be- gründung	Glück/Pech	5,8%	–	–	1	1
	Anzahl	42,5%	7	6	7	8
	Differenz	41,7%	3	9	2	8
	Verhältnis	1,7%	–	1	–	–

Tabelle 4: Entscheidungsheuristiken beim inferenzstatistischen Schließen

4 Zusammenfassung und Diskussion

In dieser explorativen Studie konnte gezeigt werden, dass bereits Schülerinnen und Schüler der 4. Jahrgangsstufe über grundlegende Fähigkeiten in den Basisproblemen der Wahrscheinlichkeitseinschätzung, des Wahrscheinlichkeitsvergleichs (bei stabiler Population) und des inferenzstatistischen Schließens von einem Merkmal einer Stichprobe auf das Merkmal der Grundgesamtheit in einem konstruierten stochastischen Problemkontext verfügen. Allerdings finden sich auch Hinweise auf ganz spezielle Heuristiken und Schwierigkeiten. Auf Grund der Stichprobengröße und -charakteristik sind natürlich keine allgemeingültigen Aussagen zu treffen, die Ergebnisse sind aber gut an bestehende Befunde anzubinden.

So kann der Wahrscheinlichkeitsvergleich bei bestimmten Aufgabenstellungen durch bekannte Heuristiken erleichtert oder beeinträchtigt werden. Während die untersuchten Schülerinnen und Schüler mit Unterereignissen erstaunlich gut umgehen konnten, nutzten sie die Heuristik, Ereignissen, die für die Beutelzusammensetzung repräsentativ scheinen, eine hohe Wahrscheinlichkeit zuzuordnen (*representativeness*). Zudem fielen den Schülerinnen und Schülern Vergleiche von einfachen und zusammengesetzten Zufallsereignissen schwer (*compound and simple events*), während sichere Teilereignisse erkannt wurden (*conjunction fallacy*). In der Tendenz verfügen die Schülerinnen und Schüler der 6. Jahrgangsstufe über ein höheres Fähigkeitsniveau, was sich vor allem in einer besseren Fähigkeit zur Begründung ihrer Entscheidungen niederschlägt.

Das Aufgabenformat zum inferenzstatistischen Schließen war für die meisten Schülerinnen und Schüler in dieser Studie gut geeignet. Allerdings erwiesen sich die Aufgaben als zu leicht, so dass Fähigkeitsniveaus in dieser Stichprobe nicht gut differenziert werden konnten. Dies ist vermutlich auf das spezielle, geschlossene Aufgabenformat und die einfache Problemstruktur zurückzuführen. Beides wurde gewählt, da bisher wenig Erkenntnisse über die Fähigkeiten von Kindern dieser Altersstufe bekannt sind und ein möglichst einfaches Aufgabenformat gesucht wurde. Somit konnte gezeigt werden, dass die meisten Schülerinnen und Schüler von einem gegebenen Datensatz auf ein Merkmal der Grundgesamtheit schließen können. Allerdings konnte in diesem Alter erwartungsgemäß eine gewisse Insensibilität in Bezug auf die Stichprobengröße aufgezeigt werden, so dass selbst in Jahrgangsstufe 6 ungefähr ein Viertel der Kinder nicht erkannte, dass eine höhere Stichprobe zu einem verlässlicheren Urteil führt. Mit Blick auf die Begründungen konnte erneut festgestellt werden, dass proportionale Strategien auch bei diesen Aufgaben selbst in Jahrgangsstufe 6 kaum genutzt werden, sondern Anzahl- und Verhältnisstrategien Vorrang haben. Es liegt nahe, dass Schülerinnen und Schüler auch bei Aufgaben zum inferenzstatistischen Schließen auf eine Repräsentativitätsheuristik zurückgreifen, was zudem verstärkt werden kann, wenn sie noch Probleme im Umgang mit Verhältnissen haben (Inhelder & Piaget, 1958). Die Insensibili-

tät in Bezug auf unterschiedliche Stichprobengrößen kann ebenfalls in diesem Zusammenhang als Indikator für Repräsentativitätsheuristiken interpretiert werden. Ähnliche Entscheidungsheuristiken wurden auch im Bereich der Kovariationsanalyse (Reiss et al., 2011; Ufer, Lindmeier & Reiss, 2011) beobachtet.

Allerdings bleibt zu betonen, dass ein Viertel der untersuchten Kinder in dieser explorativen Studie Probleme beim Verständnis der Aufgaben hatten, obwohl der Zugmodus expliziert und die Ziehung selbst über ein Video präsentiert wurde. Im Gegensatz zu anderen Untersuchungen, wurde also nicht nur das Ergebnis der Ziehung, sondern auch die Genese des Ergebnisses transparent. Durch die Videopräsentation blieb den Kindern zwar ein enaktives Erleben der Stichprobenziehung verwehrt, allerdings konnten so vergleichbare Interviewaufgaben garantiert werden und Stichprobenziehungen bis zu 20 Zügen transparent realisiert werden (siehe Diskussion Kapitel 2.4.). Möglicherweise ist der Zugmodus „Ziehen mit Zurücklegen“ für diese Kinder zu fremd, da er in alltäglichen Situationen kaum relevant ist. Das Aufgabenverständnis könnte durch die Verwendung eines anderen Zufallsgenerators, der noch deutlicher von Zug zu Zug stabil bleibt, unterstützt werden. Wollring (1994c) zeigt z. B. mit den verdeckten Schüttelurnen eine mögliche Alternative auf. Eine weitere Erklärung für die gefundenen Schwierigkeiten liegt eventuell darin, dass die Kinder über wenige Vorerfahrungen im Bereich inferenzstatistisches Schließen verfügen. Die Erfahrungen von Kindern und deren aktive Teilnahme in unterrichtlichen Aktivitäten – z. B. auf Grundlage einer *growing-samples*-Instruktionsheuristik (Bakker, 2004; Ben-Zvi, 2006) – erwies sich im Bereich der schließenden Statistik als erfolgsversprechend. So entwickelten Kinder, die an einer entsprechenden Unterrichtssequenz in der 5. Jahrgangsstufe teilnahmen ein besseres Grundverständnis, bessere Strategien und auch eine geeignetere Sprache für Situationen, in denen datenbasierte Schlussfolgerungen benötigt wurden, wie in Fallstudien aufgezeigt werden konnte (Ben-Zvi, 2006). Wollring (1994c) betont ebenfalls, dass subjektive Erfahrungen beim Begriffserwerb in frequentistischen Zusammenhängen als verständnisfördernd einzustufen sind.

Allerdings könnte bei den teilweise sehr jungen Kindern (4. Jahrgangsstufe) unserer Studien auch noch eine entwicklungspsychologisch bedingte Einschränkung vorliegen, so dass in Situationen unter Unsicherheit der Schluss von einer Stichprobe auf ein Merkmal einer Grundgesamtheit nicht von allen geleistet werden kann. Fischbein et al. (1991) fanden z. B. bei Kindern der 4. Jahrgangsstufe die Vorstellung, dass bei Zufallssituationen generell keine begründete Vorhersage möglich ist. Eine ähnliche Vorstellung könnte auch bei den in dieser Studie eingesetzten Aufgaben zum Tragen gekommen sein, so dass das Basisproblem der Generalisierung an sich nicht zugänglich ist, weil bei unsicheren Situationen für die Kinder kein Rückschluss von der Stichprobe auf die Grundgesamtheit möglich scheint. Dies ist allerdings noch durch weitere Studien zu überprüfen.

Die Interviewform erwies sich jedoch prinzipiell als geeignet, um Fähigkeiten der Schülerinnen und Schüler sichtbar zu machen. Durch den spielerischen Zugang und die Verwendung von Material konnte die neuartigen Aufgaben zum Wahrscheinlichkeitsvergleich und zum inferenzstatistischen Schließen mehrheitlich zugänglich gemacht werden. Insgesamt kann also vermutlich davon ausgegangen werden, dass grundlegende mathematische Fähigkeiten, die zur Evaluation wissenschaftlicher Evidenz benötigt werden, bereits im Primarschulalter größtenteils vorhanden sind. Bekannte Heuristiken lassen sich mit dem gewählten Zugang nachweisen, zeigen aber komplexere Wechselwirkungen als die Literatur vermuten lässt. Sowohl das frühe Auftreten dieser Heuristiken als auch die Stabilität der Befunde können als ein Hinweis gewertet werden, dass sie auf intuitiven Zugängen zum Wahrscheinlichkeitsbegriff beruhen. Die Fähigkeiten der Kinder in Jahrgangsstufe 4 und 6 unterschieden sich in dieser explorativen Studie nur graduell.

5 Unterrichtliche Implikationen

Die vorliegende Studie hatte zum Ziel, den Status Quo der Fähigkeiten von Schülerinnen und Schülern am Ende des Primar- und beginnenden Sekundarstufenalters für Basisprobleme aus dem Bereich *Daten, Häufigkeit und Wahrscheinlichkeit* zu beschreiben. Solche Untersuchungen können dazu beitragen, spezifische (Fehl-)Vorstellungen und Entscheidungsheuristiken aufzudecken, die wiederum als Ausgangspunkt für instruktionale Ansätze dienen können. So zeigt sich beim Wahrscheinlichkeitsvergleich beispielsweise, dass Repräsentativitätsvorstellungen bei ungeschickter Aufgabenwahl überdeckt werden können: Der Joker-Charakter des in der Studie eingesetzten grauen Steins erleichterte den Vergleich wesentlich, die Modellierung der Ereignisse mit unterschiedlicher Zugzahl erschwerte den Vergleich. Durch die Studien konnte z. B. auch aufgezeigt werden, dass Schülerinnen und Schüler mit Abstand am häufigsten Anzahl- und Differenzenstrategien verwenden, allerdings bei geeignetem Datenmaterial (z. B. bei einfacher multiplikativer Struktur der Daten) und strukturierten Darstellungen auch spontan Verhältnisstrategien anwenden. Dies könnte einen Ansatzpunkt zur Förderung proportionaler Strategien darstellen.

Im Bereich des informellen inferenzstatistischen Schließens wurde deutlich, dass im Elementarstufenalter der generalisierende Schluss von einer Stichprobe auf Populationsmerkmale nicht allen Kindern verständlich ist. Damit sind bereits in der Grundschule Vorläuferfähigkeiten für den aktuell in den Bildungsstandards vorgesehenen zweigleisigen Zugang zur Leitidee Daten und Zufall sehr unterschiedlich ausgeprägt. Begriffe und Methoden der Statistik sollen bei diesem Zugang konsequent mit dem klassischen Wahrscheinlichkeitsbegriff vernetzt werden. In Pfannkuchs (2005; 2006) instruktionalem Framework *towards formal inference* (für Klassenstufe 9 bis 12) wird dem Schlüsselprinzip des *sampling reasoning* – als Ar-

gumentation „von der Stichprobe zur Population“, die konträr zum klassischen Prinzip der Wahrscheinlichkeitsrechnung „von der Population zur Stichprobe“ ist – ebenfalls eine zentrale Rolle beim Aufbau formellen inferenzstatistischen Verständnisses in höheren Jahrgangsstufen zugesprochen. Allerdings bilden die üblicherweise skizzierten Aufgaben, die zum Übergang von informellen zu formellen inferenzstatistischen Methoden eingesetzt werden sollen, stets bereits deutlich komplexere Anforderungen ab als in der vorliegenden Studie untersucht. So werden z. B. vergleichende Schlüsse auf Basis von Stichproben für unterschiedliche Populationen notwendig, es sollen wiederholte Simulationen genutzt werden, um Verteilungen zu schätzen oder Vorhersagen auf Basis von Stichprobeninformationen getroffen werden (z. B. Pfannkuch, 2006; Rubin, Hammerman & Konold, 2006; Prömmel & Biehler, 2008). Es wird dabei offensichtlich angenommen, dass Schülerinnen und Schülern das Grundprinzip der Generalisierung vertraut ist, was demzufolge als ein Lernziel in den früheren Jahrgangsstufen verankert sein müsste. Wie die Studie zeigt, kann die Mehrzahl der Schülerinnen und Schüler bereits in der Primarstufe gut mit elementaren inferenzstatistischen Schlüssen umgehen, so dass sich hier die Chance bietet, dieses Basisproblem bereits in diesem Alter zu thematisieren. Mit der *growing samples* Instruktionshuristik (Ben-Zvi, 2006) liegt auch ein ganz konkreter Vorschlag für unterrichtliche Aktivitäten vor, der leicht für die Primarstufe adaptiert werden könnte.

Zusätzlich können die Aufgabenformate aus dieser Studie als Anregung verstanden werden, wie Basisprobleme der Stochastik bereits in frühen Jahren thematisiert werden können. Der formale, in Spielkontexten eingebettete Charakter der Aufgaben erwies sich dabei in unseren Studien bereits in Jahrgangsstufe 4 (bzw. teils bereits in Jahrgangsstufe 2: Lindmeier et al., 2011; Reiss et al., 2011; 2013) als geeignet. In diesem Sinne kann diese explorative Studie vielleicht dazu anregen, den Erfahrungsraum der Schülerinnen und Schüler im Bereich *Daten, Häufigkeit und Wahrscheinlichkeit* auf Basis ihrer Primärintuitionen (Greer, 2001) und im Sinne der Bildungsstandards (KMK, 2004a) bereits frühzeitig zu erweitern, indem Basisprobleme bearbeitet werden – ganz ohne Inhalte und Methoden der späteren Jahrgangsstufen vorwegzunehmen – und somit einen kumulativen Kompetenzerwerb anzubahnen.

Literatur

- Anderson, N. H. & Schlottmann, A. (1991). Developmental study of personal probability. In N. H. Anderson (Hrsg.), *Contributions to information integration theory. Vol. III: Developmental* (S. 110–134). Hillsdale, NJ: Erlbaum.
- Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, 3(2), 64–83.
- Ben-Zvi, D. (2006). Scaffolding students' informal inference and argumentation. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference*

- on *Teaching of Statistics*, Salvador, Bahia, Brazil, 2–7 July, 2006. Voorburg, The Netherlands: International Statistical Institute.
- Betz, B., Dolenc-Petz, R., Gasteiger, H., Gehrke, H., Ihn-Huber, P., Kobr, U. et al. (2012). *Zahlenzauber 4 – Mathematikbuch für die Grundschule*. Ausgabe M. München: Oldenbourg Schulbuchverlag.
- Bullock, M., Sodian, B. & Koerber, S. (2009). Doing experiments and understanding science: Development of scientific reasoning from childhood to adulthood. In W. Schneider & M. Bullock (Eds.), *Human development from early childhood to early adulthood*. Mahwah, NJ: Lawrence Erlbaum.
- Burrill, G. & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In C. Batanero, G. Burrill & C. Reading (Hrsg.), *Teaching statistics in school mathematics – challenges for teaching and teacher education: A joint ICMI/IASE Study: The 18th ICMI Study* (S. 57–69). Dordrecht: Springer.
- Eichler, A. & Vogel, M. (2012). Basic modelling of uncertainty: young students' mental models. *ZDM – The International Journal on Mathematics Education*, 44(7), 841–854.
- Falk, R. & Wilkening, F. (1998). Children's construction of fair chances: Adjusting probabilities. *Developmental Psychology*, 34(6), 1340–1357.
- Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research*, 50(2), 123–129.
- Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Dordrecht: Reidel.
- Fischbein, E. & Schnarch, D. (1997). The evolution with age of probabilistic intuitively based misconceptions. *Journal for Research in Mathematics Education*, 28, 96–105.
- Fischbein, E., Nello, M. S. & Marino, M. S. (1991). Factors affecting probabilistic judgments in children and adolescents. *Educational studies in mathematics*, 22(6), 523–549.
- Gal, I. (2004). Statistical literacy. Meanings, components, responsibilities. In D. Ben-Zvi & J. Garfield (Hrsg.), *The challenge of developing statistical literacy, reasoning and thinking* (S. 47–78). Dordrecht: Kluwer.
- Green, D. R. (1982). *Probability concepts in 11–16 year old pupils*. University of Technology, Loughborough, England. Report of research, CAMET.
- Greer, B. (2001). Understanding probabilistic thinking: The legacy of Efraim Fischbein. *Educational Studies in Mathematics*, 45(1–3), 15–33.
- Gutheil, G. & Gelman, S. A. (1997). Children's use of sample size and diversity information within basic-level categories. *Journal of Experimental Child Psychology*, 64(2), 159–174.
- Harradine, A., Batanero, C., & Rossman, A. (2011). Students and teachers' knowledge of sampling and inference. In C. Batanero, G. Burrill & C. Reading (Hrsg.), *Teaching statistics in school mathematics – challenges for teaching and teacher education: A joint ICMI/IASE Study: The 18th ICMI Study* (S. 235–246). Dordrecht: Springer.
- Inhelder, B. & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York, NY: Basic Books.
- Inhelder, B. & Piaget, J. (1975). *The origin of the idea of chance in children*. New York: NW Norton and Company.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 40(4), 237–251.

- Kahneman, D., Slovic, P. & Tversky, A. (Hrsg.). (1982). *Judgment under uncertainty. Heuristics and Biases*. New York: Cambridge University Press.
- Koerber, S., Sodian, B., Thoermer, C. & Nett, U. (2005). Scientific reasoning in young children. Preschoolers' ability to evaluate covariation evidence. *Swiss Journal of Psychology*, 64, 141–152.
- Kuhn, D. & Pearsall, S. (2000). Developmental origins of scientific thinking. *Journal of Cognition and Development*, 1, 113–129.
- Langrall, C. W. & Mooney, E. S. (2005). Characteristics of elementary school students' probabilistic reasoning. In G. A. Jones (Hrsg.), *Exploring probability in school: Challenges for teaching and learning* (S. 95–119). New York, NY: Springer.
- Lindmeier, A. M. & Reiss, K. (2013). What is in the bag? Elementary school children's understanding of accumulating evidence. In A. M. Lindmeier & A. Heinze (Hrsg.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education, Vol. 3* (S. 281–288). Kiel (Germany): PME.
- Lindmeier, A., Reiss, K., Barchfeld, P. & Sodian, B. (2012). Make your choice – students' early abilities to compare probabilities of events in an urn-context. In T.-Y. Tso (Hrsg.), *Proceedings of the 36th Conference of the International Group for the Psychology of Mathematics Education, Vol. 3* (S. 161–168). Taipei (Taiwan): PME.
- Lindmeier, A., Reiss, K., Ufer, S., Barchfeld, P. & Sodian, B. (2011). Umgang mit wissenschaftlicher Evidenz in den Jahrgangsstufen 2, 4 und 6: Stochastische Basiskonzepte und Kontingenztafelanalyse. In R. Haug & L. Holzäpfel (Hrsg.), *Beiträge zum Mathematikunterricht 2011* (S. 547–550). Münster: WTM.
- Makar, K. & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, 8(1), 82–105.
- Masnack, A. M. & Morris, B. J. (2002). Reasoning from data: The effect of sample size and variability on children's and adults' conclusions. In W. D. Gray & C. Schunn (Hrsg.), *Proceedings of the 24th annual conference of the Cognitive Science Society* (S. 643–648). Mahwah, NJ: Erlbaum.
- Nisbett, R. E., Krantz, D. H., Jepson, C. & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, 90(4), 339–363.
- Padberg, F. & Benz, C. (2011). *Didaktik der Arithmetik*. 4. Aufl. Heidelberg: Spektrum Akademischer Verlag.
- Pfannkuch, M. (2005). Probability and statistical inference: How can teachers enable learners to make the connection? In G. A. Jones (Hrsg.) *Exploring probability in school: Challenges for teaching and learning* (S. 267–294). New York, NY: Springer.
- Pfannkuch, M. (2006). Informal inferential reasoning. In A. Rossmann & B. Chance (Hrsg.) *Proceedings of the 7th International Conference on Teaching Statistics*. Salvador, Bahia, Brazil: International Association for Statistics Education.
- Prömmel, A. & Biehler, R. (2008). Einführung in die Stochastik in der Sekundarstufe I mit Hilfe von Simulationen unter Einsatz der Werkzeugsoftware FATHOM. In A. Eichler & J. Meyer (Hrsg.), *Anregungen zum Stochastikunterricht, Bd. 4* (S. 137–158). Hildesheim: Franzbecker.
- Reiss, K., Barchfeld, P., Lindmeier, A., Sodian, B. & Ufer, S. (2011). Interpreting scientific evidence: Primary students' understanding of base rates and contingency tables. In B. Ubuz (Hrsg.), *Proceedings of the 35th Conference of the International Group for the Psychology of Mathematics Education, Vol. 4* (S. 33–40). Ankara (Turkey): PME.
- Reiss, K., Lindmeier, A. M., Barchfeld, P. & Sodian, B. (2013). Developing problem solving skills in elementary school. The case of data analysis and probability. In Y. Li &

- J. N. Moschkovich (Hrsg.). *Proficiency and beliefs in learning and teaching mathematics. Learning from Alan Schoenfeld and Günter Törner* (S. 33–49). Rotterdam: Sense.
- Rubin, A., Hammerman, J. & Konold, C. (2006). Exploring informal inference with interactive visualization software. In A. Rossmann & B. Chance (Hrsg.) *Proceedings of the 7th International Conference on Teaching Statistics*. Salvador, Bahia, Brazil: International Association for Statistics Education.
- Ruffman, T., Perner, J., Olson, D. R. & Doherty, M. (1993). Reflecting on Scientific Thinking: Children's Understanding of the Hypothesis Evidence Relation. *Child Development*, 64(6), 1617–1636.
- Schnotz, W. & Bannert, M. (1999). Einflüsse der Visualisierungsform auf die Konstruktion mentaler Modelle beim Text- und Bildverstehen. *Zeitschrift für Experimentelle Psychologie*, 46(3), 217–236.
- Sedlmeier, P. & Gigerenzer, G. (1997). Intuitions about sample size: The empirical law of large numbers. *Journal of behavioral decision making*, 10(1), 33–51.
- Sedlmeier, P. & Gigerenzer, G. (2000). Was Bernoulli wrong? On intuitions about sample size. *Journal of behavioral decision making*, 13(1), 133–139.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) (2003). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss: Beschluss vom 4.12.2003*. München: Luchterhand.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) (2004a). *Bildungsstandards im Fach Mathematik für den Primarbereich. Beschluss der Kultusministerkonferenz vom 15.10.2004*. München: Luchterhand.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) (2004b). *Bildungsstandards im Fach Mathematik für den Hauptschulabschluss: Beschluss vom 15.10.2004*. München: Luchterhand.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (KMK) (2012). *Bildungsstandards im Fach Mathematik für die Allgemeine Hochschulreife: Beschluss vom 18.10.2012*.
- Shtulman, A. & Carey, S. (2007). Improbable or impossible? How children reason about the possibility of extraordinary events. *Child Development*, 78, 1015–1032.
- Sweller, J., Van Merriënboer, J. J. & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, 10(3), 251–296.
- Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological bulletin*, 76(2), 105–110.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2), 207–232.
- Ufer, S., Lindmeier, A. & Reiss, K. (2011). Würfel oder Kugel? Entscheidungsstrategien systematisieren und vergleichen. *mathematik lehren*, 168, 18–22.
- Watson, J. M. (2001). Longitudinal development of inferential reasoning by school students. *Educational Studies in Mathematics*, 47(3), 337–372.
- Watson, J. M. & Moritz, J. B. (1998). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics*, 37(2), 145–168.
- Weixler, S. (2009). *Die Entwicklung des intuitiven probabilistischen Denkens bei Schülerinnen und Schülern der Sekundarstufe I unter dem Aspekt des Conceptual Change*. [Dissertation]. München: Universität München.

- Wollring B. (1994a). Animistische Vorstellungen von Vor- und Grundschulkindern in stochastischen Situationen. *Journal für Mathematikdidaktik*, 15(1/2), 3–34.
- Wollring, B. (1994b). Fallstudien zu frequentistischen Kompetenzen von Grundschulkindern in stochastischen Situationen – Kinder rekonstruieren verdeckte Glücksräder. In H. Maier, J. Voigt (Hrsg.). *Verstehen und Verständigung* (IDM-Reihe, 19) 144–181. Köln: Aulis.
- Wollring, B. (1994c). *Qualitative empirische Untersuchungen zum Wahrscheinlichkeitsverständnis bei Vor- und Grundschulkindern*. Universität Münster: Unveröffentlichte Habilitationsschrift.
- Wollring, B. (2007). Den Zufall festhalten – Spielräume und Dokumente bei Zufallsexperimenten für die Grundschule. Lernumgebungen und Versuchsumgebungen zur Stochastik. *Beiträge zum Mathematikunterricht 2007*, 472–475.
- Yost, P. A., Siegel, A. E. & Andrews, J. M. (1962). Nonverbal probability judgments by young children. *Child Development*, 33(4), 769–780.
- Zieffler, A., Garfield, J., Delmas, R. & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 40–58.
- Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, 27, 172–223.

Anschrift der Verfasserinnen

Anke Lindmeier (Korrespondenzautorin)
IPN Kiel
Olshausenstr. 62
24118 Kiel, Deutschland
lindmeier@ipn.uni-kiel.de
www.ipn.uni-kiel.de

Kristina Reiss
TUM School of Education
Arcisstraße 21
80333 München, Deutschland
reiss@tum.de
www.edu.tum.de

Eingang Manuskript: 01.08.2013
Eingang überarbeitetes Manuskript: 14.01.2014
Online verfügbar: 14.04.2014

Anhang: Auszug aus dem Kodiermanual für die Begründungen

Die Art der Begründung wurde unabhängig von der Korrektheit kodiert. Bei Mischformen wurde zu Gunsten des Kindes der konzeptuell höherstehende Begründungsteil kodiert. Dabei gilt die aus der Tabelle ersichtliche Hierarchie.

Typ	Bezeichnung	Erläuterung
0	Keine Begründung	Keine Antwort, nicht verständliche Antwort
1	Nicht-probabilistische Begründungen	Intuitive Begründungen („Weil ich das Gefühl habe, dass es besser ist.“) Definitive Aussagen („Weil ich weiß, dass es so ist.“) Angabe von Wunsch („Weil ich will, dass es so ist.“) Animistische Erklärungen („Weil die Karte gewinnen will.“) Physikalische Erklärungen („Weil der blaue Würfel ganz verdeckt wird.“)
2	Probabilistische Begründungen	<i>Werden nach den Typen 2a–2d differenziert</i>
2a	Glück/Pech als naives Konstrukt	Es wird als Begründung auf den Zufall verwiesen, ohne dass ein Modell, wie die Struktur eines Ergebnisraums oder eine Grundgesamtheit genutzt wird. („Wenn man Glück hat, dann gewinnt diese Karte.“)
2b	Anzahlstrategie	Es werden Informationen über die Grundpopulation und/oder Ereignisse/Ergebnisse in Betracht gezogen, wobei nur absolute Anzahlen einer oder beider Farben zur Begründung genutzt werden. („Weil hier fünf blaue sind und hier auch.“)
2c	Vergleichsstrategie	Es werden Grundpopulation und/oder Ereignisse/Ergebnisse in Betracht gezogen, wobei der Vergleich zwischen den absoluten Anzahlen beider Farben genutzt wird. (Indikatoren z. B. „weniger“, „mehr“, „10 blaue aber nur 5 rote“, „5 rote und sogar 10 blaue“).
2d	Verhältnisstrategie	Es werden Grundpopulation und/oder Ereignisse/Ergebnisse in Betracht gezogen, wobei das Verhältnis der beiden Farben genutzt wird. (Indikatoren z. B. „blau zu rot wie...“, „hier sind doppelt so viele blaue“).