

Klassische Forschung aus der Psychologie als Aufhänger, um Statistik zu unterrichten: ein durch aktives Lernen geprägter Zugang¹

KAREN Y. HOLMES, NORFOLK, UND BRETT A. DODD, VIRGINIA BEACH

¹ Original ‚Teaching Statistics using classic psychology research: an activities-based approach‘ in: *Teaching Statistics* 34 (2012) 1, S. 13–17.
Übersetzung und Bearbeitung:
MANFRED BOROVČNIK, KLAGENFURT

Zusammenfassung: In diesem Beitrag stellen wir eine Sammlung von Lerneinheiten zum aktiven Lernen vor, die aus klassischen psychologischen Studien besteht, welche den sinnvollen Einsatz von Verfahren der beschreibenden und beurteilenden Statistik illustrieren. Es geht uns um einen Zugang zum Lernen, der die Lernenden in Aktivitäten und den Prozess des Lernens bewusst mit einbezieht und das nicht nur der Motivation wegen. [In der englischen Fachliteratur hat sich der Begriff „activity-based learning“ eingebürgert, der nur schwer ins Deutsche zu übertragen ist.]

1 Einleitung

Forschung in den Einführungskurs in Statistik einzubauen ist eine neuartige Strategie und soll Motivation, Selbstwirksamkeit und Leistungsfähigkeit in punkto Statistik erhöhen (Thompson 2000). [Selbstwirksamkeit geht dabei über bloßes Selbstbewusstsein hinaus und hat mit Vertrauen in die eigenen Fähigkeiten zu tun, zukünftige Situationen zu meistern.]

Forschungsaktivitäten bieten Studierenden die Möglichkeit, gemeinsam zu arbeiten, Ideen auszutauschen, Fragen zu stellen und ihre eigenen Vorstellungen vom Lernmaterial zu entwickeln, was ein tieferes begriffliches Verständnis unterstützt (Roseth et al. 2008, Saye & Brush 2004).

In diesem Beitrag beschreiben wir Forschungsaktivitäten, die durch klassische psychologische Studien angeregt sind, um die entscheidende Verbindung zwischen empirischer Forschung, deskriptiver und beurteilender Statistik aufzuzeigen. Diese Forschungsaktivitäten sind für Klassen mit 50 und weniger Studierenden geeignet und setzen einen Zeitrahmen von 20 bis 60 Minuten voraus; sie sollen an eine Einführung in die jeweiligen beschreibenden und beurteilenden Techniken anschließen.

2 Forschungsaktivität 1: Sozialer Einfluss auf Entscheidungen

[Die englischen Fachtermini lauten „informational social influence“, auch „social proof“; gemeint ist ein

Konformitätsverhalten, wonach das eigene Verhalten an jenes anderer *angepasst* wird.]

Vor dieser Lerneinheit geben wir eine kurze Erklärung des *autokinetischen Effekts*. Den Studierenden wird insbesondere mitgeteilt, dass es sich dabei um ein Phänomen der visuellen Wahrnehmung handelt. Ein stationärer Lichtpunkt wird demnach in einem dunklen Raum wahrgenommen, als ob er sich bewegte. Das Phänomen wurde zum ersten Mal durch den Astronomen Alexander von Humboldt (1799) erwähnt [Humboldt glaubte allerdings tatsächlich an eine Bewegung und nannte sie „Sternschwanken“; er unterlag also der Täuschung]; im Folgenden wurde das Phänomen von Charpentier und Aubert in unabhängigen Laborstudien im späten 19. Jh. untersucht. Es war auch Aubert (1887), der den Ausdruck „Autokinetische Empfindung“ geprägt hat – unter diesem Namen ist es bis heute bekannt (Guilford & Dallenbach 1928).

Die subjektive Natur des Phänomens hat Musafir Sherif (1935, 1937) dazu veranlasst, den Effekt zu demonstrieren, wie soziale Faktoren Entscheidungen beeinflussen. Wenn etwa eine Person beobachtet, dass ein Punkt sich drei cm bewegt, später aber hört, dass andere sagen, dass derselbe Punkt sich um sechs cm bewegt hat, wird sie sich – mit dieser Information ausgestattet – sehr wahrscheinlich dieser Norm angleichen und berichten, dass der Punkt sich sechs cm bewegt hätte. Unsere erste Aktivität entspricht der Arbeit von Sherif und ist angelehnt an eine Demonstration von Montgomery & Enzie (1971), in welcher Studierende die Länge einer Zeitspanne schätzen sollen, zuerst schriftlich und dann noch einmal laut.

Zu Beginn dieser Aktivität werden die Studierenden aufgefordert, keine Uhr oder andere Zeitmessgeräte zu benutzen und auch zu vermeiden, die Zeit leise zu zählen. Versuch 1 wird durch den Versuchsleiter durch die folgenden Worte eingeleitet: „Ich werde ‚Start‘ am Beginn und ‚Stopp‘ am Ende des Zeitintervalls sagen und ich möchte, dass ihr schriftlich festhaltet, was ihr meint, wie lange die Zeitspanne gedauert hat.“ (Es wird vorgeschlagen, die Zeitspanne mit 30 Sekunden anzusetzen.) Am Ende der Zeitspanne haben die Studierenden einige Sekunden Zeit, ihre Schätzung auf einem leeren Blatt zu notieren. Die Blätter zum ersten Versuch werden dann eingesammelt. Diese Daten werden nach Beendigung von Versuch 2 auf der Tafel notiert.

Der zweite Versuch beginnt wie der erste: Der Versuchsleiter nimmt dieselbe Zeitspanne (teilt das aber den Probanden *nicht* mit) und leitet die Aktivität mit denselben Worten ein („Ich werde ‚Start‘ ...“). Nach Ende der zweiten Versuchsspanne jedoch sagt der Versuchsleiter noch folgendes: „Einer in der Klasse hat große Erfahrung mit bereits verstrichener Zeit. Als Mitglied des Basketballteams (oder irgendeine andere plausible Erklärung) ist es für ihn wichtig, die verbleibende Zeit der Wurfuhr schätzen können. Lasst uns daher erst ‚Jürgen‘ um seine Schätzung der Zeitspanne fragen.“

[Beim Basketball wird die Angriffszeit auf 24 (35) Sekunden beschränkt, damit das Spiel schneller und attraktiver wird. Diese Zeit wird durch eine Wurfuhr angezeigt.]

Dann fragt der Versuchsleiter den studentischen „Experten“ (der natürlich vor dem Experiment eigens ausgesucht worden ist, damit er einen Schätzwert angibt, der deutlich unter oder über den 30 Sekunden liegt, die in beiden Versuchen zutrifft – wir empfehlen 45 Sekunden). Nachdem der „Experte“ seinen Wert angibt, schreiben die anderen ihre Schätzung auf ein leeres Blatt. Die Blätter werden eingesammelt und der Versuchsleiter schreibt jetzt die Daten zu den beiden Versuchen nebeneinander auf die Tafel.

Die Studierenden werden jetzt aufgefordert, über den sozialen Einfluss auf Bewertungen zu diskutieren. Speziell sollen sie hinterfragen, ob die Information des „Experten“ ihre Einschätzungen beim zweiten Versuch beeinflusst hat.

In kleinen Gruppen arbeiten nun die Studierenden, um Ideen auszutauschen, wie man die Unterschiede in den Schätzungen daraufhin prüfen kann, ob denn nun die Information des „Experten“ eine Auswirkung hatte oder nicht. Wir fordern die Studenten auf zu überlegen, welche deskriptive Kennziffern geeignet sind, die zentrale Tendenz [Lage, Niveau] und die Streuung der Schätzungen unter den zwei experimentellen Bedingungen zu beschreiben. Mittelwert und Median beschreiben das Zentrum der Verteilung, die Streuung wird dagegen durch Spannweite, Interquartilabstand oder Standardabweichung erfasst. Sofern die Verteilung annähernd normal ist, reichen Mittelwert und Standardabweichung aus (Gravetter & Wallnau 2011).

Daran anschließend regen wir an, dass die Studenten sich einen geeigneten statistischen Test überlegen, wie wir denn überprüfen können, ob die Daten einen signifikanten Unterschied in den Mittelwerten der geschätzten Zeitspannen anzeigen, die unter den beiden Bedingungen „mit“ bzw. „ohne die Information des Exper-

ten“ zustande gekommen sind. [Inzwischen wissen sie auch, dass die beiden Zeitspannen gleich waren.]

Die Studenten betrachten den Zweck und die Möglichkeiten verschiedener statistischer Tests für die Unterschiede zweier Gruppen hinsichtlich des Mittelwerts und entscheiden sich für zwei abhängige Gruppen [i. e., verbundene Stichproben]. Für alle Tests werden die Voraussetzungen besprochen; das betrifft auch das Skalenniveau der Messungen, die Zahl der Gruppen in der Untersuchung, die Forschungsstrategie und das Design.

[Unterschiede zwischen der Gruppe „Information“ und „keine Information“ sind durch das Design bereits durch die Unterschiede der einzelnen Personen abgebildet; je größer die individuellen Unterschiede, desto größer der Einfluss des Experten.]

Weil die Studie ein Design mit wiederholten Messungen verwendet, wird ein gepaarter t-Test als geeignet angesehen. Die Lerneinheit schließt damit ab, dass die Gruppen ihre Ansätze in der Klasse vorstellen, wo noch einmal darüber diskutiert wird.

[Natürlich kann man auch den Vorzeichentest anwenden. Dieser ist etwas gröber, aber durchaus geeignet zu testen, ob die Null-(Effekt-)Hypothese, dass kein Effekt des „Experten“ besteht, abgelehnt werden kann. In diesem Fall wertet man die Testgröße *Anzahl der positiven Vorzeichen der Differenzen* „Schätzung in Versuch 2 (mit Experten)“ und „Schätzung in Versuch 1 (ohne Experten)“ aus. Die Nullhypothese wird nun zu $p = 0,5$. Die Testgröße ist binomialverteilt.]

3 Forschungsaktivität 2: Zuschauer-Effekt

Vor dieser Aktivität besprechen wir kurz den Zuschauer-Effekt. Insbesondere wird den Studenten auch mitgeteilt, dass Darley & Latane (1968) gezeigt haben, dass die Wahrscheinlichkeit kleiner wird, dass ein Augenzeuge einem Opfer hilft, je mehr Leute den Notfall beobachten. Dies entsteht hauptsächlich durch eine *Verantwortungsdiffusion*, das ist eine Einstellung, dass andere ohnehin helfen werden, was zu einem geringeren Gefühl der Verantwortlichkeit und schließlich zu einer kleineren Wahrscheinlichkeit führt, helfend einzugreifen.

Zu Beginn dieser Arbeitseinheit bilden die Studenten, je nach Größe der Klasse, Gruppen zu 3 bis 5 Personen. Ein Mitglied übernimmt die Rolle des „Opfers“, das Hilfe bedarf, ein anderes registriert unauffällig die Zeit (in Sekunden), die verstreicht, bis Zuschauer Hilfe anbieten. Wir bieten den Studenten Örtlichkeiten am Campus an, wo sich genügend Passanten finden (Bibliothek, Aula, Schlafsäle etc.). Jede Gruppe bekommt

zufällig eine Zuschauer-Bedingung zugelost (1, 2 bzw. 5 oder mehr Zuschauer). Wenn die Klasse groß genug ist, können mehrere Gruppen für jede Bedingung zugeordnet werden, sodass wir für jeden Fall mehrere Beobachtungen erhalten. In kleineren Klassen kann die Situation von denselben Gruppen mehrfach (an verschiedenen Orten und Tagen) durchgespielt werden.

Ein Student (das Opfer) hat dann seine Bücher oder Unterlagen wie zufällig vor den Zuschauern (bzw. vor dem Zuschauer) fallen zu lassen, sodass sie über den Boden verstreut sind. Ein zweites Mitglied der Gruppe zeichnet auf, ob Zuschauer Hilfe anbieten und wie lange es dauert, bis Hilfe angeboten wird. Es ist ganz wichtig, dass der Beobachter dies unauffällig aus dem Hintergrund tut, sodass er nicht als weiterer Zuschauer wahrgenommen wird.

Diese Information wird dann auf ein Datenblatt übertragen (siehe Tabelle 1). Nachdem die Datensammlung komplettiert ist, werden alle Daten auf der Tafel notiert.

Zahl der Zuschauer	Helfen Zuschauer?	Zeit, die verstreicht, bis Hilfe angeboten wird
1	Ja oder nein	Wert (in Sekunden)
2		
5 oder mehr		

Tab. 1: Datenblatt einer Gruppe zum Zuschauer-Effekt

In kleinen Gruppen besprechen die Studierenden nun die üblichen Fragen:

- „Welche Kennziffern sind geeignet, um die Zeitspanne, bis Zuschauer unter den jeweiligen Bedingungen (Zahl der Zuschauer 1, 2 bzw. 5 oder mehr) Hilfe angeboten haben?“
- „Welche Kennziffern beschreiben das Verhalten der Zuschauer (Hilfe oder nicht)?“
- „Welche statistischen Tests sind geeignet, um festzustellen, ob es einen Zusammenhang gibt zwischen der Zahl der Zuschauer und dem Hilfeverhalten (ja oder nein)?“
- „Welche Tests sind geeignet, um zu bestimmen, ob die Daten ausreichen, um daraus zu schließen, dass es unter den verschiedenen Bedingungen (Zahl der Zuschauer) einen signifikanten Unterschied in der mittleren Zeit gibt, bis Hilfe angeboten wird?“

Die Studenten sollen bewusst überlegen, ob die Ziele, die mit den verschiedenen Kennziffern der zentralen Tendenz und der Streuung sowie mit den graphischen Darstellungen verbunden sind, diese Werkzeuge für die vorliegenden Fragestellungen als geeignet erscheinen lassen. Mittelwert und Standardabweichung sind jedenfalls geeignet, wenn das Merkmal auf einer

Intervall- oder Verhältnisskala gemessen wird (und die Daten annähernd einer Normalverteilung folgen).

Wie in der vorhergehenden Lerneinheit werden Mittelwert und Standardabweichung vorgezogen, wenn ein durchschnittlicher Wert in der Verteilung angezeigt und beurteilt werden soll, wie sehr die Werte von diesem Durchschnittswert abweichen. Beim Merkmal „Hilfe angeboten“ sehen die Studenten rasch, dass dies auf einer Nominalskala (ja oder nein) gemessen wird; es ist sogar dichotom [zweiwertig]. Daher ist in diesem Fall der Modus ein angemessenes Maß für das Zentrum und es eignet sich ein Stabdiagramm für die graphische Zusammenfassung der Häufigkeiten jeder Kategorie.

Bei der Suche nach geeigneten Methoden zur Beurteilung regen wir die Studenten dazu an, noch einmal den Zweck der verschiedenen Tests zu überdenken (es geht darum, eine Beziehung zwischen zwei gemessenen Merkmalen zu beurteilen); in gleicher Weise sollen die Anforderungen an die Skalen bei der Messung überprüft und die Zahl der Stichproben, die Forschungsstrategie sowie das Design der Studie bedacht werden.

In der vorliegenden Untersuchung wird die Zahl der Zuschauer auf einer ordinalen Skala gemessen, das Hilfeverhalten (ob Hilfe angeboten wurde oder nicht) auf einer nominalen Skala. Jedes Durchspielen der Situation ist ein Fall. Der Wert (die Anzahl) der Zuschauer ist durch Randomisierung bestimmt worden. Wir müssen festhalten, in wie vielen Fällen dann *Hilfe ja* und *Hilfe nein* aufgetreten ist. Etwa sei die Situation mit *einem* Zuschauer 20 Mal durchgespielt worden: In 5 Fällen sei Hilfe angeboten, in 15 sei keine Hilfe angeboten worden.

Wenn wir prüfen, ob ein Zusammenhang zwischen der Versuchsbedingung (Zahl der Zuschauer) und dem Hilfeverhalten (ja oder nein) besteht, müssen wir diese Daten in einer Kontingenztafel zusammenstellen; der Chi-Quadrat-Test auf Unabhängigkeit ist dann eine geeignete Methode, die Unabhängigkeit zu prüfen [im Falle der Ablehnung kann man einen Zusammenhang zwischen den Merkmalen Anzahl der Zuschauer und Hilfe annehmen].

Untersucht man zusätzlich, ob die mittlere Zeit, die vergeht, bis Hilfe angeboten wird, in den Gruppen der einzelnen Versuchsbedingungen verschieden ist, so ist die Varianzanalyse eine geeignete Methode. Die Zuordnung zu den Gruppen ist ja durch Zufall erfolgt. In diesen Gruppen ergeben sich entsprechende Messwerte für die Zeiten.

Hinweis: Die Varianzanalyse setzt voraus, dass es keinerlei fehlende Werte gibt (das sind hier Fälle, in denen keine Hilfe angeboten wurde). Der Untersu-

chungsleiter kann zusätzliche Hilfen anbieten, wie man bei fehlenden Werten in den Gruppen vorgeht.

Auch diese Einheit schließt damit, dass die Gruppen ihre Ideen und Ergebnisse in der Klasse vortragen.

[Anstelle der Varianzanalyse kann man auch mit Hilfe von Boxplots für die unterschiedlichen Versuchsbedingungen (1, 2, 5 oder mehr Zuschauer) die Zeit, bis Hilfe angeboten wird, untersuchen. Man wird zwar keine Signifikanzaussagen im Sinne der beurteilenden Statistik herausbekommen, aber schön illustrieren können, wie die Zeit zunimmt, wenn die Zahl der Zuschauer ansteigt. Solche Boxplots sind auch für die Praxis der Forschung sehr hilfreich, wenn man – nach dem erfolgten Nachweis eines signifikanten Unterschieds in den Versuchsgruppen durch die Varianzanalyse – der Frage nachgeht, welcher Art und Größe die signifikanten Unterschiede denn sind.]

4 Forschungsaktivität 3: Assoziationsketten und Gedächtnis

Einleitend erörtern wir experimentelle Studien zum Kurzzeitgedächtnis. Insbesondere gehen wir auf Zusammenhänge zwischen Assoziationsketten und erfolgreichem Abrufen aus dem Gedächtnis ein. Nach solchen Studien wird das Memorieren dadurch erleichtert, dass man ein Wort mit anderen Worten assoziiert, das einen Sinn hat.

Wir besprechen auch die Arbeit von Hermann Ebbinghaus (1885/1913) über sinnwidrige Ausdrücke. Solche Worte haben typischerweise einen Vokal zwischen zwei Konsonanten, sind also von der Bauart „KVK“. Diese Forschungsaktivität soll prüfen, wie der Versuch, die sinnwidrigen Worte mit Bedeutung zu verbinden, das Abrufen dieser aus dem Kurzzeitgedächtnis erleichtert und verbessert.

Dann teilen wir die Studenten willkürlich in zwei Gruppen ein: die Wiederholungs- und die Wortverkettungsgruppe. Wir beginnen das Experiment mit der Wiederholungsgruppe und bitten die andere Gruppe, den Raum vorübergehend zu verlassen.

Die Wiederholungsgruppe erhält eine Liste mit sinnwidrigen Worten (etwa wie in Tabelle 2) mit der folgenden Anleitung, die auf dasselbe Blatt obenauf geschrieben ist:

Du hast drei Minuten Zeit, um die unten stehenden sinnwidrigen Ausdrücke zu lernen. Um das Abrufen dieser Ausdrücke zu erleichtern, musst du diese Ausdrücke leise für dich selbst wiederholen; dazu hast du drei Minuten Zeit. Danach musst du diese Seite an den Versuchsleiter zurück geben und dann möglichst viele der sinnwidrigen Worte, wie dir eben einfallen, auf ein leeres Blatt schreiben.

1.	KOP	6.	REK
2.	ZOK	7.	CET
3.	VAP	8.	NUL
4.	MIP	9.	TOX
5.	SEK	10.	DAX

Tab. 2: Beispiele für sinnwidrige Ausdrücke

Nachdem die Anleitung auch laut vorgelesen worden ist, haben die Studenten drei Minuten Zeit, um die Ausdrücke auf der Liste zu lernen. Nach diesen drei Minuten erhalten sie eine weitere Minute, um so viele Ausdrücke wie möglich auf ein leeres Blatt zu schreiben. Die Zahl der von den jeweiligen Studenten korrekt aus dem Gedächtnis abgerufenen Worte wird auf der Tafel notiert.

Dann wird die Wortverkettungsgruppe in den Raum gebeten; die Teilnehmer erhalten eine Liste mit sinnwidrigen Ausdrücken mit der Anleitung, die nun wie folgt aussieht:

Du hast drei Minuten Zeit, um die unten stehenden sinnwidrigen Ausdrücke zu lernen. Um das Abrufen dieser Ausdrücke zu erleichtern, musst du jeden Ausdruck über eine Assoziationskette mit einem anderen Wort verbinden, das für dich einen Sinn hat. Zum Beispiel ‚SEF‘ kann mit dem Wort ‚SEE‘ assoziiert werden. In drei Minuten musst du diese Aufgabe erledigen. Danach musst du [...].

[Die weitere Vorgangsweise ist identisch mit der anderen Experimentalgruppe. Nachdem nun die Daten beider Gruppen auf der Tafel notiert sind, geht es an die Auswertung.]

Die Studenten arbeiten gemeinsam an der Analyse der Daten und machen sich wieder die grundsätzlichen Fragen bewusst, die da lauten:

- „Welche Kennziffern sind geeignet, um die Zahl der sinnwidrigen Worte, welche unter den beiden Versuchsbedingungen (Wiederholungs- lernen, Wortverkettung) richtig wiedergegeben wurden, summarisch zusammenzufassen?“
- „Gibt es einen signifikanten Unterschied in der mittleren Anzahl der richtig wiedergegebenen Worte in den beiden Experimentalgruppen?“

Auch hier erweisen sich für die erste Frage Mittelwert und Standardabweichung als geeignet, um das Zentrum der Daten und die Streuung in den beiden Datensätzen zu erfassen.

[Die Autoren sprechen von annähernd normalverteilten Daten, was nicht der Fall sein muss. Jedenfalls aber sollten die Forderungen an die Messskala (Intervall- bzw. Verhältnisskala) erfüllt sein), weil sonst die weitere Datenanalyse sinnlos wird.]

Für die Beantwortung der zweiten Frage beginnen die Studenten wieder über den Zweck der verschiedenen

Methoden der beurteilenden Statistik zu diskutieren. Hier geht es um die Differenz der Mittelwerte und die Frage, ob diese als von null verschieden betrachtet werden kann, die Frage nach den Voraussetzungen, die den jeweiligen Tests unterliegen und, ob die Messungen die Erfordernisse an die Skalen erfüllen.

Auch die Studie selbst und ihr Design werden besprochen. Hier wird das unabhängige Merkmal der Lernstrategie, welches die Gruppen erzeugt, auf einer Nominalskala, die Zahl der sinnwidrigen Worte, die richtig aus dem Gedächtnis abgerufen wurden, auf einer Verhältnisskala gemessen. Berücksichtigt man, dass die zwei Datensätze von zwei getrennten Gruppen von Einzelpersonen stammen und den Umstand, dass die Teilnehmer willkürlich einer der beiden Experimentalgruppen zugeordnet wurden, so erhält man ein „between-subjects“-Design. Daher ist der t-Test für unabhängige Stichproben geeignet, die Frage nach signifikanten Unterschieden im Mittelwert zwischen den beiden Experimentalgruppen zu beantworten.

Die Lerneinheit schließt wieder damit, dass die Gruppen ihre Ergebnisse und ihre Vorgangsweise in der ganzen Klasse vorstellen.

[Auch hier könnte man sich mit Boxplots der Daten unter beiden Bedingungen begnügen. Einflussgrößen nennt man auch Faktoren. Einflussgröße in Aktivität 3 ist die *Art des Lernens*. Die *Faktorstufen* entsprechen den Experimentalbedingungen.

- Ein „within subjects“-Design realisiert alle Faktorstufen (hier Lernbedingungen) für alle Individuen (verbundene Stichproben). Für jedes Individuum kann dann direkt die Differenz in der Zielvariablen (hier die Zahl der korrekt wiedergegebenen Worte) als „Behandlungseffekt“ interpretiert werden. Das war in Aktivität 1 der Fall. Jeder Proband hat unter beiden Bedingungen die Zeit, die verstrichen ist, geschätzt.
- Ein „between subjects“-Design separiert die Teilnehmer (unabhängige Stichproben); die einen erhalten nur die Behandlung der einen Faktorstufe, die anderen nur die der anderen Faktorstufe (im Beispiel haben wir nur zwei Gruppen zu vergleichen).

Als Vorteil des „between subjects“-Designs sieht man die Unabhängigkeit der Messungen an. Etwa könnten die Teilnehmer lernen und dadurch in der zweiten Runde bessere Ergebnisse erzielen. Als Nachteil ergibt sich, dass die Testgröße eine größere Varianz hat und damit die statistische Macht (Power) geringer wird, allfällige Unterschiede zwischen den Gruppen zu erkennen; d. h., dass man die Nullhypothese (dass

kein Unterschied zwischen den Gruppen besteht) schwerer ablehnen kann, falls tatsächlich doch ein solcher Unterschied besteht.]

Literatur

- Darley, J. M.; Latane, B. (1968): Bystander intervention in emergencies: Diffusion of responsibility. In: *Journal of Personality and Social Psychology* 8(4), S. 377–383.
- Ebbinghaus, H. (1885/1913): *Memory: A contribution to experimental psychology*. New York: Columbia University Press (Neudruck 1964. New York: Dover).
- Gravetter, F. J.; Wallnau, L. B. (2011): *Essentials of statistics for the behavioral sciences* (7. Aufl.). Belmont, CA: Wadsworth/Cengage Learning.
- Guilford, J. P.; Dallenbach, K. M. (1928): A study of the autokinetic sensation. In: *The American Journal of Psychology* 40(1), S. 83–91.
- Montgomery, R. L.; Enzie, R. F. (1971): Social influence and the estimation of time. In: *Psychonomic Science* 22(2), S. 77–78.
- Roseth, C. J.; Garfield, J. B.; Ben-Zvi, D. (2008): Collaboration in learning and teaching statistics. In: *Journal of Statistics Education* 16(1). www.amstat.org/publications/jse/v16n1/roseth.html (Zugriff: 29.11.2013).
- Saye, J. W.; Brush, T. A. (2004): Scaffolding problem-based teaching in a traditional social studies classroom. In: *Theory and Research in Social Education* 32(3), S. 349–378.
- Sherif, M. (1935): A study of some social factors in perception. In: *Archives of Psychology* 27(187), S. 1–60.
- Sherif, M. (1937): An experimental approach to the study of attitudes. In: *Sociometry* 1(1/2), S. 90–98.
- Thompson, W. B. (2000): Making data analysis realistic: Incorporating research into statistics courses. In: Ware, M.; Johnson, D. (Hrsg.): *Handbook of demonstrations and activities in the teaching of psychology, Vol. 1: Introductory statistics, research methods, and history* (2. Aufl.). New Jersey: Lawrence Erlbaum, S. 63–65.

Anschrift der Verfasser

Karen Y. Holmes
Norfolk State University
Norfolk, USA
kyholmes@nsu.edu

Brett A. Dodd
Regent University
Virginia Beach, USA
brettdod@regent.edu

Weitere Beiträge in *Stochastik in der Schule* zu psychologischer Forschung als Anlass für Unterricht:

König, G. (2011): Motivation bei Wettbewerben: Stochastische Aspekte einer Diskussion in der Zeitschrift „Psychological Science“. In: *Stochastik in der Schule* 31(3), S. 22–25.

Richardson, M.; Reischman, D. (2011): Die magische Zahl 7. In: *Stochastik in der Schule* 31(3), S. 26–29.