

**Einflussfaktoren auf Kompetenz, Sichtweisen und motivationale  
Variablen im Bereich von Statistical Literacy**

Dissertation

zur Erlangung des Grades eines Doktors  
der Philosophie (Dr. phil.)

der Pädagogischen Hochschule Ludwigsburg

vorgelegt von Ute Sproesser aus Stuttgart

Ludwigsburg

2015

Erstgutachter: Prof. Dr. Joachim Engel

Zweitgutachter: Prof. Dr. Sebastian Kuntze

Weitere Gutachter: Prof. Dr. Laura Martignon, Prof. Dr. Hans-Christoph Nürk

Datum des Abschlusses der mündlichen Prüfung: 11.12.2015

## **Danksagung**

An dieser Stelle möchte ich die Gelegenheit nutzen, mich bei einigen Personen zu bedanken, die mich in der Zeit meiner Promotion unterstützt und begleitet haben.

Mein besonderer Dank gilt Prof. Dr. Joachim Engel und Prof. Dr. Sebastian Kuntze, die mir diese Perspektive eröffneten und meine Promotion betreuten. Während meiner Promotionszeit standen sie mir mit ihrer Expertise und Erfahrung sowie mit Geduld und aufmunternden Worten zur Seite, wann immer es nötig war. Auch angesichts von Problemen und Rückschlägen haben sie mich stets unterstützt und so wesentlich zum Gelingen dieser Arbeit beigetragen.

Auf diesem Weg möchte ich auch den Mitgliedern des Kooperativen Promotionskollegs „Effektive Lehr-Lernarrangements“ danken. Aus diesem Kreis haben mich neben vielen anderen insbesondere Prof. Dr. Laura Martignon und Prof. Dr. Hans-Christoph Nürk in vielen Punkten beraten und stets in meinem Tun unterstützt. Dafür sowie für ihre Bereitschaft, meine Dissertation zu begutachten, möchte ich mich bedanken. Ebenso danke ich den anderen Doktoranden des Kollegs für viele hilfreiche, aufbauende Gespräche. Mein Dank gilt auch dem Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg sowie der PH Ludwigsburg, ohne deren finanzielle Unterstützung dieses Projekt nicht hätte realisiert werden können.

Ein besonderes Dankeschön geht an meine Kollegen des Instituts für Mathematik und Informatik der PH Ludwigsburg, insbesondere an Christoph Till, Anika Dreher, Marita Friesen, Claudia Wörn und Marc Zimmermann für ein immer offenes Ohr, eine positive Arbeitsatmosphäre, anregende Diskussionen und – falls alles andere nicht mehr half – einen großen Vorrat an Süßigkeiten. Für die Hilfe bei der Dateneingabe möchte ich mich bei den Hilfskräften Cornelia Dieterich und Georg Schlutter bedanken. Ein großes Danke gilt auch den Mitarbeitern der Forschungsförderungsstelle der PH Ludwigsburg, die mich bei vielen Anträgen beraten und somit zur Finanzierung der Projektes beigetragen haben.

Weiterhin danke ich allen Schülerinnen und Schülern sowie ihren Lehrerinnen und Lehrern, die bei dieser Studie beteiligt waren. Ohne sie wäre diese Arbeit nicht möglich gewesen.

Mein persönlicher Dank geht an meinen Mann Martin Dieterich, meine Eltern Barbara und Ludwig Sproesser sowie meine Geschwister Gudrun und Gunther, die immer an mich geglaubt und mich bedingungslos unterstützt haben. Ihr habt mir die notwendige Kraft zur Fertigstellung dieser Arbeit gegeben. Auch meinen Freunden und insbesondere meinen Mitspielerinnen der KSV Hoheneck Fußballmannschaft möchte ich für das entgegengebrachte Verständnis und die willkommene Abwechslung danken.

## Überblick über Publikationen und Tagungsbeiträge im Rahmen dieser Dissertation

Teilergebnisse dieser Arbeit wurden bereits in folgenden Beiträgen veröffentlicht bzw. präsentiert:

*Publikationen (nach Erscheinungsjahr geordnet)*

Sproesser, U., Engel, J. & Kuntze, S. (angenommen). Fostering self-concept and interest for statistics through statistics-specific learning environments. *Statistics Education Research Journal*.

Sproesser, U., Kuntze, S. & Engel, J. (eingereicht). Is statistical literacy interrelated to reading comprehension and cognitive abilities? A multilevel analysis. Manuskript eingereicht zur Publikation.

Sproesser, U., Engel, J. & Kuntze, S. (2015). Einflussfaktoren auf die Entwicklung von Statistical Literacy und von Sichtweisen auf zufallsbedingte Variabilität in der achten Realschulklasse. *Lernen und Lernstörungen* 4, 115-129.

Sproesser, U. (2014). Informelles statistisches Schließen anbahnen – Die Sicht von Achtklässlern auf Variabilität. In U. Sproesser, S. Wessolowski & C. Wörn (Hrsg.), *Daten, Zufall und der Rest der Welt. Didaktische Perspektiven zur anwendungsbezogenen Mathematik* (S. 235-246). Heidelberg: Springer.

Sproesser, U., Engel, J. & Kuntze, S. (2014). Der Einfluss einer statistikbezogenen Unterrichtseinheit auf Selbstkonzept und Motivation bei Achtklässlern. In J. Roth & J. Ames, *Beiträge zum Mathematikunterricht 2014* (S. 1159-1162). Münster: WTM.

Sproesser, U., Kuntze, S. & Engel, J. (2014). A multilevel perspective on factors influencing students' statistical literacy. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014)*, Flagstaff, Arizona, USA. Voorburg, The Netherlands: International Statistical Institute. Verfügbar unter [http://icots.info/9/proceedings/pdfs/ICOTS9\\_7E2\\_SPROESSER.pdf](http://icots.info/9/proceedings/pdfs/ICOTS9_7E2_SPROESSER.pdf) (Invited paper – refereed) [23.5.2015].

Sproesser, U., Kuntze, S. & Engel, J. (2014). Einflussfaktoren auf Statistical Literacy. *Die Perspektiven verbinden. Abstractband der 2. Jahrestagung der GEBF* (S. 180). Verfügbar unter <http://www.gebf2014.de/abstractband> [23.5.2015].

Sproesser, U., Kuntze, S. & Engel, J. (2013). Einflussfaktoren auf Statistical Literacy – erste Ergebnisse einer Studie mit Schülerinnen und Schülern der 8. Realschulklasse. In G.

Greefrath, F. Käpnick & M. Stein (Hrsg.), *Beiträge zum Mathematikunterricht 2013* (S. 962-965). Münster: WTM.

Sproesser, U., Kuntze, S. & Engel, J. (2013). General cognitive and reading abilities as impact factors on statistical literacy. In A.M. Lindmeier & A. Heinze (Eds.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 5 (p. 170). Kiel, Germany: PME.

Sproesser, U. & Kuntze, S. (2013). Statistical Thinking and Language – A Qualitative Analysis. In B. Ubuz, C. Haser, & M.A. Mariotti, *Proceedings of the Eighth Congress of the European Society for Research in Mathematics Education* (pp. 901-911). Antalya, Turkey: Middle East Technical University. Verfügbar unter <http://www.mathematik.uni-dortmund.de/~erme/index.php?slab=proceedings> [23.5.2015].

#### *Konferenzbeiträge (chronologisch geordnet)*

Sproesser, U., Kuntze, S. & Engel, J. (2014, Juli). A multilevel perspective on factors influencing students' statistical literacy. Vortrag auf der ICOTS9. Flagstaff, Arizona, USA.

Sproesser, U., Engel, J. & Kuntze, S. (2014, März). Der Einfluss einer statistikbezogenen Unterrichtseinheit auf Selbstkonzept und Motivation bei Achtklässlern. Vortrag auf der Jahrestagung der GDM. Koblenz, Germany.

Sproesser, U., Engel, J., Kuntze, S. & Nürk, H.-C. (2014, März). Einflussfaktoren auf Statistical Literacy. Posterpräsentation auf der 2. Jahrestagung der GEBF. Frankfurt, Germany.

Sproesser, U., Kuntze, S. & Engel, J. (2013, November). Reduktion und Variabilität als Förderansätze von Statistical Literacy: Erste Ergebnisse aus ReVa-Stat. Vortrag beim Arbeitskreis Stochastik der GDM. Paderborn, Germany.

Sproesser, U., Kuntze, S. & Engel, J. (2013, Juli). General cognitive and reading abilities as impact factors on statistical literacy. Vortrag auf der 37. PME-Tagung. Kiel, Germany.

Sproesser, U., Kuntze, S. & Engel, J. (2013, Juni). ReVa-Stat. Entwicklung von Vorstellungen zu datenbezogener Reduktion und statistischer Variabilität als Förderansatz zum Aufbau von Statistical Literacy. Vorstellung des Projekts und erste Ergebnisse. Vortrag im Rahmen des Doktorandenkolloquiums zwischen der PH Ludwigsburg und der PH Schwäbisch Gmünd. Ludwigsburg, Germany.

Sproesser, U., Kuntze, S. & Engel, J. (2013, März). Einflussfaktoren auf Statistical Literacy – erste Ergebnisse einer Studie mit Schülerinnen und Schülern der 8. Realschulklasse. Vortrag auf der Jahrestagung der GDM. Münster, Germany.

Sproesser, U. & Kuntze, S. (2013, Februar). Statistical Thinking and Language – A Qualitative Analysis. Vortrag auf der 8. CERME-Tagung. Antalya, Türkei.

## **Inhaltsverzeichnis**

Danksagung .....	i
Überblick über Publikationen und Tagungsbeiträge im Rahmen dieser Dissertation.....	ii
Inhaltsverzeichnis .....	1
Abbildungsverzeichnis .....	6
Tabellenverzeichnis.....	8
Zusammenfassung .....	9
1 Einleitung .....	11
2 Theoretischer Hintergrund .....	14
2.1 Statistical Reasoning, Statistical Thinking und Statistical Literacy .....	15
2.2 Die Sicht dieser Studie auf Statistical Literacy: Ziele und wesentliche Aspekte .....	19
2.3 Die Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ .....	26
2.4 (Bereichsspezifische) motivationale Variablen .....	33
2.4.1 Selbstkonzept .....	33
2.4.2 Interesse.....	34
2.5 Zur Beschreibung der Intervention dieser Studie .....	35
3 Das Projekt ReVa-Stat: Überblick und Forschungsinteresse der Teilstudien.....	37
3.1 Zusammenfassende Darstellung der Forschungsfragen der drei Teilstudien .....	39
4 Methode.....	40
4.1 Stichprobe und Design der Studie .....	40
4.2 Testinstrumente .....	43
4.2.1 Fachlicher Kompetenztest .....	44
4.2.2 Test zur Erfassung von Sichtweisen auf Variabilität .....	44
4.2.3 Test zur Erfassung bereichsspezifischer motivationaler Variablen .....	45
4.2.4 Lesegeschwindigkeits- und -verständnistest .....	46
4.2.5 Test zur Erfassung kognitiver Fähigkeiten.....	47
4.2.6 Test zur Erfassung des Arbeitsgedächtnisses .....	48
4.2.7 Erfassung weiterer Kovariaten .....	48
4.3 Übergreifende Bemerkung zu den statistischen Analysen .....	50
5 Teilstudie 1 .....	51

5.1 Abstract.....	51
5.2 Deutsche Zusammenfassung .....	52
5.3 Introduction .....	53
5.4 Theoretical Background .....	53
5.4.1 Statistical literacy .....	53
5.4.2 Statistical Literacy Competency Construct (SLCC) - a competency construct for using models and representations in statistical contexts .....	56
5.4.3 Theoretical considerations and empirical findings concerning variables potentially related to SLCC .....	61
5.5 Research questions .....	63
5.6 Methods .....	64
5.6.1 Sample and context of the study .....	64
5.6.2 Measures and covariates.....	64
5.6.3 Statistical Analyses .....	66
5.7 Results .....	68
5.7.1 Correlations between the consulted variables .....	68
5.7.2 The relationship between SLCC and reading comprehension as well as cognitive abilities .....	69
5.7.3 The relationship between SLCC and cognitive abilities controlled for grades in mathematics, gender and the socio-economic status.....	71
5.8 Discussion and conclusion.....	73
5.9 References .....	78
5.10 Additional analyses of the data.....	85
5.10.1 Data cleaning and checking of assumptions for linear models .....	85
5.10.2 Supplementary analyses of the relationship between the variables: CART.....	86
6 Teilstudie 2.....	88
6.1 Zusammenfassung deutsch .....	88
6.2 Ausführliche englische Zusammenfassung .....	89



6.2.1 Background .....	89
6.2.2 Aims .....	90
6.2.3 Methods .....	90
6.2.4 Results .....	91
6.2.5 Discussion .....	91
6.3 Einleitung.....	93
6.3.1 Statistical Literacy.....	93
6.3.2 „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ – ein Kompetenzmodell im Bereich von Statistical Literacy.....	95
6.3.3 Sichtweisen bezüglich zufallsbedingter Variabilität .....	96
6.3.4 Lernchancen durch die Auseinandersetzung mit zufallsbedingter Variabilität und datenbezogener Reduktion .....	97
6.3.5 Empirische Befunde zu möglichen Einflussfaktoren auf kognitive Variablen im Bereich von Statistical Literacy .....	97
6.4 Forschungsfragen.....	98
6.5 Methodik.....	99
6.5.1 Forschungsdesign und Stichprobe.....	99
6.5.2 Messinstrumente.....	100
6.6 Ergebnisse.....	102
6.6.1 Der Einfluss von individuellen Voraussetzungen auf die Entwicklung der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ .....	102
6.6.2 Der Einfluss der vier Treatments auf die Entwicklung der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ .....	103
6.6.3 Der Einfluss von individuellen Voraussetzungen und Treatmentvariablen auf Sichtweisen bezüglich zufallsbedingter Variabilität .....	104
6.7 Diskussion .....	105
6.8 Infobox Forschungsmethoden .....	109
6.8.1 Zuweisung zu den Treatmentgruppen .....	109
6.8.2 Korrelations- und Regressionsanalysen .....	109

6.9 Infobox Implikationen für die Praxis.....	110
6.10 Danksagung und Deklaration von Interessenkonflikten.....	111
6.11 Literatur .....	112
6.12 Elektronisches Supplement.....	116
7 Teilstudie 3.....	117
7.1 Abstract.....	117
7.2 Introduction .....	118
7.3 Theoretical background .....	118
7.3.1 Self-concept.....	119
7.3.2 Interest .....	120
7.3.3 Self-concept and interest: the case of mathematics and statistics .....	121
7.3.4 The statistics-related intervention of the current study .....	123
7.4 Research questions .....	124
7.5 Methods .....	125
7.5.1 Design and sample .....	125
7.5.2 Measures.....	126
7.5.3 Analyses .....	127
7.6 Results .....	129
7.6.1 Empirical distinction of mathematics- and statistics-related self-concept and interest .....	129
7.6.2 Fostering domain-specific self-concept and interest through a statistics-oriented intervention.....	131
7.6.3 The impact of prior statistics-related competency on the development of self-concept and interest in this domain .....	133
7.7 Discussion.....	134
7.8 References .....	138
8 Gesamtdiskussion.....	144
8.1 Zusammenfassung .....	144

8.2 Zusammenhänge zwischen der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ und Eingangsvoraussetzungen der Schüler .....	146
8.3 Zur Wirkung der Treatments auf das Kompetenzkonstrukt, Sichtweisen bezüglich zufallsbedingter Variabilität und motivationale Variablen .....	151
8.4 Weitere Forschung.....	154
8.5 Grenzen der Studie .....	157
8.5.1 Einschränkungen der Testinstrumente .....	157
8.5.2 Repräsentativität der Stichprobe .....	158
8.5.3 Beständigkeit der geschilderten Entwicklungen .....	158
8.6 Abschließende Bemerkungen .....	159
9 Darlegung des eigenen Anteils an dieser Studie .....	159
10 Literatur .....	161
11 Anhang .....	178
11.1 Lernumgebung „M&Ms“ des Treatments „Basis-Training“ .....	178
11.2 Lernumgebung „M&Ms“ des Treatments „Reduktion“ .....	182
11.3 Lernumgebung „M&Ms“ des Treatments „Variabilität“ .....	186
11.4 Lernumgebung „M&Ms“ des Treatments „Reduktion und Variabilität“ .....	190

## Abbildungsverzeichnis

Abbildung 2-1: Beispielaufgabe zur Illustration des Umgangs mit Variabilität .....	23
Abbildung 2-2: Beispielaufgabe zur Illustration wesentlicher Überlegungen bezüglich der Reduktion von Daten .....	25
Abbildung 2-3: Beispielaufgaben zu den drei Stufen des „datenbezogenen Lesens“ in Anlehnung an Curcio (1987).....	27
Abbildung 2-4: Elemente der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ (siehe Kuntze et al., 2008a, S. 114) .....	28
Abbildung 2-5: Beispielaufgabe auf Kompetenzniveau III (siehe z.B. Lindmeier, Kuntze & Reiss, 2007, S. 4) .....	29
Abbildung 2-6: Beispielaufgabe auf Niveaustufe IV (siehe z.B. auch siehe Kuntze et al., 2008a, S. 117) .....	3030
Abbildung 2-7: Beispielaufgabe auf Kompetenzniveau V (siehe z.B. Kuntze, 2013, S. 80) .	31
Abbildung 4-1: Aufgaben aus dem Basis-Training .....	42
Abbildung 4-2: Reflexionsaufgaben des Treatments „Variabilität“ .....	42
Abbildung 4-3: Reflexionsaufgaben des Treatments „Reduktion“ .....	43
Abbildung 4-4: Beispiel-Item zur Erfassung von Sichtweisen bezüglich zufallsbedingter Variabilität .....	45
Abbildung 4-5: Beispielitem zur Erfassung von aufgabenspezifischem Selbstkonzept und Interesse .....	46
Abbildung 4-6: Beispielitem des LGVT (Schneider et al., 2007, Testheft S. 3) .....	47
Abbildung 4-7: Beispielitems für verbale und nonverbale kognitive Fähigkeiten (Heller & Perleth, 2000; S. 114 bzw. S. 121) .....	47
Abbildung 4-8: Überblick der eingesetzten Instrumente, des zeitlichen Ablaufs und der drei Teilstudien des Projekts .....	49
Figure 5-1: Elements of the competency “using models and representations in statistical contexts” (cf. Kuntze et al., 2008b, p. 2) .....	57
Figure 5-2: Student’s answer to an item at Level III (see e.g. Lindmeier, Kuntze & Reiss, 2007, p. 4) .....	58
Figure 5-3: Student’s answer to an item at Level IV (see also Kuntze et al., 2008b, p. 4) .....	59

Figure 5-4: Student's answer to an item at Level V (Kuntze, 2013, p. 80) .....	60
Figure 5-5: Example of an LGVT item (cf. Schneider et al., 2007, translation by Leiss, Schukajlow, Blum, Messner & Pekrun, 2010, p. 127) .....	66
Figure 5-6: Scatterplot Matrix of Density Estimates, Linear Regression Lines and Scatterplot Smoothers with Confidence Bands .....	85
Figure 5-7: Regression tree including grades in mathematics, verbal and nonverbal cognitive abilities .....	87
Abbildung 6-1: Beispielaufgabe „Variabilität“ .....	94
Abbildung 6-2: Beispiel-Item zur Erhebung von Sichtweisen auf Variabilität .....	101
Abbildung 6-3: Entwicklung von Sichtweisen auf Variabilität nach Treatmentgruppen getrennt .....	105
Abbildung 6-4: Entwicklung von ESM 1. Word-Dokument. Zeitlicher Ablauf der Studie. .	116
Figure 7-1: Sample item for task-specific self-concept and interest .....	127
Figure 7-2: Confirmatory factor analysis for the pretest .....	131

## **Tabellenverzeichnis**

Tabelle 2-1: Kompetenzmodell zum „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ (siehe Kuntze, 2013, S. 7) .....	28
Table 5-1: Examples of items included in the cognitive abilities subscales used in this study (translated by the authors) .....	65
Table 5-2: Bivariate correlations between the consulted variables .....	68
Table 5-3: Model 1 including reading comprehension .....	69
Table 5-4: Model 2 including reading comprehension, verbal and nonverbal cognitive abilities .....	70
Table 5-5: Model 3 including verbal and nonverbal cognitive abilities .....	70
Table 5-6: Model 4 including verbal and nonverbal cognitive abilities as well as grades in mathematics .....	71
Table 5-7: Model 5 including cognitive abilities, grades in mathematics and gender .....	72
Table 5-8: Model 6 including cognitive abilities, grades in mathematics, gender and the socio-economic status .....	73
Tabelle 6-1: Kompetenzmodell zum Nutzen von Darstellungen und Modellen in statistischen Kontexten (Kuntze, 2013, S. 7) .....	96
Tabelle 6-2: Der Einfluss von individuellen Voraussetzungen und Treatmentvariablen auf die Entwicklung der Kompetenz .....	104
Table 7-1: Global fit indices for the tested models (pretest) .....	129
Table 7-2: Reliabilities and sample-items of mathematics- and statistics-related self-concept and interest .....	130
Table 7-3: Students' self-concept and interest related to mathematics and statistics in pre- and posttesting .....	133
Table 7-4: Multiple regression model predicting statistics-related self-concept (posttest) ..	134
Table 7-5: Multiple regression model predicting statistics-related interest (posttest) .....	134
Tabelle 8-1: Übersicht über Zielsetzungen und Ergebnisse der einzelnen Teilstudien .....	144

## **Zusammenfassung**

Für die informierte Teilhabe an der Gesellschaft sind grundlegende Fähigkeiten des Verstehens und Interpretierens von Daten unumgänglich. Insbesondere werden diese benötigt, um evidenzbasierte Entscheidungen im privaten und öffentlichen Leben treffen zu können. Ebenso spielen sie eine Rolle, um wichtige und komplexe soziale Themen wie Armut, Migration, Gesundheit, Kriminalität oder Bildung zu verstehen. Für die Entwicklung diesbezüglicher Fähigkeiten sind – wie für jegliche Lernprozesse – auch motivationale Komponenten wie Interesse und Selbstkonzept von großer Bedeutung. Trotz der Bedeutung datenbezogener Fähigkeiten im Alltag vieler Menschen gibt es kaum empirische Befunde dazu, welche allgemeinen Lernermerkmale<sup>1</sup> mit diesen zusammenhängen bzw. sie bedingen. Ebenso fehlt es an empirischer Evidenz zur Wirksamkeit von Lernarrangements, die auf die Förderung von kognitiven und motivationalen Variablen im Bereich Statistik abzielen. An diesem Punkt setzt die vorliegende Dissertation mit drei Teilstudien an, die jeweils unterschiedliche Zielsetzungen verfolgen. Das erste Ziel der Dissertation besteht in der Erforschung von Zusammenhängen zwischen statistikbezogener Kompetenz und bestimmten Lerner- bzw. Kontextmerkmalen. Das zweite Ziel ist es, die Wirkung einer statistikspezifischen Intervention auf diese Kompetenz sowie auf Sichtweisen bezüglich Variabilität zu untersuchen. Die dritte Zielsetzung der Dissertation liegt in der Analyse von Effekten dieser Intervention auf bereichsspezifische motivationale Variablen. Mit diesen drei Zielsetzungen fokussiert die Dissertation auf einen Erkenntnisgewinn, der sowohl für die Unterrichtspraxis als auch für die fachdidaktische Theoriebildung von Bedeutung ist. Eine Besonderheit dieser Arbeit stellt die Berücksichtigung der gruppierten Datenstruktur in allen drei Teilstudien durch die Verwendung geeigneter Analyseverfahren dar.

In einem ersten Schritt (Teilstudie 1) wurde in 25 achten Realschulklassen durch Mehrebenenregressionen die Beziehung zwischen der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ und Leseverständnis sowie allgemeinen kognitiven Fähigkeiten untersucht. Zusätzlich wurden individuelle und klassenbezogene Eingangsvoraussetzungen der Lernenden als Kontrollvariablen einbezogen. Das Kompetenzkonstrukt stellte sich als relativ eigenständig von den eher allgemeinen Schülervariablen Leseverständnis und kognitiven Fähigkeiten heraus. Dieses Ergebnis legt einerseits nahe, dass das in der vorliegenden Studie verwendete Testinstrument geeignet ist, Kompetenz im Bereich Statistik in-

---

<sup>1</sup> Im folgenden Text wird zugunsten der Lesbarkeit darauf verzichtet beide Geschlechter explizit zu benennen. Selbstverständlich sind darunter sowohl Mädchen als auch Jungen zu verstehen.

haltsspezifisch zu messen. Andererseits deutet dies darauf hin, dass die textgestützte Anlage der Lernmaterialien der Intervention für Schüler mit schwächerem Leseverständnis beziehungsweise mit niedrigeren kognitiven Fähigkeiten nicht grundsätzlich ein Hindernis darstellen muss. Außerdem ergaben sich signifikante aber moderate Zusammenhänge mit den zusätzlich einbezogenen Kovariaten Mathematiknote und Geschlecht auf Individualebene sowie mit dem sozioökonomischen Status auf Klassenebene.

Teilstudie 2 nahm in den Blick, inwiefern die individuellen Eingangsvoraussetzungen Leseverständnis, allgemeine kognitive Fähigkeiten, die Mathematiknote sowie das Geschlecht die Entwicklung der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ sowie die Sichtweisen der Lernenden auf zufallsbedingte Variabilität im Verlauf einer Intervention beeinflussten. Während vier Unterrichtsstunden bearbeiteten die Lernenden aus 25 achten Realschulklassen aufgeteilt in jeweils vier Treatmentgruppen speziell für diese Studie konzipierte Materialien. In ähnlicher Weise wie in Teilstudie 1 stellte sich das Leseverständnis trotz textgestützter Anlage der Lernmaterialien als wenig bedeutsam für die Kompetenzentwicklung heraus, so dass Lerner mit schwachem Leseverständnis in vergleichbarem Maße von der Intervention profitieren konnten wie ihre stärkeren Peers. Dagegen spielten die kognitiven Fähigkeiten sowie die Mathematiknote der Lernenden eine bedeutsamere Rolle für die Kompetenzentwicklung. Insgesamt verzeichneten die teilnehmenden Jungen einen höheren Kompetenzzuwachs als die Mädchen. Im Vergleich zu diesen individuellen Eingangsvoraussetzungen war die Zugehörigkeit zu den verschiedenen Treatments der Intervention relativ unbedeutend für die Entwicklung von statistikbezogener Kompetenz. Insbesondere konnte keine Verbesserung durch die Anregung von spezifischen Reflexionen im Vergleich zu einem eher algorithmischen Basis-Training festgestellt werden. Dagegen führte die Zugehörigkeit zu einem dieser reflexionsorientierten Treatments dazu, dass die Lernenden nach erfolgter Intervention zufallsbedingte Variabilität im Vergleich zu den Lernenden des Basis-Trainings signifikant stärker berücksichtigten als zuvor. Während die verschiedenen Treatments also kaum zu Unterschieden in der Kompetenzentwicklung führten, hatten sie bedeutsame Unterschiede in den Sichtweisen auf Variabilität zur Folge.

Neben der Entwicklung kognitiver Variablen stellt auch die Förderung motivationaler Variablen ein wichtiges Ziel schulischen Lehr-Lernhandelns dar. Durch Teilstudie 3 wurde der Fokus dieser Dissertation entsprechend erweitert und der Einfluss der Intervention auf das bereichsspezifische Selbstkonzept und Interesse analysiert. Dabei diente eine Baseline-Gruppe aus vier zusätzlichen Klassen als Referenzpunkt über den zeitlichen Verlauf der Intervention.



Die Schüler dieser Baseline-Gruppe absolvierten lediglich in analogen Zeitabständen die Tests, nahmen aber an Stelle der Intervention an ihrem regulären Mathematik-Unterricht teil. Zunächst wurde durch konfirmatorische Faktorenanalysen abgesichert, dass die verwendeten Skalen zur Erhebung von Selbstkonzept und Interesse bezogen auf Mathematik und Statistik empirisch trennbare Faktoren darstellten. Zudem legte es der Vergleich jeweils einer allgemeinen und einer aufgabenspezifischen Skala bezüglich des statistikbezogenen Selbstkonzepts und Interesses nahe, dass die Lernenden den Begriff Statistik mit dem Umgang mit Tabellen, Diagrammen und Kennwerten assoziierten. Im Verlauf der statistikbezogenen Intervention blieben die Durchschnittswerte der mathematikbezogenen Skalen recht stabil, während sich die Durchschnitte der statistikbezogenen Skalen in allen vier Treatmentgruppen im Vergleich zur Baseline-Gruppe signifikant erhöhten. Die verschiedenen Treatments zogen keine signifikanten Unterschiede nach sich. Es scheint so, dass die Auseinandersetzung mit statistischen Inhalten in einer schülerzentrierten Arbeitsform die Verstärkung des Selbstkonzepts und des Interesses bezogen auf den Bereich Statistik bewirkte – unabhängig von unterschiedlichen inhaltlichen Schwerpunkten der Treatments. Insbesondere Schüler, die bereits zum Zeitpunkt des Vortests über relativ hohe Kompetenz im Bereich Statistik verfügten, konnten durch die Intervention in ihrem statistikbezogenen Selbstkonzept und tendenziell auch in ihrem Interesse gefördert werden. Dieses Ergebnis unterstreicht die wechselseitige Beziehung zwischen Leistungs- und motivationalen Variablen und somit die Bedeutung der Förderung beider Bereiche durch geeignete Materialien.

Durch die Untersuchung von statistikbezogener Kompetenz, Sichtweisen auf Variabilität und motivationalen Variablen im Verlauf einer Intervention nimmt die vorliegende Dissertation eine breite Perspektive an Lernervariablen im Bereich Statistik in den Blick. Mit dieser Herangehensweise wird der Tatsache Rechnung getragen, dass schulischer Unterricht nicht nur auf den Aufbau von Kompetenz abzielen sollte, sondern auch die Persönlichkeitsentwicklung der Lernenden, die wiederum Lernprozesse determiniert, von wesentlicher Bedeutung ist. Die vorliegenden Ergebnisse liefern über eine Verbreiterung der fachdidaktischen Basis hinaus eine evidenzbasierte Einschätzung, inwiefern die untersuchten Variablen durch die Auseinandersetzung mit schülerzentrierten Lernmaterialien, wie sie im Rahmen dieser Studie entwickelt wurden, gefördert werden können.

## **1 Einleitung**

In unserer modernen Gesellschaft sind Daten in vielen Zusammenhängen allgegenwärtig (Gal, 2002; Ben-Zvi & Garfield, 2005). Täglich werden wir in den Medien mit Informationen bei-

spielsweise über die Wirtschaftskrise, den Zusammenhang zwischen CO<sub>2</sub>-Ausstoß und dem Abschmelzen der Pole oder die auflebende Vorliebe der Deutschen für regionale Produkte bombardiert. Häufig besingt die Werbung ihre Produkte nicht mehr, sondern sie berichtet von Studien, die deren Qualität wissenschaftlich nachweisen sollen. Und auch der berufliche Alltag großer Teile der Bevölkerung ist von Daten geprägt, wenn zum Beispiel statistische Darstellungen verwendet oder Trends ausgemacht werden müssen. Dabei können Daten nicht nur eingesetzt werden, um neutral zu informieren, sondern auch um bewusst zu polarisieren oder zu manipulieren. Insofern wird im privaten und beruflichen Umfeld von informierten Bürgern verlangt, Daten eigenständig zu verstehen, zu interpretieren und zu hinterfragen (Gal, 2002). Fähigkeiten des Umgangs mit Daten sind also im Sinne einer modernen Demokratie unerlässlich, um ohne auf fremde Interpretationen von Daten angewiesen zu sein an der Gesellschaft teilhaben zu können und Entscheidungen in einer von Ungewissheit geprägten Welt evidenzbasiert treffen zu können und somit handlungsfähig zu sein. Das Fehlen solcher grundlegenden Fähigkeiten identifizierte Wallman bereits 1993 als möglichen Auslöser von Missverständnissen, Täuschungen, Misstrauen und Befürchtungen in der Bevölkerung.

In Anbetracht der alltagsrelevanten Bedeutung diesbezüglicher Fähigkeiten ist es nicht verwunderlich, dass diese in den letzten Jahrzehnten in eine Vielzahl von Curricula aufgenommen wurden (Shaughnessy, 2007). Seit 1989 sind Inhalte im Bereich von Daten und Wahrscheinlichkeit Bestandteil der US-amerikanischen NCTM-Standards (National Council of Teachers of Mathematics, 1989), nur kurze Zeit später zogen vergleichbare Inhalte auch in die australische (Australian Education Council, 1991) und neuseeländische Curricula (Ministry of Education, 1992) ein. Die curriculare Aufnahme dieser Inhalte führte zu erheblich mehr Beachtung insbesondere des Lehrens von Statistik durch die (nationale) fachdidaktische Forschung (Shaughnessy, 2007), so dass die genannten Länder bis heute in diesem Bereich eine Art Vorreiterrolle einnehmen. In Deutschland dagegen wurde „Daten und Zufall“ flächendeckend erst 2003 als eine von fünf inhaltlichen Leitideen in die nationalen Bildungsstandards eingeführt (siehe KMK, 2003; vgl. auch Kaun, 2006). Beispielsweise in der Realschule in Baden-Württemberg erfolgte die Aufnahme dieses Inhaltsbereichs praktisch aus dem Nichts heraus (vgl. Land Baden-Württemberg, 2004; 1994). Wie in vielen Curricula anderer Länder orientieren sich die Inhalte von „Daten und Zufall“ an den fünf von Holmes (1980) genannten Komponenten statistischer Untersuchungen, so dass von der Grundschule bis zum Abitur im Sinne eines Spiralprinzips Daten erhoben, dargestellt, reduziert / zusammengefasst und interpretiert werden sowie das Thema Wahrscheinlichkeit behandelt wird. Auch wegen der verhältnismäßig späten flächendeckenden Aufnahme von statistischen Inhalten in die deutschen

Bildungsstandards besteht hierzulande gerade im internationalen Vergleich Nachholbedarf in der entsprechenden fachdidaktischen Forschung (Biehler & Hartung, 2006).

Trotz der inzwischen großen internationalen Beachtung von Fähigkeiten des Verstehens und Interpretierens von Daten sowie den damit verbundenen Bestrebungen, diese in schulischen Kontexten zu fördern, sind Probleme von Lehrenden und Lernenden in diesem Bereich gut dokumentiert. Beispielsweise berichtet Shaughnessy (2007) auf Grundlage der in den „National Assessments of Educational Progress“ erzielten Ergebnisse, dass Acht- bzw. Zwölfklässler unter anderem Schwierigkeiten beim Lesen bestimmter Diagrammformate oder bei der sinnvollen Auswahl und dem Verstehen von Kennwerten haben. Die Interpretation und Verwendung von in Diagrammen und Tabellen dargestellter Information fielen den Lernenden demzufolge noch schwerer. Auch Schield (1998) dokumentiert Probleme älterer Lernender sowohl im Bereich der deskriptiven als auch der Inferenz-Statistik. In einem Vergleich zwischen Studenten, Lehrern und Statistikern zeigten sich in allen Teilstichproben beachtliche Schwierigkeiten beim Beschreiben und Vergleichen von in Tabellen, Diagrammen und verbal gegebenen Daten (Schield, 2011). Garfield, delMas und Zieffler (2012) berichten aus einer Studie unter knapp 14.000 Studenten, dass sich die durchschnittlichen Leistungen in Statistikeinführungskursen von 2005 bis 2011 trotz der verstärkten Beachtung dieser Inhalte durch verschiedene Bildungseinrichtungen nicht verbessert haben und auf niedrigem Niveau verbleiben. Laut einer internationalen Studie von ICME (International Commission of Mathematics Instruction) und IASE (International Association for Statistical Education) fühlen sich auch viele Lehrer im Bereich der Statistik nur unzureichend ausgebildet (Batanero, Burril & Reading, 2011). Batanero und Kollegen schildern hier zahlreiche Schwierigkeiten und Fehlvorstellungen auf Seiten dieser Lehrpersonen, die wiederum einen entscheidenden Einfluss auf die Entwicklung von Schülerkompetenz haben. Da in manchen deutschen Bundesländern wie Baden-Württemberg bis vor einigen Jahren Statistik weder im Schulunterricht (z.B. Land Baden-Württemberg, 1994) noch in vielen Lehramtsstudiengängen (z.B. Land Baden-Württemberg 2003; 1999) als Pflichtthema behandelt werden musste, kann hierzulande mindestens von ähnlich großen Problemen ausgegangen werden.

Auch wenn durch die internationale Beachtung und curriculare Aufnahme von Inhalten, Wissenselementen und Fähigkeiten im Bereich der Statistik geeignete Rahmenbedingungen für die Vermittlung einer statistischen Grundbildung gegeben sind, zeigt die Vielzahl von Problemen Lernender, dass eine solche noch lange nicht flächendeckende Realität ist. Verschiedene Forscher weisen darauf hin, dass bestehende Konzeptionen von Lehrveranstaltungen kaum

geeignet sind, um eine solche allgemeine Grundbildung in Bezug auf den Umgang mit Daten umzusetzen (z.B. Garfield et al., 2012; Schield, 1998; Shaughnessy, 2007) beziehungsweise dass Lernende beim Verlassen der Schule noch weit von einer solchen entfernt sind (Watson, 2011). Das in dieser Dissertation dargestellte Promotionsprojekt setzt am Forschungsbedarf bezüglich der Ausrichtung von Förderkonzepten im Bereich des Umgangs mit Daten an. Dabei wird die Wirkung von speziell konzipierten Förderansätzen nicht nur auf Kompetenz im Bereich Statistik untersucht, sondern es werden auch Effekte auf weitere Zielvariablen in den Blick genommen: Einerseits erfolgte eine Ausweitung auf Sichtweisen auf zufallsbedingte Variabilität, welche als bedeutsam beim Aufbau von Fähigkeiten des Umgangs mit Daten angesehen werden (Engel & Sedlmeier, 2005; Wild & Pfannkuch, 1999). Andererseits wurden auch bereichsspezifische motivationale Variablen in den Blick genommen, deren Bedeutung im Lernprozess inzwischen sowohl im Allgemeinen (z.B. Pekrun & Zirngibl, 2004) als auch bezogen auf Statistik (z.B. Gal, 2002; Gal, Ginsburg & Schau, 1997) betont wird. Die hier gewonnenen Einblicke in die Wirksamkeit von Gestaltungsmerkmalen von Lernmaterialien im Bereich Statistik können zur Verbesserung von Unterrichtsqualität in Materialien für die Schule umgesetzt werden. Über diese Implikationen für die Schulpraxis hinaus liefert die Studie einen Beitrag zur fachdidaktischen Theorieentwicklung, indem empirische Erkenntnisse über Zusammenhänge zwischen Kompetenz im Bereich Statistik und bestimmten Lernvariablen sowohl vor als auch im Verlauf einer Intervention gewonnen wurden.

## **2 Theoretischer Hintergrund**

Die Beweggründe und Ziele dieses Forschungsprojekts wurden bereits in den letzten Abschnitten skizziert. In diesem Kapitel wird der zu Grunde liegende theoretische Rahmen vorgestellt. Im ersten Teilkapitel werden die drei miteinander vernetzten Konstrukte Statistical Reasoning, Statistical Thinking und Statistical Literacy unter Beachtung der Unterschiede und Überschneidungsbereiche vorgestellt. Auf Grundlage dieses theoretischen Überblicks werden im zweiten Teilkapitel die Ziele und wesentlichen Aspekte von Statistical Literacy aus Sicht dieser Studie dargelegt, die im verwendeten Kompetenzmodell ihren Niederschlag finden. Dieses Modell sowie das auf dessen Grundlage entwickelte Testinstrument werden im vierten Teilkapitel erläutert. Anschließend wird der theoretische Rahmen auf motivationale Variablen ausgeweitet. Den Abschluss dieses Kapitels bildet die Beschreibung der Intervention dieser Studie.

## 2.1 Statistical Reasoning, Statistical Thinking und Statistical Literacy

Wie die Vielzahl an theoretischen und empirischen Arbeiten der letzten 20 Jahre zeigt, hat sich das Interesse der fachdidaktischen Forschung am Bereich Statistik deutlich erhöht (Shaughnessy, 2007). Dabei haben sich in Abgrenzung zu der zuvor weit verbreiteten Betonung algorithmischer Fähigkeiten und formaler Wahrscheinlichkeitsrechnung beim Lehren von Statistik inzwischen die drei Konstrukte Statistical Reasoning, Statistical Thinking und Statistical Literacy etabliert, die stärker auf ein inhaltliches Verständnis fokussieren (Broers, 2006; Chance, 2002; delMas, 2002). Da alle drei Konstrukte Wissens- und Fähigkeits-Elemente im Bereich von Statistik beinhalten, ist eine genaue Grenzziehung schwierig. Teilweise werden die einzelnen Begriffe uneinheitlich verwendet, teilweise werden sie als gegeneinander austauschbar angesehen (Bidgood, 2014; Chance, 2002; delMas, 2002). Im Folgenden werden einige der in der Fach-Community verbreiteten Beschreibungen von Statistical Reasoning, Statistical Thinking und Statistical Literacy vorgestellt und zueinander in Beziehung gesetzt. Dies soll zunächst einen breit gefächerten Überblick über in der Fach-Community diskutierte Konstrukte, die Wissen und Fähigkeiten beim Umgang mit Daten zusammenfassen, geben. Auf dieser Grundlage werden anschließend die für die vorliegende Arbeit relevanten Aspekte von Statistical Literacy genauer beschrieben sowie das hier zu Grunde liegende Kompetenzmodell vorgestellt.

Ben-Zvi und Garfield (2005) beschreiben Statistical Reasoning als die Art und Weise, wie Menschen mit statistischen Konzepten argumentieren und wie sie aus statistischen Daten Sinn machen. Nach delMas (2002) geht es hierbei im Wesentlichen darum, wie und warum ein Lerner angesichts einer Aufgabe zu einem bestimmten Ergebnis gekommen ist. Auch laut Shaughnessy (2007) fasst das Konstrukt Statistical Reasoning zusammen, wie Menschen über bestimmte statistische Konzepte wie den Durchschnitt oder Variabilität denken beziehungsweise wie sie mit diesen argumentieren. Insofern stellen Modelle von Statistical Reasoning häufig Kategorien oder hierarchische Stufen des Verständnisses heraus, anhand derer verbale Schülerantworten auf offene Fragen eingeordnet und analysiert werden können. Dies hat den Vorteil, ein ganzes Spektrum an Wissen und Fähigkeiten abzubilden, anstatt Antworten nur als richtig oder falsch zu beurteilen (ibid.). Ein mögliches Beispiel, das auch im Bereich der Statistik häufig verwendet wird, stellt das „SOLO-Modell“ (Biggs & Collis, 1982) dar<sup>2</sup>: Aufbauend auf Arbeiten von Piaget (z.B. 1954; 1962) wird die Komplexität von Schülerantworten

---

<sup>2</sup> Für nähere Informationen bezüglich dieses und weiterer Modelle sei exemplarisch auf Jones, Langrall, Mooney und Thornton (2005) verwiesen.

auf offene Fragen auf einer fünfstufigen Skala (sensumotorisch, ikonisch, konkret-symbolisch, formal und postformal) eingeordnet. Auf jeder dieser Stufen kann die Komplexität der Äußerung noch konkreter beschrieben werden, indem beispielsweise angegeben wird, ob bei der Bearbeitung der Problemstellung nur ein Aspekt einbezogen wurde („unistructural“, z.B. der Mittelwert einer Verteilung), ob mehrere Aspekte getrennt voneinander betrachtet wurden („multistructural“, z.B. der Mittelwert und die Streuung) oder ob mehrere Aspekte miteinander verbunden wurden („relational“, z.B. Verbindung von Mittelwert und Streuung). Das Konstrukt Statistical Reasoning beschreibt also Arten des Denkens bzw. Argumentierens bei der Bearbeitung statistischer Aufgabenstellungen, ohne normativ bestimmte Anforderungen an Lernende zu nennen. Statistical Thinking fasst zusammen, was Menschen als Lerner, Konsumenten und Produzenten von Statistik aus Sicht von Statistik-Experten wissen und können sollten (Shaughnessy, 2007). Ein umfassendes und viel beachtetes Modell von Statistical Thinking legten Wild und Pfannkuch (1999) vor. Demnach kann Statistical Thinking von Lernenden durch die konkrete Durchführung von statistischen Untersuchungen vergleichbar mit der Arbeit professioneller Statistiker entwickelt werden. Das Kernstück dieses Modells bildet der „PPDAC-Cycle“, der den idealtypischen Verlauf einer statistischen Untersuchung von der Problemstellung bis zum Ziehen von Schlussfolgerungen darstellt. Dabei geht der Auf- und Ausbau von Statistical Thinking mit der Entwicklung von allgemeinen und statistikspezifischen Arten des Denkens wie dem Modellieren oder dem Beachten von Variabilität einher. Zusätzlich wird die Bedeutung metakognitiver Strategien bei der Produktion und Analyse von Daten betont, da sowohl die Daten als auch das eigene Vorgehen ständig hinterfragt und reflektiert werden müssen. Um Statistical Thinking in diesem Sinn zu entwickeln, benötigen Lernende wie in anderen Bereichen des Problemlösens grundlegende Dispositionen wie einen gesunden Skeptizismus, Neugierde und Durchhaltevermögen.

Auch Ben-Zvi und Garfield (2005) verbinden Statistical Thinking mit Fähigkeiten, die zur Durchführung komplexer statistischer Untersuchungen erforderlich sind. Die Autoren nennen hier beispielsweise das Verständnis für Stichprobenziehen sowie für damit verbundene Schlüsse von einer Stichprobe auf eine Population. Solche komplexen Wissens- und Fähigkeitselemente gehen über das hinaus, was der Durchschnittsbürger in seinem Alltag benötigt und was somit als statistische Grundbildung in der Schule vermittelt werden sollte. Grundlegende Kompetenzen im Bereich Statistik werden häufig mit dem Begriff Statistical Literacy in Verbindung gebracht. Dieses Konstrukt wird im Folgenden näher beschrieben.

Modelle von Statistical Literacy identifizieren laut Shaughnessy (2007) das Mindestmaß an Wissen und Fähigkeiten im Bereich Statistik, das Schüler sowie Erwachsene brauchen, um als informierte Bürger datenbasierte Entscheidungen treffen zu können. Dies umfasst auch eine kritische Grundhaltung gegenüber Verkäufern, Wissenschaftlern oder Politikern, so dass jeder Mensch – sofern er davon betroffen ist – in der Lage sein sollte, auf Daten basierende Fehler, Halbwahrheiten oder Lügen als solche zu identifizieren. Während einige Fachdidaktiker Statistical Literacy lediglich als Verstehen grundlegender statistischer Begriffe betrachten (Garfield & delMas, 2010; Garfield et al., 2005), beziehen andere auch anspruchsvollere Fähigkeiten wie das Interpretieren und kritische Hinterfragen von statistischer Information mit ein (Gal, 2002; Wallman, 1993). Andere Ansätze wiederum zählen für bestimmte Zielgruppen eine Reihe von konkreten, teilweise auch sehr komplexen Fähigkeiten wie ein Verständnis für Umfragen oder für das Verhalten zufälliger Stichproben auf, die Statistical Literacy einschließen soll (Franklin et al., 2007).

Gal (2002) entwarf ein Modell, das wesentliche, vielfach miteinander verwobene Komponenten von Statistical Literacy zusammenfasst. Demnach benötigen Menschen als Konsumenten von Statistiken sogenannte „Literacy Skills“, also allgemeine Kenntnisse und Fähigkeiten des Entnehmens und Verarbeitens von Information aus schriftlichen Darstellungen wie Listen, Tabellen oder Grafiken (S. 7ff; siehe auch die Auffächerung von Statistical Literacy in „Document Literacy“, „Prose Literacy“ und „Quantitative Literacy“, beispielsweise bei Shaughnessy, 2007, S. 964). Daneben ist spezifisches Wissen im Bereich der Statistik (S. 9ff; z.B. grundlegende Begriffe bezüglich deskriptiver Statistiken oder Darstellungen von Daten), der Mathematik (S. 13ff; z.B. mathematische Prozeduren zur Bestimmung von Kennwerten wie dem Mittelwert oder dem Median) und des jeweiligen Datenkontexts (S. 15; z.B. ob alle relevanten Informationen eines Studiendesigns dokumentiert sind, ob statistische Schlüsse auf Grundlage dieses Designs zulässig sind, ...) von Bedeutung. Durch das separate Aufführen von sogenannten kritischen Fähigkeiten und einer kritischen Haltung betont Gal, dass über eine grundsätzlich kritische Einstellung gegenüber statistischen Informationen und Schlussfolgerungen hinaus spezielle Kenntnisse beispielsweise über mögliche Verzerrungen durch das Verwenden bestimmter Kennwerte oder Darstellungen erforderlich sind (S. 15ff). Um diese Elemente mit dem Ziel eines adäquaten Umgangs mit Daten zu verbinden, sind motivationale Variablen wie die positive Einschätzung der eigenen Fähigkeiten sowie das Interesse und die Wertschätzung gegenüber statistischer Verfahren notwendig (S. 17ff), die somit das letzte Element von Gal's Modell bilden.

Die Komplexität des oben skizzierten Modells macht deutlich, dass Lernende Statistical Literacy im Sinne von Gal nur schrittweise und über einen längeren Zeitraum entwickeln. Watson und Callingham (1997) konnten durch Rasch-Analysen zeigen, dass Statistical Literacy trotz großer Vielschichtigkeit ein eindimensionales Konstrukt mit sechs hierarchischen Niveaustufen ist. Anhand eines in diesem Zusammenhang entwickelten Modells können (Schüler-)Fähigkeiten von einer sehr basalen Ausprägung bis hin zu Expertenniveau klassifiziert werden. In Einklang mit einer früheren Studie von Watson (1997) sowie der Arbeit von Gal (2002) spielen für die Zuordnung zu einer Niveaustufe über den Umgang mit grundlegenden statistischen Begriffen und Techniken hinaus der Einbezug des Kontexts sowie der Grad an kritischem Hinterfragen von statistischen Informationen und Schlussfolgerungen eine wesentliche Rolle. Die von Watson und Callingham (ibid.) eingesetzten Aufgaben fokussierten neben grundlegenden Statistik-Inhalten des australischen Curriculums insbesondere auf das Verständnis von Variabilität, dessen Ausprägung mitentscheidend für die Einstufung von Lernertfähigkeiten ist.

Diese Ausführungen machen deutlich, dass Statistical Literacy im Sinne von Wallman (1993), Gal (2002) oder Watson und Callingham (2003) über das hinausgeht, was Shaughnessy (2007) oder Garfield und Kollegen (Garfield et al., 2005; Garfield & delMas, 2010) darunter verstehen. An den genannten Beispielen ist zu erkennen, dass es in der Fachcommunity keinen allgemeinen Konsens über das Konstrukt Statistical Literacy und insbesondere damit verbundene Inhalte bzw. Anforderungen gibt.

Zudem zeigen die sich teilweise überschneidenden Beschreibungen insbesondere von Statistical Thinking und Statistical Literacy auf, dass eine klare Grenzziehung zwischen den Konstrukten kaum möglich ist. Beispielsweise das kritische Hinterfragen von statistischer Information sowie die Beachtung des Kontextes werden als kennzeichnend für Statistical Thinking als auch für Statistical Literacy angesehen (z.B. Gal, 2002; Wild & Pfannkuch, 1999). Ebenso scheint das Verständnis für Variabilität, also für Unterschiede in Daten und deren Ursprung, in einem Überschneidungsbereich zu liegen (Watson, 1997; Watson & Callingham, 2003; Wild & Pfannkuch, 1999). Sowohl Statistical Thinking als auch Statistical Literacy wird als bedeutsam für das Treffen von datenbasierten Entscheidungen angesehen (Shaughnessy, 2007; Wallman, 1993; Wild & Pfannkuch, 1999), wenn auch in Bezug auf unterschiedliche Zielgruppen (Grundbildung für alle Menschen versus Kompetenzen hauptsächlich professioneller Statistiker).



Die Uneinheitlichkeit an Beschreibungen von Statistical Reasoning, Thinking und Literacy wird in der Fach-Community durchaus diskutiert und zeigt sich auch darin, wie die Konstrukte zueinander in Beziehung gesetzt werden. DelMas (2002) nennt drei Perspektiven, wie man das Verhältnis zwischen den drei Konstrukten darstellen kann. So liegt vielen Arbeiten ein hierarchisches Modell zu Grunde, das Statistical Literacy als Start- und Statistical Thinking als Endpunkt eines Wissens- und Fähigkeitsspektrums mit gewissen Überschneidungsbereichen auffasst. Andererseits kann Statistical Literacy auch als Oberbegriff betrachtet werden, der Statistical Reasoning und Thinking miteinschließt. Zuletzt kann man die Unterscheidung der drei Konstrukte darin sehen, auf welche statistikbezogenen Aktivitäten mit einer Problemstellung fokussiert wird. So kann es beispielsweise im Zusammenhang mit dem komplexen Inhalt „Konfidenzintervall“ darum gehen, solche zu interpretieren (Statistical Literacy), die diesbezüglichen Denkweisen von Lernern offen zu legen (Statistical Reasoning) oder diese kritisch zu überprüfen (Statistical Thinking) (ibid., S. 6ff).

Die in den letzten Abschnitten aufgezeigte fehlende Einheitlichkeit an Konstruktbeschreibungen und –abgrenzungen macht deutlich, warum einige Wissenschaftler für theoretische und empirische Arbeiten in diesem Bereich empfehlen, zunächst eine klare Arbeitsdefinition des zu untersuchenden Konstrukts vorzulegen (z.B. Ben-Zvi & Garfield, 2005; Shaughnessy, 2007). Entsprechend dieser Empfehlung wird im folgenden Abschnitt erläutert, welches Verständnis von Statistical Literacy dieser Studie zu Grunde liegt.

## **2.2 Die Sicht dieser Studie auf Statistical Literacy: Ziele und wesentliche Aspekte**

Statistical Literacy wird in dieser Studie als eine bereichsspezifische Kompetenz im Sinne Weinerts (2001) aufgefasst, die neben Wissen und Fähigkeiten in einem bestimmten Bereich auch individuelle Dispositionen wie die Bereitschaft und Motivation sich mit verschiedenen Problemstellungen auseinander zu setzen, impliziert. Vergleichbar mit Mathematical Literacy (Klieme, Neubrand & Lüdtke, 2001; OECD, 2003) zielt Statistical Literacy also nicht auf die Beherrschung isolierter Regeln, Verfahren oder Fertigkeiten ab (vgl. auch Ben-Zvi & Garfield, 2005), sondern darauf, handlungsfähig zu sein und Probleme in vielfältigen Alltagskontexten lösen zu können (Gal, 2002; Wallman, 1993). Im Fall von Statistical Literacy geht es demnach um das Lösen von alltagsbezogenen Problemen in statistischen Kontexten.

Das Lösen von Problemen ist im Wesentlichen dadurch charakterisiert, dass eine Barriere zwischen Ausgangs- und erwünschtem Zielzustand überwunden werden muss, wobei zu Beginn nicht bekannt ist, wie dies erreicht werden kann (Frensch & Funke, 1995, S. 8). Das Anwenden von Routinen und Algorithmen reicht nicht aus, um ein Problem zu lösen. Vielmehr

muss die Komplexität des Problems schrittweise reduziert werden, bis eine Lösung erreicht und vor dem Hintergrund der ursprünglichen Problemstellung überprüft werden kann (Klieme, Funke, Leutner, Reimann & Wirth, 2001).

Bezogen auf das Lösen realitätsbezogener Probleme mit mathematischen Mitteln spielt das (mathematische) Modellieren eine wesentliche Rolle (z.B. Blum et al, 2004; Klieme, Neubrand et al., 2001): Einerseits muss das in einer konkreten Sachsituation gegebene Problem so weit vereinfacht werden, dass es mit mathematischen Mitteln bearbeitet werden kann. Man spricht von der Bildung eines mathematischen Modells, was gegebenenfalls schrittweise über ein Situations- und ein Realmodell erfolgt (z.B. Blum & Leiss, 2005). Andererseits ist es notwendig, die ermittelte mathematische Lösung wieder auf die ursprüngliche Sachsituation zurück zu beziehen und zu reflektieren, inwiefern die mathematische Lösung im Sachkontext Sinn macht und hilft, das ursprüngliche Problem besser zu verstehen. Dieses Vorgehen validiert das Modell und die mathematische Lösung vor dem Hintergrund der Sachsituation (ibid.). Das Wechseln zwischen konkreter Sachsituation und mathematischem Modell, das beim Lösen realitätsbezogener Probleme vollzogen werden muss, stellt ein zentrales Element des (mathematischen) Modellierens dar (ibid.). In vergleichbarer Weise müssen Lernende beim Lösen von Problemen in statistischen Kontexten zwischen der Sachsituation und einem statistischen Modell (umfasst im Vergleich zum mathematischen Modell z.B. statistikspezifische Darstellungen) hin- und herwechseln, also modellieren (Engel & Kuntze, 2011; Garfield et al., 2012; Lehrer, 2014).

Wie beim mathematischen Modellieren erfordern auch Modellierungsaktivitäten in statistischen Kontexten, dass vielfältige Darstellungen konkreter Sachsituationen verstanden, erstellt, verändert und hinterfragt werden können. Darstellungen werden hier als semiotische Repräsentationen von beispielsweise mathematischen Objekten verstanden, über die diese abstrakten Objekte erst zugänglich und bearbeitbar werden (Duval, 2006). Im Falle der Statistik handelt es sich z.B. um Darstellungen von Daten wie Diagramme oder Tabellen, die unter anderem von den Medien oder im beruflichen und privaten Rahmen benutzt werden. Da durch einzelne statistische Darstellungen immer nur bestimmte Aspekte einer Sachsituation vermittelt werden können, führen multiple Perspektiven und der Wechseln zwischen verschiedenen Darstellungsformen der Daten oft zu tieferen Einsichten. Wild und Pfannkuch (1999) haben für diesen Wechsel der Darstellungsformen den Begriff Transnumeration geprägt. Ein solcher Darstellungswechsel erfordert Modellierungsaktivitäten, wenn zwischen statistischem Modell und dessen Bedeutung in der realen Welt übersetzt wird. Ebenso muss modelliert werden,

wenn ein Lerner eine im Sinne einer Problemstellung angemessene statistische Darstellung auswählt und erstellt, denn dazu muss er sich einerseits darüber bewusst sein, welche konkrete Information er vermitteln will (Vereinfachung der ursprünglichen Situation durch Identifikation der relevanten Daten) und andererseits die Eigenschaften verschiedener Darstellungen kennen, um sich sinnvoll für eine zu entscheiden und diese zu gestalten (ggf. als Teil der Bildung eines geeigneten statistischen Modells). Auch nach der Erstellung einer Darstellung muss überprüft werden, ob die ursprüngliche Absicht durch das erzielte Ergebnis erreicht wurde oder noch einmal überdacht werden muss (Teilschritt der Validierung des Modells). Insofern ist beim Lösen von Problemen in statistischen Kontexten das Modellieren kaum vom Nutzen statistischer Darstellungen zu trennen.

Modellierungsaktivitäten sind insbesondere erforderlich, wenn mit statistischer Variabilität umgegangen wird. Variabilität wird von einer Vielzahl von Experten als das wesentliche Merkmal statistischer Daten hervorgehoben (z.B. Cobb & Moore, 1997; Moore, 1997; Watson, 2011; Watson & Callingham, 2003; Wild & Pfannkuch, 1999). Shaughnessy (2007) definiert Variabilität als „propensity of something to change“ (S. 972), also als grundlegende Eigenschaft statistisch erhobener Daten, Stichproben und Verteilungen, sich voneinander zu unterscheiden. Variabilität ist – zumindest bei genauer Betrachtung – überall in der Realität anzutreffen (Wild & Pfannkuch, *ibid.*): Nicht nur Menschen unterscheiden sich hinsichtlich ihrer physischen Besonderheiten, ihrer kognitiven Voraussetzungen oder ihrer finanziellen Situation, sondern selbst nach genauen Vorgaben und Maßen maschinell gefertigte Objekte wie Schuhe oder Nägel sind – wenn man genau hinschaut – nicht identisch. In ähnlicher Weise erhält man bei mehrfacher Messung eines Merkmals unter gleichen Bedingungen kaum dasselbe Ergebnis (Moore, 1990). Diese Unterschiede können laut Wild und Pfannkuch (*ibid.*) systematische Ursachen haben, also für einen Sachverhalt charakteristisch sein (z.B. in der Regel steigende Durchschnittsgrößen von Schülern der Klassen 5 bis 10) oder dem Zufall zugeschrieben werden (z.B. nicht erklärbare interindividuelle Unterschiede oder Abweichungen z.B. als Folge von Messfehlern oder weiterer nicht fassbarer Einflüsse). Aufgrund von Variabilität wurden teils anspruchsvolle statistische Methoden entwickelt, um relevante Informationen (z.B. systematische Unterschiede) aus durch Variabilität geprägten Daten zu ermitteln (Cobb & Moore, 1997; Wild & Pfannkuch, *ibid.*). Ob Unterschiede als zufällig attribuiert werden oder nicht hängt insofern auch davon ab, welche statistischen Methoden und welches Kontextwissen verfügbar sind. Die Beurteilung solcher Unterschiede als zufällig oder nicht hat meist praktische Folgen: Beispielsweise bei der Identifizierung von Trends können

dem Zufall attribuierte Unterschiede vernachlässigt werden, während Unterschiede als Folge systematischer Ursachen berücksichtigt werden sollten (Wild & Pfannkuch, *ibid.*).

Um die Herkunft von Unterschieden im Sinne einer Problemstellung angemessen bewerten zu können, wird in der Regel nach Strukturen in Daten bzw. nach Zusammenhängen zwischen den betrachteten Variablen gesucht. Will man beispielsweise einen Zusammenhang zwischen Alter und Körpergröße von Schülern herausfinden, so kann man beide Variablen in einem Streudiagramm gegeneinander abtragen. Bei dieser Problemstellung ist es naheliegend Unterschiede innerhalb der einzelnen Altersstufen zu vernachlässigen und sich auf die systematischen Unterschiede zwischen den Altersstufen zu konzentrieren. Um den systematischen Zusammenhang höheres Alter – größere Körpergröße zu erkennen, muss zwischen einem statistischen Modell und dessen Bedeutung in der konkreten Sachsituation gewechselt, also modelliert werden. Solche Modellierungsaktivitäten sind häufig erforderlich, wenn mit Variabilität umgegangen wird. Durch die Modellierung von Variabilität kann oft nicht nur die Art (systematisch oder zufällig) der vorliegenden Unterschiede geklärt, sondern es können auch zukünftige Entwicklungen vorhergesagt oder zu Grunde liegende Prozesse gesteuert werden (Wild & Pfannkuch, *ibid.*).

In der Beispielaufgabe in Abbildung 2-1 (rechte Seite) erkennt man, dass die Verbrauchswerte von Herrn Arnold in den Monaten Juni bis Februar zwar nicht identisch, aber doch relativ ähnlich sind (attribuiert als zufällige Variabilität). Somit müssen diese Unterschiede für die Ermittlung eines Pauschalpreises, beispielsweise durch Bilden des Durchschnitts, nicht speziell beachtet werden. Dagegen weicht der Verbrauch im Juli deutlich von den anderen Monaten ab (dieser Unterschied wird als systematisch angesehen), so dass dieser Wert bei der Bildung eines Festpreises gesondert behandelt werden sollte. Bei dieser einfachen Aufgabe sind sowohl der notwendige Wechsel zwischen Sachsituation („Ist es gerecht, einen zukünftigen Festpreis für jeden Mitarbeiter auf Grundlage der Daten des Monats Juli zu ermitteln?“) und dem statistischem Modell (z.B. grafische oder numerische Darstellung zur Beurteilung von Unterschieden), als auch die Möglichkeit der Vorhersage der zukünftigen Verbrauchsentwicklung durch die Modellierung von Variabilität erkennbar. Zudem wird klar, dass der Umgang mit Variabilität nicht immer komplexe statistische Verfahren wie die Berechnung von Streumaßen erfordert, sondern auch durch sorgfältiges Beobachten der Werte sowie durch kritisches Hinterfragen, was die Abweichungen im vorliegenden Kontext bedeuten, erfolgen kann.

Abbildung 2-1

Beispielaufgabe zur Illustration des Umgangs mit Variabilität

Name	Kaffee-Liste Juli	Monat	Kaffee-Verbrauch Arnold
Arnold	<del>///</del> <del>///</del> <del>///</del> <del>///</del> IIII	Juli	<del>///</del> <del>///</del> <del>///</del> <del>///</del> IIII
Baier	II	Juni	<del>///</del> III
Christian	<del>///</del> <del>///</del> III	Mai	<del>///</del> I
Domold	<del>///</del> I	April	<del>///</del> IIII
Eisner		März	<del>///</del> II
Flick	III	Februar	<del>///</del> III

Ausgehend von der „Kaffee-Liste Juli“ soll für jeden Mitarbeiter ein monatlicher Pauschalpreis ermittelt werden. Herr Arnold findet das nicht gerecht.

Da der kritische und hinterfragende Umgang mit Variabilität in vielen Bereichen des statistischen Arbeitens wie dem Erheben, Darstellen, Reduzieren und Interpretieren von Daten eine Rolle spielt (Watson, 2011), ist es leicht nachvollziehbar, dass viele Forscher ihm eine große Bedeutung zumessen. Diese hohe Relevanz zeigt sich auch in einem von Watson und Callingham (2003) entwickelten Kompetenzmodell bezüglich Statistical Literacy, das das Verständnis für Variabilität als zentrales Element zur Klassifizierung von Lernerfähigkeiten herausstellt.

Leider kann nicht davon ausgegangen werden, dass Lernende sich über das Phänomen der Variabilität bewusst sind (Shaughnessy, 2007; Watson, 2011). Wild und Pfannkuch (1999) stellen in diesem Zusammenhang fest, dass Menschen dazu neigen, selbst dann nach kausalen Ursachen für Unterschiede zu suchen, wenn die vorliegenden Daten unter Berücksichtigung von Zufallsvariabilität noch gut innerhalb der Grenzen des zu Erwartenden liegen. Solche Sichtweisen auf Variabilität berücksichtigen kaum zufällige Unterschiede in den Daten. Das gilt insbesondere für Schüler, deren Hang zu einem mechanistisch-deterministischen Weltbild durch verschiedene Studien (Engel & Sedlmeier, 2005; Green, 1982, 1986, 1990; Piaget & Inhelder, 1975) gut dokumentiert ist. In diesen Studien zeichneten die Lernenden bei der Vorhersage der Verteilung von Regentropfen oft regelmäßige und somit „variabilitätslose“ Muster. Engel und Sedlmeier (ibid.) stellten fest, dass diese Vorliebe für regelmäßige Strukturen auch mit zunehmendem Alter nicht zu verschwinden scheint. Gründe dafür sehen Engel und Sedlmeier (ibid.) wie auch Green (1982) und Fischbein (1975) im Einfluss der Schule, die häufig ein deterministisches Denken vermittele und kaum Raum für Erfahrungen mit zufallsbedingter Variabilität lasse. Aufgrund dieser deterministischen Denkweisen rechnen Engel

und Sedlmeier (2005) mit Schwierigkeiten bei der Entwicklung angemessener Sichtweisen bezüglich Variabilität. Um im Sinne von Statistical Literacy sinnvoll mit Variabilität umgehen zu können, sind aber Sichtweisen auf Variabilität, die auch den Zufall als mögliche Ursache von Unterschieden in Daten zulassen, von Bedeutung und sollten von den Lernenden entwickelt werden.

Zufällige von systematischer Variabilität zu unterscheiden, ist eine naheliegende menschliche Reaktion auf die oft überwältigende Vielfalt von relevanten und irrelevanten Informationen in unserer Umwelt (Wild & Pfannkuch, 1999). Das Herausfiltern bedeutsamer Strukturen und Muster kann mit der Metapher des Suchens nach dem Signal im Rauschen (Konold & Pollatsek, 2002) beschrieben werden. Dabei stellt das Signal eine Reduzierung der Daten auf die für die jeweilige Aufgabenstellung wesentlichen Informationen dar, wohingegen das Rauschen die nicht erklärbare, dem Zufall attribuierte Variabilität darstellt (Engel, Sedlmeier & Wörn, 2008; Vogel, 2010). Wird beispielsweise ein bestimmter Zusammenhang (z.B. zwischen Alter und Körpergröße) als systematische Ursache für Variabilität identifiziert, so sollten die Abweichungen davon, also das Rauschen, zufällig sein. Die Modellierung von Variabilität macht folglich häufig das Ausblenden von Details durch Datenreduktion mit dem Ziel des Bildens sinnvoller Überblicke notwendig (Kröpfl, Peschek & Schneider, 2000). Dies kann beispielsweise durch numerische Kennwerte wie den Durchschnitt oder durch graphische Zusammenfassungen erfolgen. Ungeachtet der Art der Reduktion ist es wichtig, die Balance zwischen voller Information und maximaler Vereinfachung je nach Aufgabenstellung zu wahren. Nur wenn die Lösung des zu Grunde liegenden Problems eine deutliche Verringerung an Komplexität erforderlich macht, ist der mit Reduktionsprozessen verbundene Verlust an Information zu rechtfertigen. Dabei kann es sich anbieten, den Informationsgehalt von Daten durch schrittweise Reduktionen zu verringern, denn nur ein angemessener Grad an Reduktion unterstützt das Erkennen von Mustern, während eine zu starke Reduzierung bedeutsame Unterschiede in den Daten verschleiern und somit die Qualität von Schlussfolgerungen gefährden kann. Insofern ist es wichtig, dass Lernende nicht nur algorithmische Fertigkeiten im Bereich der Datenreduktion erwerben, sondern entsprechende inhaltliche Vorstellungen sowie ein Bewusstsein für angemessene Arten der Reduzierung entwickeln (Kröpfl, Peschek & Schneider, 2000; Shaughnessy, 2007). Ebenso umfassen Fähigkeiten des Reduzierens von Daten das Wissen über den damit verbundenen Informationsverlust.

Die Bedeutung des Reduzierens von Daten für das sogenannte statistische Verständnis insgesamt konnte in einer Studie von Reading (2002) empirisch bestätigt werden. Fähigkeiten der Datenreduktion bildeten hier ein breites Spektrum des statistischen Verständnisses ab.

In den Beispielaufgaben aus Abbildung 2-2 werden Vor- und Nachteile sowie die Angemessenheit der reduzierten Angabe „Durchschnitt“ hinterfragt. Die Lernenden können in diesem Zusammenhang erkennen, dass der Durchschnitt eine sinnvolle Überblicksangabe ist, wenn beispielsweise abgeschätzt werden soll, wie viel Kaffee für den nächsten Monat eingekauft werden muss. Auf der anderen Seite soll hier entdeckt werden, dass der Durchschnitt nur wenig über die Breite und Streuung einer Verteilung aussagt und besonders in kleinen Stichproben sehr anfällig für Ausreißer (Extremwerte in einer monatlichen Auflistung oder in der zeitlichen Verbrauchsentwicklung einzelner Mitarbeiter) sein kann. Die Schüler sollen also je nach Problemstellung abwägen, ob der Durchschnitt einen angemessenen oder wenig geeigneten Kennwert einer Verteilung darstellt.

Abbildung 2-2

Beispielaufgaben zur Illustration wesentlicher Überlegungen bezüglich der Reduktion von Daten

Name	Kaffee-Liste Juli
Arnold	☞ ☞ ☞ ☞ ☞
Baier	II
Christian	☞ ☞ III
Domold	☞ I
Eisner	
Flick	III

Im Durchschnitt trank jeder Mitarbeiter im Juli etwa 9 Tassen Kaffee.

- Was sind Vorteile / Nachteile dieser „reduzierten“ Angabe?
- Ist der Durchschnitt ein geeigneter Wert, um den typischen Verbrauch der einzelnen Mitarbeiter zu illustrieren?
- In welcher Situation ist die Angabe des Durchschnitts sinnvoll, wann nicht?

Verschiedene Forscher bzw. Forschungsgruppen haben Testinstrumente entwickelt, mit denen Fähigkeiten des Umgangs mit Daten ermittelt werden sollen. Diese beinhalten teilweise auch oben genannte Elemente wie den Umgang mit Variabilität oder das Reduzieren von Daten. Die meisten dieser Testinstrumente werden jedoch im Hochschulbereich eingesetzt beziehungsweise zielen nicht auf Statistical Literacy ab, wie es oben beschrieben wurde (z.B. del-Mas, Garfield, Ooms & Chance, 2007; Garfield et al., 2012; Ziegler, 2014). Das Testinstrument von Watson und Callingham (2003) wurde für Lernende der Klassenstufen drei bis neun entwickelt. Dass im Zuge der Auswertung einige der Schülerantworten doppelt kodiert werden, manche Items also nicht unabhängig voneinander sind, kann als Nachteil dieses Ansatzes

gesehen werden. Zudem spielt das Verständnis für Variabilität hier eine dominante Rolle. Kuntze und Kollegen (z.B. 2008a, b; 2010) greifen den Aspekt der Variabilität auf und ergänzen ihn in einem Kompetenzmodell um weitere bedeutsame Elemente von Statistical Literacy wie oben beschrieben. Im folgenden Teilkapitel werden dieses Kompetenzmodell und das entsprechende Testinstrument vorgestellt, auf welche in der vorliegenden Studie zurückgegriffen wurde.

### **2.3 Die Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“**

Kirsch, Jungeblut und Mosenthal (1998, zitiert nach Shaughnessy, 2007, S. 964) legten ein Instrument vor, mit dem Fähigkeiten des Entnehmens von Informationen aus Graphen und Tabellen gemessen werden können. In mehreren Studien wurde gezeigt, dass sich die diesbezüglichen Fähigkeiten von Erwachsenen durch fünf hierarchische Kategorien beschreiben lassen: „locating“ (Identifizieren konkreter Elemente der Fragestellung), „cycling“ (Identifizieren einzelner Elemente nach bestimmten Vorgaben), „integrating“ (Erkennen eines gegebenen Zusammenhangs zwischen verschiedenen Elementen), „generating“ (Bilden von Zusammenhängen zwischen mehreren Elementen) und „making inferences“ (aus gegebener Information Folgerungen ableiten). Curcio (2010, S. 9, vgl. auch 1987, S. 384) entwickelte ein ähnliches Modell des Lesens in Graphen, das mit den drei Hierarchiestufen „reading the graph“, „reading between the graph“ und „reading beyond the graph“ vom einfachen Ablesen einzelner Datenpunkte über das Herstellen von Verbindungen zwischen mehreren Datenpunkten bis hin zur Vorhersage nicht gegebener Daten reicht. Shaughnessy (2007, S. 989f) ergänzte Curcios Modell um eine vierte, auf den Kontext fokussierende Stufe „reading behind the graph“. Insbesondere die Kategorie „locating“, aber auch „cycling“ von Kirsch et al. (ibid.) zeigt Parallelen zu Curcios erster Hierarchiestufe auf, die Kategorien „integrating“ und „generating“ können mit Curcios zweiter Hierarchiestufe und die Kategorie „making inferences“ mit der dritten Hierarchiestufe zur Passung gebracht werden. Die Ähnlichkeit der Ansätze von Kirsch et al. und von Curcio, validiert die beiden unabhängigen Arbeiten und legt die Existenz eines Konstrukts „datenbezogenes Lesen“ nahe (Shaughnessy, ibid.). Betrachtet man Lernende als Konsumenten von Daten (vgl. z.B. Gal, 2002), so umfasst dieses Konstrukt weite Teile des Umgangs mit Daten wie beispielsweise das Entnehmen, Verarbeiten oder Interpretieren von Informationen aus statistischen Darstellungen. Mögliche Aufgabenstellungen zu den drei auf Curcio (1987) zurückgehenden Stufen des „datenbezogenen Lesens“ werden in Abbildung 2-3 angeregt.



Abbildung 2-3

Ideen für mögliche Aufgaben zu den drei Stufen des „datenbezogenen Lesens“ in Anlehnung an Curcio (1987)

Name	Kaffee-Liste Juli
Arnold	<del>III</del> <del>III</del> <del>III</del> <del>III</del> IIII
Baier	II
Christian	<del>III</del> <del>III</del> III
Domold	<del>III</del> I
Eisner	
Flick	III

*Stufe 1:*  
Ablezen des größten oder kleinsten Wertes

*Stufe 2:*  
Vergleichen oder Aufsummieren einzelner Werte

*Stufe 3:*  
Vorhersage eines Trends ausgehend von den gegebenen Werten

Wie in früheren Forschungsarbeiten gezeigt, eignen sich Fähigkeiten des Reduzierens von Daten (Reading, 2002) sowie der Grad des Verständnisses für Variabilität (Watson & Callingham, 2003) als Indikatoren für Kompetenz bezüglich Statistical Literacy. Entsprechend der oben beschriebenen Beziehung zwischen Variabilität und Reduktion ist es sinnvoll, beide Elemente zu integrieren, um Kompetenz im Bereich von Statistical Literacy zu beschreiben und zu evaluieren. Kuntze und Kollegen (2008a, b; 2010; Kuntze, 2013) verbinden diese beiden Elemente anhand der Metapher des „datenbezogenen Lesens“ (vgl. Curcio, 1987). Zusätzlich erweitern sie Curcios Ansatz um die Idee, dass das Lesen von statistischer Information aus Darstellungen wie Tabellen, Diagrammen oder numerischen Zusammenfassungen das Nutzen statistischer Modelle miteinbezieht. Auf der Grundlage dieses verbreiterten Verständnisses des datenbezogenen Lesens etablierten Kuntze und Kollegen (ibid.) die Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“. Die Betonung des Umgangs mit Darstellungen und Modellen verbindet dieses Kompetenzmodell auch mit den deutschen Bildungsstandards, die diese Elemente als allgemeine mathematische Kompetenzen auflisten (KMK, 2003). Abbildung 2-4 zeigt einen Überblick über die schematisch zusammengefassten Überlegungen, die in der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ zusammenwirken.

Das von Kuntze und Kollegen entwickelte Kompetenzmodell wurde schrittweise von zunächst drei Kompetenzniveaus (Lindmeier et al., 2007) über vier Niveaus (Kuntze et al., 2008a, b) auf zuletzt fünf Niveaus (Kuntze et al., 2010; Kuntze, 2013) erweitert. Die hierarchische Struktur des jeweiligen Modells wurde über entsprechende Testinstrumente in verschiedenen Stichproben empirisch bestätigt. Ein Überblick über die Kompetenzniveaus und

deren Abgrenzung voneinander findet sich in Tabelle 2-1. Im folgenden Text werden die Anforderungen der einzelnen Stufen teils auch anhand von Beispielaufgaben erläutert.

Abbildung 2-4

Elemente der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ (Kuntze, 2013, S. 7)

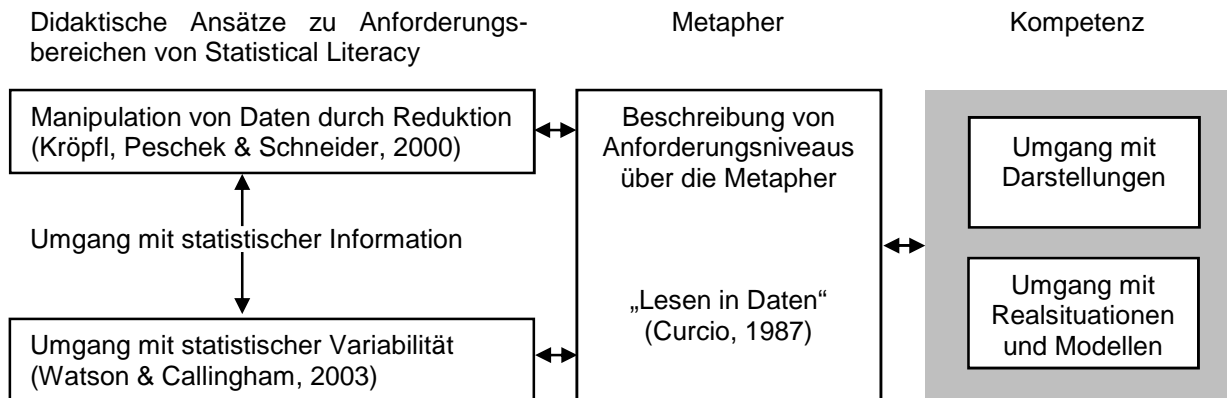


Tabelle 2-1

Kompetenzmodell zum „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ (Kuntze, 2013, S. 7)

Kompetenzniveau I	Einschrittiges Nutzen einer Darstellung oder Arbeiten innerhalb eines gegebenen Modells (z.B. Ablesen eines gegebenen Werts aus einem Diagramm, Vervollständigen eines gegebenen Diagramms bei gegebener Datentabelle)
Kompetenzniveau II	Zwei- oder mehrschrittiges Nutzen von Darstellungen oder Wechsel zwischen zwei gegebenen Modellen (z.B. Vergleichen von Daten unter Einschluss eines transformierenden Schrittes oder von Begriffswissen)
Kompetenzniveau III	Mehrschrittiges Nutzen von Darstellungen einschließlich der Nutzung eines nicht gegebenen Modells (z.B. eigene Modellierungsaktivitäten zur Unterstützung einer kumulativen Interpretation von in Diagrammen gegebenen Daten)
Kompetenzniveau IV	Mehrschrittiges Nutzen von Darstellungen bzw. Nutzung eines nicht gegebenen Modells einschließlich des adäquaten Umgangs mit statistischer Variabilität (z.B. eigene Modellierungsaktivitäten auf der Basis von Diagrammdarstellungen, die den Umgang mit statistischer Variabilität mit einschließen)
Kompetenzniveau V (Projekt RIKO-STAT)	Mehrschrittiges Nutzen von Darstellungen bzw. Nutzung eines nicht gegebenen Modells einschließlich des adäquaten Umgangs mit statistischer Variabilität und des Modellierens/mentalen Konstruierens nicht gegebener Daten (z.B. mögliche zugrunde liegende Daten modellieren beim Beurteilen oder Interpretieren gegebener Aussagen)

Kompetenzniveau I erfordert ein lediglich einschrittiges Umgehen mit statistischen Darstellungen und Modellen, wie beispielsweise beim Entnehmen einzelner Werte aus einer Tabelle oder einem Diagramm. Dies entspricht dem grundlegenden Nutzen eines gegebenen Modells, wenn zum Beispiel graphische Elemente zur Skala eines Diagramms in Verbindung gesetzt werden müssen, um eine Größe aus dem Diagramm abzulesen.

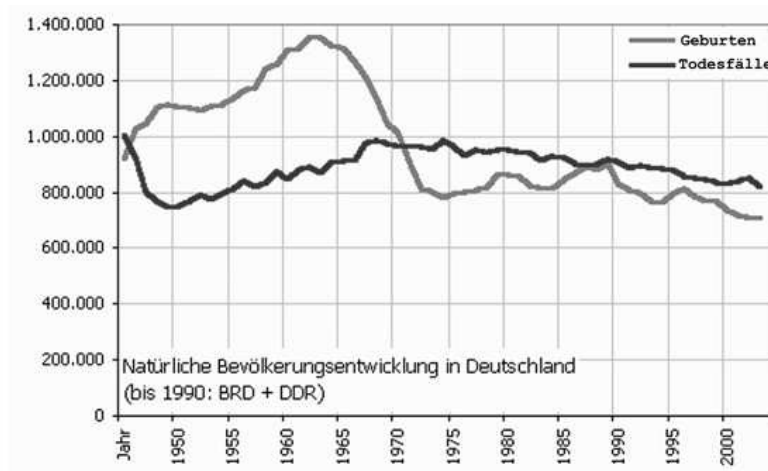
Die Anforderungen auf Kompetenzniveau II umfassen dagegen ein zwei- oder mehrschrittiges Nutzen von Darstellungen und Modellen. Dies ist beispielsweise notwendig, wenn zwischen zwei gegebenen Darstellungen oder Modellen gewechselt werden muss. Items auf dieser Niveaustufe verlangen es zum Beispiel, dass einzelne Werte eines Diagramms ausgewählt und aufsummiert oder dass die Daten zweier Diagramme miteinander verglichen werden.

Auf diesen beiden untersten Kompetenzniveaus sind die zur Lösung der Aufgabe erforderlichen Modelle bereits gegeben und müssen von den Lernenden nicht selbst gebildet werden. Eigene Modellierungsaktivitäten sind dagegen ab Kompetenzniveau III zusätzlich zum mehrschrittigen Nutzen von statistischen Darstellungen erforderlich. Abbildung 2-5 zeigt eine Beispielaufgabe des Kompetenzniveaus III: Um herauszufinden, ab welchem Jahr die Bevölkerung zurückging, müssen beide Kurven zueinander in Bezug gesetzt werden. Es kann beispielsweise ein Modell der Bevölkerung als eine Art Bestand gebildet werden, zu dem Neugeborene hinzukommen und von dem Todesfälle abgezogen werden.

Abbildung 2-5

Beispielaufgabe auf Kompetenzniveau III (vgl. Lindmeier, Kuntze & Reiss, 2007, S. 4)

Betrachte folgendes Diagramm zur Bevölkerungsentwicklung in Deutschland:



Ab welchem Jahr ging die Bevölkerung zurück? Begründe.

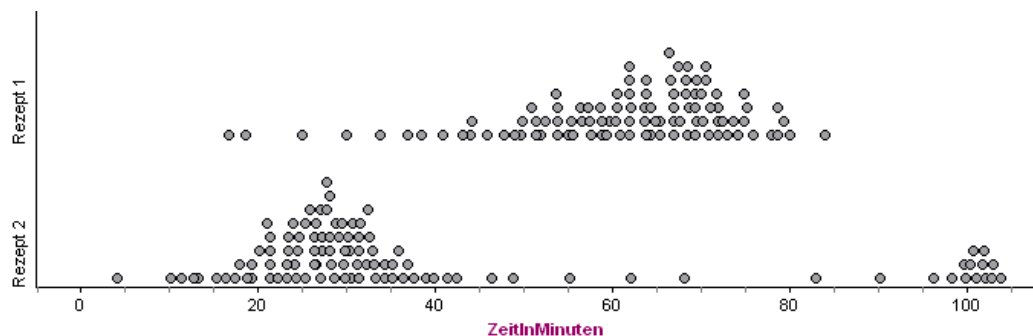
Ab Kompetenzniveau IV wird zusätzlich zum mehrschrittigen Nutzen von Darstellungen und eigenen Modellierungsaktivitäten der angemessene Umgang mit Variabilität erforderlich. Abbildung 2-6 illustriert ein Beispielitem auf Niveaustufe IV, bei dem zwei Verteilungen des Wirkungsbeginns von Kopfschmerztabletten miteinander verglichen werden sollen. Da sich diese Verteilungen sowohl in ihren Mittelwerten als auch in ihrer Streuung unterscheiden, muss hier Variabilität modelliert werden, um die geforderten Argumente liefern zu können. Es können anspruchsvolle Modelle (z.B. Betrachtung von Streuung oder Varianz), aber auch einfache Modelle (z.B. Vergleichen von Häufigkeiten in bestimmten Zeitintervallen) zur Lösung der Aufgabe gebildet werden.

Abbildung 2-6

Beispielaufgabe auf Niveaustufe IV (vgl. Kuntze et al., 2008a, S. 117)

Ein Hersteller von Tabletten gegen Kopfschmerzen hat zwei verschiedene Rezepte entwickelt. Die Tabletten sind in einem Labor an je 100 Personen mit Kopfschmerzen getestet worden.

In der Grafik ist dargestellt, wie lange es gedauert hat, bis die Kopfschmerzen verschwunden waren. Ein Punkt entspricht einer Person.



Doktor Huber meint:

Rezept 2 ist besser als Rezept 1!

Finde Argumente **dafür** und **dagegen**!

Die höchste Niveaustufe V erfordert auf der Grundlage eines mehrschrittigen Nutzens von Darstellungen sowie eigenen Modellierungsaktivitäten unter Einbezug von statistischer Variabilität insbesondere das mentale Generieren nicht gegebener Daten, die bestimmten Bedingungen genügen müssen. In der Beispielaufgabe in Abbildung 2-7 müssen zu einer vorgegebenen Durchschnittslänge mögliche Datenpunkte unter Berücksichtigung von Variabilität erzeugt und angeordnet werden. Zur Lösung der Aufgabe kann beispielsweise der Durchschnitt als eine Art Ausgleichswert modelliert werden, der die Abweichungen nach unten und

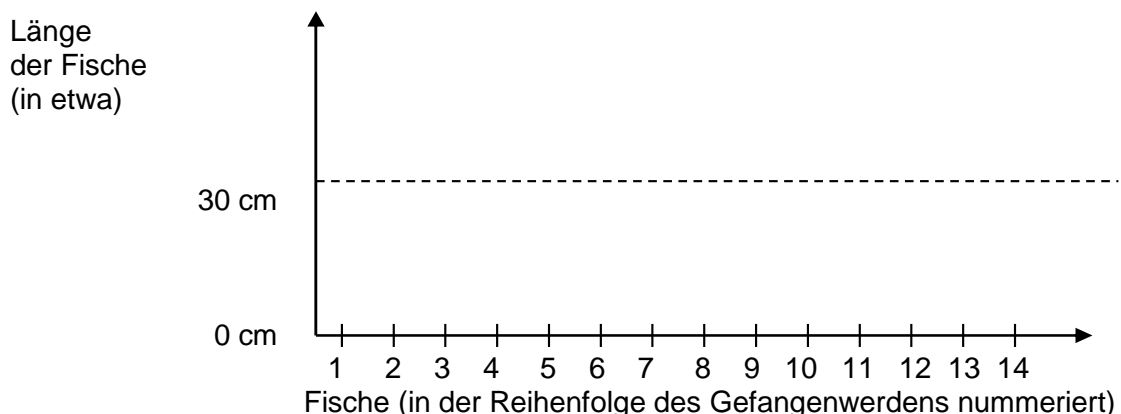
oben ausbalanciert. Des Weiteren muss bei der Erzeugung der Datenpunkte berücksichtigt werden, dass es mehr Abweichungen vom Mittelwert nach unten als nach oben gibt.

#### Abbildung 2-7

Beispielaufgabe auf Kompetenzniveau V (Kuntze, 2013, S. 8)

Ein Angler fängt Fische. Er stellt fest, dass Abweichungen von der Durchschnittslänge nach oben (übergroße Fische) viel seltener auftreten als nach unten (unterdurchschnittlich große Fische).

Entwickle für eine Durchschnittsgröße von 30 cm eine typische Verteilung, die zu dieser Beschreibung passt, und stelle diese im folgenden Diagramm dar:



Die abgebildeten Beispielaufgaben sind Bestandteil eines Testinstruments bezüglich der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“, das in der momentanen Form mit fünf Niveaus insgesamt 15 Aufgaben umfasst. Die hierarchische Struktur des Kompetenzkonstrukts entlang einer Fähigkeitsdimension konnte in verschiedenen Stichproben durch Rasch-Analysen des jeweiligen Testinstruments (mit Aufgaben zu drei, vier und fünf Kompetenzniveaus) sowie durch relative Lösungshäufigkeiten der Aufgaben auf den einzelnen Kompetenzniveaus empirisch bestätigt werden (z.B. Fröhlich, Kuntze & Lindmeier 2007; Kuntze, Engel, Martignon & Gundlach 2010; Kuntze, Lindmeier & Reiss 2008a, b; Lindmeier et al., 2007). Die Rasch-Skalierbarkeit ermöglicht Vergleiche der Leistungen sowie der Leistungsentwicklung zwischen Lernenden(gruppen).

Trotz des offensichtlichen Überschneidungsbereichs zwischen der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ und der Leitidee „Daten und Zufall“ der deutschen Bildungsstandards (KMK, 2003) unterscheiden sich diese in wesentlichen Punkten. „Daten und Zufall“ fasst konkrete Inhalte, Wissens Elemente und Fähigkeiten des Umgangs mit Daten zusammen, wohingegen das vorgestellte Kompetenzkonstrukt verschiedene Stufen einer spezifischen Kompetenz im Bereich von Statistical Literacy beschreibt.

Außerdem unterscheiden sich das Kompetenzkonstrukt und „Daten und Zufall“ in Bezug auf wesentliche inhaltliche Aspekte. Beispielsweise spielt „Wahrscheinlichkeit“ eine bedeutsame Rolle innerhalb der Leitidee „Daten und Zufall“, während dies kein explizites Element des Kompetenzmodells darstellt. Andererseits ist der Umgang mit Variabilität ein wesentlicher Aspekt des Kompetenzmodells, wird aber in „Daten und Zufall“ nicht konkret aufgeführt. In ähnlicher Weise unterscheidet sich das Kompetenzkonstrukt „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ auch von der PISA-Kategorie „Unsicherheit“ (siehe z.B. Blum et al., 2004). Aus diesen Gründen können Ergebnisse aus der Forschung bezüglich „Daten und Zufall“ oder „Unsicherheit“ nicht direkt auf die dieser Studie zu Grunde liegende Kompetenz übertragen werden. In der vorliegenden Dissertation soll mit dem „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ vielmehr eine Kernkompetenz im Bereich von Statistical Literacy untersucht werden.

Bezüglich der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ liegen bereits Vorläuferstudien vor (Projekt KOMMA, Projekt RIKO-STAT), auf die in dieser Arbeit aufgebaut werden kann. Über die Entwicklung und Erprobung des Kompetenzmodells bzw. entsprechender Testinstrumente hinaus (z.B. Kuntze et al., 2010; Kuntze et al., 2008a, b; Lindmeier et al., 2007) wurden in diesen Studien mögliche Bedingungsvariablen der Kompetenz beleuchtet. Beispielsweise wurden Zusammenhänge mit benachbarten Bereichen von Begriffswissen durch Korrelations- bzw. Regressionsanalysen untersucht, wobei sich Begriffswissen zu Wahrscheinlichkeit als etwas enger mit dem Kompetenzkonstrukt verbunden herausstellte als Wissen zu Funktionen oder Risiko (Kuntze, 2013; Kuntze et al., 2010). Die hier ermittelten bivariaten Korrelationen lagen zwischen  $r = 0,3$  und  $r = 0,4$ . Außerdem wurden Zusammenhänge zwischen der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ und auf Mathematik bzw. Statistik bezogenen motivationalen Variablen analysiert (Gundlach, Kuntze, Martignon & Engel, 2010; Kuntze, 2013). Diese stellten sich im Vergleich zu anderen Inhaltsbereichen als in erwartungsgemäßer Größenordnung heraus (vgl. Helmke & Weinert, 1997). Auf Grundlage der genannten Vorgängerarbeiten wurden in der vorliegenden Dissertation weitere Erkenntnisse bezüglich der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ gewonnen. Zunächst wurde analysiert, inwieweit diese Kompetenz mit allgemeinen Lernervariablen wie Leseverständnis und allgemeinen kognitiven Fähigkeiten zusammenhängt. Die Untersuchung von Zusammenhängen mit solchen allgemeinen Lernermerkmalen weitet die Befunde bisheriger Arbeiten aus und dient u.a. dem Zweck empirisch zu bestätigen, dass es sich um eine bereichsspezifische Kompetenz handelt, die nicht vollständig von allgemeinen Lernervoraussetzungen determi-

niert wird. Außerdem wurden in der vorliegenden Arbeit Effekte einer statistikspezifischen Intervention auf diese Kompetenz untersucht. Dies soll Erkenntnisse über Gestaltungsmerkmale lernförderlicher Materialien hervorbringen.

Die in diesem Projekt entwickelte Intervention sollte nicht nur die Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ fördern, sondern auch Effekte auf bereichsspezifische motivationale Variablen haben. Im folgenden Abschnitt wird der theoretische Rahmen bezüglich der hier fokussierten motivationalen Variablen skizziert.

## **2.4 (Bereichsspezifische) motivationale Variablen**

Lange Zeit wurden motivationale Variablen in der Mathematikdidaktik nur wenig beachtet (McLeod, 1992). Inzwischen werden sie auch deshalb als bedeutsame Ziele von (Mathematik)Unterricht angesehen, da sie den Lernprozess und somit Leistungsvariablen beeinflussen können (Marsh, Trautwein, Lüdtke, Köller & Baumert, 2005; Nagy, Trautwein, Baumert, Köller, & Garrett, 2006; Pekrun & Zirngibl, 20004). Obgleich Gal und Kollegen bereits 1997 die Bedeutung motivationaler Variablen auch bezogen auf den Bereich Statistik betont haben, gibt es hier im Vergleich zur Mathematik bislang nur wenige Studien, die diesbezügliche Variablen in den Blick genommen haben (z.B. Bond, Perkins & Ramirez, 2012; Ramirez, Schau & Emmiöglu, 2012). Die vorliegende Dissertation fokussierte in Teilstudie 3 auf die Förderung von bereichsspezifischem Selbstkonzept und Interesse durch eine statistikbezogene Intervention. Im Folgenden werden diese Variablen und deren Zusammenhang zu Leistungsvariablen näher beschrieben.

### **2.4.1 Selbstkonzept**

Das akademische Selbstkonzept drückt das Vertrauen einer Person in die eigenen Fähigkeiten bezogen auf einen bestimmten Bereich wie Mathematik oder Statistik aus (Bong & Skaalvik, 2003). Es wird hauptsächlich durch frühere Kompetenzerfahrungen gebildet und gilt als relativ stabil (ibid.). Insofern erscheint es plausibel, dass hohe Kompetenz in der Regel auch zu hohen Ausprägungen von Selbstkonzept führt. Andererseits kann aber auch ein hohes Selbstkonzept Leistung in positiver Weise beeinflussen (Valentine, DuBois & Cooper, 2004, S. 113f): Beispielsweise nehmen Lernende mit einem höheren Selbstkonzept erfolgsorientierte Verhaltensweisen (z.B. Üben) eher auf bzw. behalten diese bei. Dadurch erhalten sie mehr korrekatives Feedback und wählen auch schwierigere Aufgaben, was wiederum zu besseren Leistungen führen kann. Auch empirisch konnten solche wechselseitigen Zusammenhänge zwischen akademischem Selbstkonzept und Leistung bereits bestätigt werden (z.B. Marsh et al., 2005).

Zudem haben empirische Studien (z.B. Hansford & Hattie, 1982; Marsh, Trautwein, Lüdtke, Köller & Baumert, 2006) gezeigt, dass Korrelationen zwischen Leistungsvariablen und Selbstkonzept umso höher ausfallen, je enger das untersuchte Selbstkonzept auf den jeweiligen Leistungsbereich bezogen ist. Auch aus diesem Grund wird für empirische Studien empfohlen, das akademische Selbstkonzept so bereichsspezifisch wie möglich zu erheben (Marsh & Craven, 1997). Diese Einschätzung passt zu Befunden von O'Mara und Kollegen (2006, S. 197ff), die in einer Meta-Analyse festgestellt haben, dass Interventionen zur Verbesserung des Selbstkonzepts besonders wirksam sind, wenn sich Intervention und Selbstkonzept auf den gleichen Inhaltsbereich beziehen. Des Weiteren zeigte sich in dieser Studie, dass besonders Interventionen, die auf die Gabe von Feedback abzielen oder bereichsspezifische Fähigkeiten fördern, positive Effekte auf das akademische Selbstkonzept haben. Bei der Konzeption der Intervention im Rahmen dieser Dissertation wurden solche Merkmale, die sich als günstig für die Förderung des Selbstkonzepts gezeigt haben, aufgenommen.

#### **2.4.2 *Interesse***

Interesse gilt als eine spezifische Beziehung zwischen einer Person und einem bestimmten Inhalt oder Objekt (z.B. Krapp, 2007, S. 8) und ist somit eine inhalts- bzw. bereichsspezifische Variable. Laut Schiefele (1992) schreiben Personen ihren Interessensobjekten eine hohe subjektive Bedeutung zu (wertbezogene Valenzen) und verbinden positive Gefühle damit (gefühlbezogene Valenzen). Insofern wird Interesse als ein Auslöser für intrinsische Motivation angesehen. Nur wenn wert- und gefühlbezogene Valenzen positiv erlebt werden, kann sich Interesse entwickeln (Krapp, 2005). Dies erfordert laut der Selbstbestimmungstheorie der Motivation (z.B. Deci & Ryan, 2002) die Befriedigung dreier psychologischer Grundbedürfnisse: Individuen wollen sich als kompetent und autonom bei der Ausübung einer Tätigkeit wahrnehmen. Zudem ist es wichtig, dass sie sich sozial eingebunden fühlen. Aufgrund der Bedeutung dieser drei Grundbedürfnisse für die Entwicklung von Interesse sollten Interventionen, die auf die Förderung dieser Variable abzielen, das Erleben von Kompetenz, Autonomie und sozialer Eingebundenheit ermöglichen. Bei der Entwicklung der Intervention der vorliegenden Arbeit wurde dies berücksichtigt.

Entsprechend der Selbstbestimmungstheorie von Deci und Ryan (ibid.) kann u.a. (wahrgenommene) Kompetenz das Interesse an einem bestimmten Objekt beeinflussen. Aber auch das bestehende Interesse an einem Gegenstand kann Einfluss auf Leistungsvariablen haben (Schiefele, 1991, S. 307ff): Interessierte Lerner nehmen eher intensive und sinnorientierte Lernaktivitäten auf, sie nutzen elaboriertere Lernstrategien und investieren insgesamt mehr



Zeit und Energie für das Lernen. Diese wechselseitige Beeinflussung von Interesse und Leistung zeigt sich auch in empirisch ermittelten Zusammenhängen zwischen diesbezüglichen Variablen (z.B. Marsh et al., 2005).

Die vielfach empirisch nachgewiesenen Zusammenhänge zwischen Leistungsvariablen und Interesse sowie Selbstkonzept (z.B. Baumert & Köller, 2000; Klieme, Neubrand & Lüdke, 2001) verdeutlichen, warum diese motivationalen Variablen inzwischen als wichtige Ziele von Unterricht angesehen werden. Wie oben erwähnt gibt es bezogen auf den Bereich Statistik noch wenige Untersuchungen, die sich mit diesen Variablen beschäftigt haben. In Vorgängerstudien dieser Dissertation wurden Instrumente zur Erhebung statistikspezifischer motivationaler Variablen entwickelt (z.B. Gundlach et al., 2010), auf die hier zurückgegriffen wird. Mit diesen soll die Entwicklung des bereichsspezifischen Selbstkonzepts und Interesses im Verlauf einer statistikbezogenen Intervention ermittelt werden. Im Folgenden wird die im Rahmen dieser Arbeit konzipierte Intervention vorgestellt.

## **2.5 Zur Beschreibung der Intervention dieser Studie**

Die Intervention dieser Dissertation berücksichtigte in ihrer grundsätzlichen Anlage Merkmale, die als förderlich für die Entwicklung von Selbstkonzept und Interesse angesehen werden: Die Aufgaben der Intervention wurden von den Lernenden in festen Zweiertteams bearbeitet, also einer kooperativen und sozialen Lernform, die das Grundbedürfnis der sozialen Eingebundenheit berücksichtigt (vgl. Deci & Ryan, 2002). Des Weiteren erhielten die Schüler regelmäßig Feedback zu ihren Bearbeitungen in Form von standardisierten Musterlösungen und Hilfekarten, die jederzeit zur Verfügung standen sowie in Form von individueller Rückmeldung. Durch die Gabe von Feedback konnte die eigene Kompetenz erfahrbar werden, was sowohl für die Entwicklung von Interesse (Krapp, 2005) als auch von Selbstkonzept (Bang & Skaalvik, 2003) als bedeutsam erachtet wird. Außerdem waren die Schüler in ihrem Lernprozess weitgehend autonom und konnten diesen zu großen Teilen (z.B. Lerntempo, Art der Bearbeitung) selbst steuern. Auch dies gilt im Sinne der Selbstbestimmungstheorie von Deci & Ryan (ibid.) als wichtige Voraussetzung für die Entwicklung von Interesse. Die Intervention zielte zudem darauf ab, Kompetenz im Bereich von Statistical Literacy zu entwickeln. Im Folgenden wird die inhaltliche Ausrichtung und der Intervention näher erläutert.

Die Relevanz des Umgangs mit Variabilität und datenbezogener Reduktion für das Arbeiten mit statistischer Information spricht dafür, dass die Förderung dieser Aspekte den Aufbau von Statistical Literacy insgesamt begünstigen kann. In den deutschen Bildungsstandards wird der Umgang mit Variabilität nicht explizit benannt (KMK, 2003), weshalb davon auszugehen ist,

dass dies hierzulande kaum Gegenstand von (Mathematik-)Unterricht darstellt. Fishbein (1975) sieht den schulischen Unterricht aufgrund der Vermittlung eines deterministischen Denkens sogar als kontraproduktiv für den Aufbau eines Verständnisses für zufallsbedingte Variabilität an:

„The child is taught [in school] that explanation consists in specifying a cause; that a scientific prediction must be a certainty; that ambiguity and uncertainty are not acceptable in scientific reasoning and so on. Even if all this is not explicitly stated, it is implied in all that is taught in school.“ (ibid., S. 71)

Green (1990; 1986; 1982) sowie Engel und Sedlmeier (2005) kommen in jüngeren Studien zu ähnlichen Schlussfolgerungen, was die gezielte unterrichtliche Auseinandersetzung mit zufallsbedingter Variabilität nahe legt. Ohne die Thematisierung dieses Phänomens mag auch die Einsicht in die Notwendigkeit und Sinnhaftigkeit von Datenreduktionen, die grundsätzlich in Form von numerischen oder grafischen Zusammenfassungen in den Bildungsstandards genannt werden, nur schwer zu vermitteln sein: Wenn sich Lernende beispielsweise nicht damit auseinandersetzen, aus welchen Gründen sich Daten in einem bestimmten Kontext unterscheiden könnten oder bis zu welchem Grad vorliegende Unterschiede in Daten als zufällig betrachtet werden können, werden Datenreduktionen auf einem eher algorithmischen Bilden von Kennwerten oder Erstellen von graphischen Zusammenfassungen verbleiben und es wird kaum über eine - je nach Kontext – sinnvolle Art der Datenreduktion reflektiert. Insofern erscheint die Verbindung dieser zwei Elemente bei der Erstellung von Lernmaterialien als sinnvoll.

Die im Rahmen dieser Interventionsstudie konzipierten Lernmaterialien greifen die Idee auf, Kompetenz im Bereich von Statistical Literacy speziell durch Reflexionen bezüglich statistischer Variabilität und datenbezogener Reduktion zu fördern. Dazu wurden Aufgaben des elementaren und eher algorithmischen Umgangs mit statistischen Darstellungen und Modellen (z.B. Bilden von statistischen Kennwerten, Lesen und Erstellen von Grafiken und Tabellen), wie er auch Gegenstand von Schulbuchaufgaben ist (z.B. Erheben, Darstellen und Reduzieren von Daten bei Baum & Klein, 2004; Böttner et al., 2005; Griesel, Postel & Vom Hofe, 2004), durch darauf aufbauende Reflexionsanlässe bezüglich Reduktion und Variabilität, sogenannte „Prompts“, ergänzt (Aufgabenbeispiele siehe Abbildungen 4-1 bis 4-3 im Methodenteil). Um zu ermitteln, inwieweit diese Reflexionen über die grundlegenden Aufgaben hinaus zur Kompetenzentwicklung beitragen können sowie um differenzielle Effekte der Reflexionsanlässe zu erfassen, sah das Design der vorliegenden Studie vier verschiedene Treatments im 2x2-

Design vor: Über die Aufgaben des „Basis-Trainings“ hinaus wurden die Lernenden der weiteren drei Treatments („Reduktion“, „Variabilität“ sowie „Reduktion und Variabilität“) dazu aufgefordert, im jeweiligen Aufgabenkontext über Variabilität oder Reduktion oder beide Aspekte nachzudenken. Der Gedanke, dass es im Umgang mit Daten wichtig ist, die Balance zwischen Repräsentativität und Variabilität zu wahren (Kröpfl, Peschek & Schneider, 2000; Watson & Kelly, 2004), legt es nahe, dass eine optimale Förderung beide Aspekte miteinander verbindet und somit das kombinierte Treatment zur besten Kompetenzentwicklung führt.

Obwohl der in dieser Arbeit verwendete Kompetenztest Aufgaben zum Umgang mit statistischer Variabilität enthält, kann man auf Grundlage der Bearbeitung dieser Aufgaben nicht direkt auf die Sichtweisen der Lernenden auf Variabilität schließen. Aus diesem Grund wurden Effekte der Intervention nicht nur auf die Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ untersucht, sondern auch auf die Sichtweisen der Lernenden bezüglich Variabilität. Die Erfassung dieser Sichtweisen erfolgte in Anlehnung an bestehende diesbezügliche Forschung (Engel, 2011; Kuntze, 2012) und wird im Methodenteil exemplarisch erläutert.

### **3 Das Projekt ReVa-Stat: Überblick und Forschungsinteresse der Teilstudien**

Aus der alltagsrelevanten Bedeutung von Statistical Literacy sowie teils gravierender diesbezüglicher Probleme von Schülern und Erwachsenen ergibt sich der Wunsch nach der Förderung verschiedener Variablen in diesem Bereich. An diesem Punkt setzte das Forschungsprojekt ReVa-Stat mit drei Teilstudien an. Im Sinne der multikriterialen Zielerreichung von Unterricht (Pekrun & Zirngibl, 2004) zielte ReVa-Stat neben der Unterstützung kognitiver Variablen darauf ab, das bereichsspezifische Selbstkonzept und Interesse der Lernenden zu fördern. Die Wirksamkeit der entwickelten Intervention wird somit in Hinblick auf verschiedene Zielvariablen hin evaluiert und soll für die unterrichtliche Praxis nutzbar sein. Darüber hinaus wurden durch diese Arbeit Erkenntnisse zur Verbreiterung der fachdidaktischen Theoriebasis gewonnen. Im Folgenden wird ein Überblick über die drei Teilstudien und deren Zielsetzungen gegeben.

Teilstudie 1 befasste sich mit der Untersuchung der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ im Vorfeld der Intervention. Mittels des oben beschriebenen Testinstruments, das in allen Teilstudien dieses Projekts eingesetzt wurde, sollte zunächst empirisch überprüft werden, ob sich das zu Grunde liegende Kompetenzkonstrukt als spezifisch für den entsprechenden Inhaltsbereich bestätigte oder ob das Testinstrument im Wesentlichen allgemeine Lernervoraussetzungen abbildete. Durch Mehrebenenanalysen wur-

de untersucht, inwieweit das Leseverständnis und allgemeine kognitive Fähigkeiten mit dem Kompetenzkonstrukt zusammenhängen. Als zusätzliche Kontextvariablen wurden die Mathematiknote, das Geschlecht und der sozioökonomische Status auf der Individual- bzw. Klassenebene miteinbezogen. Durch die Beleuchtung von Zusammenhängen zwischen der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ und den genannten Lernervariablen konnte das Kompetenzkonstrukt empirisch von diesen abgegrenzt und somit für den spezifischen Inhaltsbereich bestätigt werden.

In Teilstudie 2 wurde zunächst in den Blick genommen, über welche Kompetenzausprägung bezüglich des „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ die Schüler der vorliegenden Stichprobe vor und nach der Intervention verfügten. Darüber hinaus wurde die Wirkung von vier verschiedenen Treatments (2x2-Design) einer statistikspezifischen Intervention auf diese Kompetenz sowie auf Sichtweisen bezüglich zufallsbedingter Variabilität untersucht. Es wurden sowohl individuelle Voraussetzungen wie das Leseverständnis, kognitive Fähigkeiten, die Mathematiknote und das Geschlecht als auch die Zugehörigkeit zum jeweiligen Treatment als potentielle Einflussfaktoren in multiple Regressionsmodelle aufgenommen. Durch dieses Vorgehen sollte analysiert werden, welche Variablen die Entwicklung von Sichtweisen bezüglich Variabilität und der beschriebenen Kompetenz im Verlauf der Intervention begünstigten. Die Ergebnisse dieser Teilstudie sollten über die Verbreiterung der theoretischen Basis hinaus Erkenntnisse über Gestaltungskriterien förderlicher Lernmaterialien liefern, die auch in der unterrichtlichen Praxis genutzt werden können.

Ergänzend zur Fragestellung von Teilstudie 2 beschäftigte sich Teilstudie 3 mit den Effekten der Intervention auf domänenspezifische motivationale Lernervariablen. Zunächst wurde analysiert, inwieweit die Skalen zur Erhebung des mathematik- und statistikbezogenen Selbstkonzepts und Interesses empirisch trennbare Faktoren darstellten. Da das akademische Selbstkonzept und Interesse bezüglich spezifischer Inhaltsbereiche wichtige unterrichtliche Zielvariablen darstellen, wurde untersucht, inwieweit die statistikbezogene Intervention deren Entwicklung förderte. Um die Veränderungen dieser Variablen einordnen zu können, wurden die Mittelwerte von Vor- und Nachtest jeweils mit denen einer Baseline-Kontrollgruppe als Referenzpunkt verglichen. Zusätzlich wurde analysiert, inwieweit eine hohe Kompetenz bezüglich des „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ (Vortest) die Entwicklung des statistikbezogenen Selbstkonzepts und Interesses im Verlauf der Intervention noch unterstützte. Neben Erkenntnissen über den Zusammenhang zwischen motivationalen Variablen und Kompetenz im Bereich von Statistical Literacy, konnten auch aus dieser Teil-

studie Erkenntnisse über Gestaltungsmerkmale von förderlichen Unterrichtsmaterialien erlangt werden.

Im Folgenden sind die in den einzelnen Teilstudien verfolgten Forschungsfragen explizit aufgelistet. Zudem finden sich diese selbstverständlich bei den Kapiteln zu den jeweiligen Teilstudien.

### **3.1 Zusammenfassende Darstellung der Forschungsfragen der drei Teilstudien**

Teilstudie 1 befasste sich mit möglichen Einflussfaktoren auf die Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ im Vorfeld der Intervention. Die Forschungsfragen lauteten:

- 1) Wie ist der Zusammenhang zwischen Leseverständnis und der Kompetenz “Nutzen von Darstellungen und Modellen in statistischen Kontexten”?
- 2) Inwieweit hängt die Kompetenz mit allgemeinen kognitiven Fähigkeiten zusammen?
- 3) Inwieweit modellieren die Mathematiknote, das Geschlecht und der sozioökonomische Status den Zusammenhang zwischen der Kompetenz und Leseverständnis sowie allgemeinen kognitiven Fähigkeiten?

Teilstudie 2 nahm die Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ sowie Sichtweisen auf Variabilität im Verlauf der Intervention in den Blick. Dabei wurden folgende Forschungsfragen untersucht:

- 1) Über welche Kompetenz des „Nutzens von Darstellungen und Modellen in statistischen Kontexten“ verfügen die Lernenden?
- 2) Inwieweit begünstigten die Lernervoraussetzungen Leseverständnis, kognitive Fähigkeiten, die Mathematiknote und das Geschlecht die Entwicklung dieser Kompetenz?
- 3) Welche Treatments des 2x2-Designs wirkten besonders förderlich auf die Kompetenz?
- 4) Inwieweit beeinflussten die Lernervoraussetzungen Leseverständnis, kognitive Fähigkeiten, die Mathematiknote und das Geschlecht die Entwicklung von Sichtweisen auf Variabilität?
- 5) Welche Treatments wirkten besonders förderlich auf die Entwicklung von Sichtweisen auf Variabilität?

Teilstudie 3 fokussierte auf die Effekte der Intervention auf bereichsspezifische motivationale Variablen. Die diesbezüglichen Forschungsfragen lauteten:

- 1) Inwieweit sind mathematik- und statistikbezogenes Selbstkonzept und Interesse empirisch trennbar?
- 2) Kann das bereichsspezifische Selbstkonzept und Interesse von Schülern durch eine statistikbezogene Intervention gefördert werden? Wenn ja, gibt es Unterschiede zwischen den vier Treatments?
- 3) Welche Rolle spielt die Ausprägung der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ zum Vortestzeitpunkt für die Entwicklung des statistikbezogenen Selbstkonzepts und Interesses?

## **4 Methode**

### **4.1 Stichprobe und Design der Studie**

Insgesamt nahmen über 600 Schüler aus 29 achten Klassen am Projekt ReVa-Stat teil. Darunter absolvierten die 25 Interventionsklassen im Schuljahr 2012/13 alle im Folgenden beschriebenen Tests und die Intervention, während die vier Baseline-Kontrollklassen im Schuljahr 2013/14 lediglich eine Auswahl an Testteilen bearbeiteten. Die Lernenden zwischen 12 und 16 Jahren, davon knapp die Hälfte weiblich, besuchten acht Realschulen in Baden-Württemberg. Obwohl bei der Rekrutierung darauf geachtet wurde, dass sowohl Schulen mit eher ländlichem als auch mit eher städtischem Einzugsgebiet in die Studie integriert wurden, sind aufgrund der Freiwilligkeit der Studienteilnahme Selektionseffekte auf Schüler-, Klassen- und Schulebene nicht ausgeschlossen. Teilweise unterscheiden sich die den drei Teilstudien zu Grunde liegenden Stichprobengrößen, da nicht alle Probanden zu sämtlichen Testzeitpunkten anwesend waren beziehungsweise an der Intervention teilnahmen. Genaue Angaben zur jeweiligen Stichprobe finden sich im Methodenteil der entsprechenden Teilstudie. Sowohl die Implementierung der Testungen als auch der Intervention wurden von einem Forscher und einem regulär unterrichtendem Lehrer beaufsichtigt.

Die in der Intervention verwendeten Materialien wurden im Schuljahr 2011/12 entwickelt und unter Studierenden der Pädagogischen Hochschule Ludwigsburg sowie in einem zweiten Schritt in drei achten Realschulklassen pilotiert. Diese Vorstudie diente dem Zweck, Rückmeldung über die Klarheit der Formulierungen, Fragen und Abbildungen zu erhalten und die Materialien dementsprechend optimieren zu können. Des Weiteren wurde über anonyme Fragebögen erhoben, ob sich die Schüler einer Klasse bei paralleler Beschäftigung mit verschiedenen Treatment-Materialien über diese austauschen würden. Da sich keine Übertragungseffekte zeigten, konnte für die Hauptstudie an der Bildung von vier Treatmentgruppen innerhalb jeder Klasse festgehalten werden.

Im Rahmen der Hauptstudie wurden die Schüler der 25 Interventionsklassen innerhalb von jeder Klasse parallelisiert nach ihren Vortestleistungen randomisiert auf vier Treatmentgruppen im 2x2-Design verteilt. Innerhalb dieser Gruppen bestimmten die jeweiligen Lehrer Paare, die die Lernumgebungen in Partnerarbeit absolvierten. Während der vierstündigen Intervention (4 x 45min) erhielten die Probanden des „Basis-Trainings“ (1) Aufgaben, die auf grundlegende Fähigkeiten des Arbeitens mit statistischen Darstellungen wie Tabellen, Diagrammen und Kenngrößen abzielten. Einen Teil dieser Grundaufgaben bearbeiteten auch die Schüler der anderen drei Treatments, wobei hier überdies in sogenannten Prompts spezielle Reflexionen angeregt wurden. Die Prompts des Treatments „Reduktion“ (2) forderten dazu auf, sich Gedanken über die Datenreduktion (z.B. Verlust an Information zugunsten höherer Übersichtlichkeit) zu machen. Die Lernenden des Treatments „Variabilität“ (3) dagegen sollten in speziellen Prompts über statistische Variabilität im jeweiligen Kontext nachdenken. Das Treatment „Reduktion und Variabilität“ (4) enthielt Reflexionsanlässe beider Aspekte und verband so die beiden verwandten Konzepte miteinander. Durch die Aufnahme sich wiederholender Aufgabenteile speziell im Basis-Training waren die Lernzeiten aller Treatmentgruppen parallelisiert.

In den Abbildungen 4-1 bis 4-3 sind Beispielaufgaben des Kontexts „M&M-Packungen“ (vgl. auch Eichler & Vogel, 2012; Engel & Vogel, 2005) aus den verschiedenen Treatments abgedruckt. Die komplette Lernumgebung findet sich im Anhang. Alle Schüler bearbeiteten die grundlegenden Aufgaben aus Abbildung 4-1, bei denen eigene und fremde Vorhersagen bezüglich des Inhalts kleiner M&M-Packungen mit deren tatsächlichem Inhalt verglichen werden sollten sowie zur Reduktion der vorliegenden Daten durch eine Summen- und Durchschnittsbildung aufgefordert wurde.

## Abbildung 4-1

### Aufgaben aus dem Basis-Training

Marie meint:

„Anscheinend sind in einer Packung immer 24 M&Ms. Wenn ich davon ausgehe, dass von jeder der 6 Farben gleich viele M&Ms hergestellt werden, kann ich mit 4 M&Ms jeder Farbe rechnen.“

1. Mache nun eine eigene Vorhersage, wie viele M&Ms von jeder Farbe in deiner Packung enthalten sein werden (**Tabellenspalte „Meine Vorhersage“**).
2. Öffne deine Packung und notiere in der Tabelle die Anzahlen der jeweiligen Farbe.
3. Nimm dir 9 Packungs-Kärtchen und schaue dir die angegebenen Häufigkeiten auch im Vergleich zu deiner Packung an.
4. Addiere nun die Anzahlen der roten / orangen / gelben / ... M&Ms **der 9 Kärtchen sowie deiner eigenen Packung** und notiere sie in der Tabelle
5. Berechne für jede Farbe die durchschnittliche Anzahl in allen 10 Packungen und schreibe sie in die letzte Spalte der Tabelle.

Farbe	Maries Vorhersage	Meine Vorhersage	Anzahl in 1 Packung	Gesamtanzahl in allen 10 Packungen	Durchschnitt in allen 10 Packungen
Braun	4				
Grün	4				
Rot	4				
Blau	4				
Gelb	4				
Orange	4				
<b>Summe</b>	24	24	24	240	---

Ausgehend von diesen Aufgabenstellungen erhielten die Lernenden des Treatments „Variabilität“ den expliziten Auftrag, die Unterschiede zwischen den Packungsinhalten zu reflektieren, also im Kontext über Variabilität nachzudenken (siehe Abbildung 4-2).

## Abbildung 4-2

### Reflexionsaufgaben des Treatments „Variabilität“

1. Sind Maries Vorhersage, deine Vorhersage und die tatsächlichen Anzahlen identisch?
2. Wenn nein: Woran liegt das? Hast du bei der Vorhersage einen Fehler gemacht?
3. Wie könnte die Farbverteilung einer weiteren Packung aussehen?

Braun: \_\_\_\_\_, Grün: \_\_\_\_\_, Rot: \_\_\_\_\_, Blau: \_\_\_\_\_, Gelb: \_\_\_\_\_, Orange: \_\_\_\_\_

4. Die von dir berechneten durchschnittlichen Anzahlen sind nicht alle gleich. Glaubst du, dass Maries Annahme („von allen Farben werden gleich viele M&Ms hergestellt“) falsch ist?



Die Lernenden im Treatment „Reduktion“ dagegen beschäftigten sich mit der Reflexion von Vor- und Nachteilen der Datenreduktion, hier durch Berechnen von durchschnittlichen Anzahlen über mehrere M&M-Packungen hinweg (siehe Abbildung 4-3). Das Treatment „Reduktion und Variabilität“ enthielt beide in Abbildung 4-2 und 4-3 dargestellten Prompts.

#### Abbildung 4-3

##### Reflexionsaufgaben des Treatments „Reduktion“

Wenn man alle Anzahlen von M&Ms nebeneinander vor sich sieht, kann man leicht den Überblick verlieren. Darum macht es Sinn, die Gesamtanzahl und den Durchschnitt zu berechnen, denn damit erfasst man für jede M&M-Farbe einen Wert und kann besser vergleichen.

Vor allem der Durchschnitt ist ein wichtiger Wert, um sich einen Überblick zu verschaffen. Es gehen aber andererseits auch viele Informationen verloren, wenn man ihn berechnet.

1. Welche Informationen gehen hier beim Bilden des Durchschnitts verloren?
2. Welche Vorteile siehst du darin den Durchschnitt zu bilden?

Um den Einfluss des Lehrers so gering wie möglich zu halten und die parallele Beschäftigung der Lernenden mit den unterschiedlichen Treatmentmaterialien zu ermöglichen, waren diese schülerzentriert angelegt. In Partnerarbeit bearbeiteten die Schüler die Aufgaben zunächst in ihrem Tempo und verglichen anschließend selbstständig mit Musterlösungen. Bei Problemen konnten sie auf Hilfe-Karten zurückgreifen. Nach der Bearbeitung eines bestimmten Umfangs an Materialien erhielten sie persönliche Rückmeldung über die Qualität ihrer Ansätze und Ergebnisse und wurden gegebenenfalls zur Überarbeitung von Teilaufgaben aufgefordert.

## 4.2 Testinstrumente

Die Lernenden der 25 Interventionsklassen durchliefen kurz vor und nach der Intervention sowie im Abstand von circa acht Wochen verschiedene Tests. Die einzelnen Testteile werden im Folgenden näher beschrieben. Abgesehen von den Tests zur Erfassung des Arbeitsgedächtnisses, kognitiver Fähigkeiten und des Leseverständnisses bzw. der Lesegeschwindigkeit bearbeiteten die Lernenden aus 4 Baseline-Kontrollklassen in ähnlichen Abständen die gleichen Fragebögen, um als Referenzpunkt über den zeitlichen Verlauf der Intervention zu dienen.

Für eine Übersicht, welche Testteile jeweils bei den drei Messzeitpunkten eingesetzt wurden, sei auf Abbildung 4-8 verwiesen. Dieser ist außerdem der zeitliche Ablauf des Projektes sowie ein Überblick über die Teilstudien zu entnehmen. Hieraus wird auch ersichtlich, dass aufgrund der großen Fülle an Datenmaterial bisher nicht alle erhobenen Daten in den Auswer-

tungen berücksichtigt werden konnten. Insofern gibt es noch ausreichend Datenmaterial für zukünftige Forschung.

#### **4.2.1 Fachlicher Kompetenztest**

Zu allen drei Messzeitpunkten absolvierten die Schüler einen Kompetenztest zum „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ (vgl. Gundlach et al., 2010; Kuntze, 2013; Kuntze, Engel, Martignon & Gundlach, 2010; Kuntze et al., 2008a, b). Dieser enthielt 15 Items bezüglich Statistical Literacy, die sich gemäß theoretischer Überlegungen auf fünf hierarchischen Niveaustufen einordnen lassen. Die empirischen Lösungshäufigkeiten bestätigten diese Hierarchie in der vorliegenden Stichprobe. Zudem sprachen Analysen des Testinstruments mit Mplus 7.1 (Muthén & Muthén, 1998 – 2012) für eine eindimensionale Struktur (Ergebnisse der Vortests: BIC eindimensional: 4413; BIC zweidimensional: 4424; BIC dreidimensional: 4442) und die Rasch-Modellierbarkeit (BIC Rasch: 4413; BIC Birnbaum: 4464). Die in ConQuest 2.0 (Wu, Adams, Wilson, & Haldane, 2007) ermittelten Item-Fit-Werte stellten sich als angemessen heraus (gewichtete Abweichungsquadrate innerhalb der Konfidenzintervalle:  $0.97 \leq \text{MSNQ} \leq 1.04$ ; t-Werte  $< 1,0$ ). Im Gegensatz zu Vorgängerstudien kam es in dieser relativ homogenen Stichprobe allerdings aufgrund von Decken- und Bodeneffekten zu reduzierten Trennschärfen, was wohl auch die Ursache für die eher mäßige EAP/PV-Reliabilität von 0.47 (Vortest) ist.

#### **4.2.2 Test zur Erfassung von Sichtweisen auf Variabilität**

Die Studienteilnehmer wurden außerdem dazu aufgefordert, ihre Sichtweisen bezüglich zufallsbedingter Variabilität zum Ausdruck zu bringen. Dazu verlangten vier Items das Markieren einer zufälligen Verteilung von Objekten, wobei die entsprechenden Antworten nicht gemäß der Kategorien richtig / falsch gewertet, sondern als Ausdruck davon gesehen wurden, wie stark das Denken der Lernenden Erscheinungsformen von Variabilität einbezog. Im Beispielitem (siehe Abbildung 4-4) sollten die Schüler eine aus ihrer Sicht typische Verteilung von Zahnspangen einzeichnen. Die 16 linear hintereinander abgebildeten Figuren waren in Vierergruppen arrangiert, so dass die Antworten der Lernenden in drei Kategorien eingeteilt werden konnten (siehe Kuntze, 2012): Zeigten die Lernenden durch eine komplett regelmäßige Markierung keine Beachtung von Variabilität, so wurde ihre Antwort mit dem Code 0 versehen. Für die teilweise Berücksichtigung von Variabilität, beispielsweise durch unregelmäßiges Auslassen einer Figur in den einzelnen Vierergruppen, erhielten sie den Code 1. Beachteten die Schüler Variabilität über alle Teilgruppen hinweg, indem sie die Objekte ohne erkennbare Regelmäßigkeit markierten, so wurde der Code 2 vergeben. Dabei spielte es keine

Rolle, ob die Lernenden genau 12 Objekte markierten und somit das durchschnittliche Verhältnis (3 von 4 Kinder, entsprechend 12 von 16 Figuren) berücksichtigten oder ob sie davon abwichen. Diese Kategorisierung ermöglichte es, Abstufungen in den Sichtweisen der Schüler auf Variabilität zu ermitteln. Da höhere Code-Ausprägungen einer stärkeren Berücksichtigung von Variabilität entsprechen, kann der einfache Summenwert der Codes dieser vier Aufgaben als grober Indikator für die Sichtweisen der Lernenden auf Variabilität gesehen werden. Anhand von diesem Kategorisierungssystem konnten Summenwerte zwischen null und acht Punkten erzielt werden.

Abbildung 4-4

Beispiel-Item zur Erfassung von Sichtweisen bezüglich zufallsbedingter Variabilität

### Zahnspangen

Karin weiß, dass Kieferorthopäde Dr. Müller **durchschnittlich drei von vier Patienten** eine Zahnspange verpasst. Sie befragt an einem Tag vor seiner Praxis die 16 nacheinander herauskommenden Patienten.

Zeichne in die abgebildeten Smileys die Spangen so ein, wie du es in der Reihenfolge der herauskommenden Patienten für typisch halten würdest.



8.30 Uhr  17.00 Uhr

### 4.2.3 Test zur Erfassung bereichsspezifischer motivationaler Variablen

Um motivationale Variablen im Bereich von Mathematik und Statistik zu messen, wurden Multiple-Choice-Fragebögen mit fünfstufigen Likert-Skalen (1 = stimmt gar nicht ... 5 = stimmt genau) verwendet. Die mathematik-spezifischen Items zur Erfassung des akademischen Selbstkonzepts (6 Items, Beispielitem: „Mathe-Aufgaben kann ich gut lösen.“), des Interesses und der intrinsischen Motivation (insgesamt 6 Items, Beispielitem: „Für Mathematik interessiere ich mich.“) sowie der Leistungsmotivation (3 Items, Beispielitem: „In Mathematik strenge ich mich an, weil ich nicht versagen möchte.“) entstammen etablierten Skalen (Pekrun et al., 2002; 2003). Um Selbstkonzept (3 Items, Beispielitem: „Statistik-Aufgaben kann ich lösen.“), Interesse und intrinsische Motivation (insgesamt 3 Items, Beispielitem: „Ich interessiere mich für Statistik.“) auch auf Statistik bezogen zu erheben, wurden die Skalen von Gundlach und Kollegen (2010) benutzt, die einige der mathematikbezogenen Items auf den Bereich Statistik übertragen. Explorative Faktorenanalysen dieser Vorgängerforschung hatten bereits die empirische Trennbarkeit der mathematik- und statistikbezogenen Skalen gezeigt. Zusätzlich wurden den Schülern insgesamt zehn konkrete Aufgaben bezüglich statis-

tischer Darstellungen wie Tabellen, Diagramme und Kennwerten (Beispielitem siehe Abbildung 4-5) aber auch bezüglich des Ziehens von Stichproben, Variabilität und Reduktion vorgelegt, zu denen sie Angaben über ihre Überzeugung, diese richtig zu lösen und ihr Interesse (bzw. ihre intrinsische Motivation) daran machen sollten (vgl. auch Zirngibl, Götz, Pekrun, vom Hofe, & Kleine, 2005).

Abbildung 4-5

Beispielitem zur Erfassung von aufgabenspezifischem Selbstkonzept und Interesse

<b>Wie denkst du über diese Aufgabe?</b>							
Die Schülerinnen und Schüler der 8d haben im letzten Schuljahr folgende Mathenoten erreicht.							
Berechne die durchschnittliche Mathenote!	Note	1	2	3	4	5	6
	Häufigkeit	1	7	12	5	2	0

	stimmt genau	stimmt weitgehend	stimmt etwas	stimmt kaum	stimmt gar nicht
Ich bin mir sicher, dass ich diese Aufgabe korrekt bearbeiten kann.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich fände es interessant, mich mit dieser Aufgabenstellung zu beschäftigen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Diese Aufgabe zu bearbeiten würde mir Spaß machen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

#### 4.2.4 Lesegeschwindigkeits- und -verständnis

Das Leseverständnis und die Lesegeschwindigkeit der Lernenden wurde über den im Rahmen von PISA 2000 entwickelten „Lesegeschwindigkeits- und -verständnis LGVT 6-12“ (Schneider, Schlagmüller & Ennemoser, 2007) erhoben. Dabei muss innerhalb von vier Minuten möglichst viel von einem kontinuierlichen Texts gelesen werden, in dem in regelmäßigen Abständen je ein Wort fehlt. Die Schüler müssen entscheiden, welches von drei zur Auswahl gestellten Wörtern die Lücke im Gesamtkontext am sinnvollsten füllt (siehe Beispielitem in Abbildung 4-6). Für richtige Ergänzungen erhalten die Lernenden jeweils zwei Punkte, für falsche Markierungen wird ein Punkt abgezogen, woraus sich ein Summenwert für das individuelle Leseverständnis ergibt. Zusätzlich geben die Lernenden an, bis zu welcher Stelle sie gelesen haben, so dass die Anzahl der gelesenen Worte als Indikator für die Lesegeschwindigkeit dienen kann. Für beide Testteile liegen Vergleichswerte aus verschiedenen Normierungsstichproben vor. Die Retestrelabilität beträgt  $r = 0.87$  für Leseverständnis und  $r = 0.84$  für Lesegeschwindigkeit (ibid.).

Abbildung 4-6

Beispielitem des LGVT 6-12 (Schneider et al., 2007, Testheft S. 3)

Die Giraffe ist eines der größten Säugetiere auf der Welt, sie kann bis zu sechs Meter groß werden. Auf ihrem Kopf befinden sich zwei oder drei Hörner, abhängig davon, ob sie aus dem Norden oder Süden kommt. Diese [Köpfe, Flecken, Hörner] sind von einer Haut bedeckt und enden in Haarbüscheln.

#### 4.2.5 Test zur Erfassung kognitiver Fähigkeiten

Kognitive Fähigkeiten wurden über zwei Subskalen des Kognitiven Fähigkeitstests KFT 4-12+R (Heller & Perleth, 2000), einer deutschen Version des „Cognitive Abilities Test“ von Thorndike und Hagen (1971), erfasst. Für beide Testteile liegen Normwerte aus verschiedenen Altersgruppen und Schularten vor, so dass jede Subskala für sich aussagekräftig ist. Die Subskala „Figurale Analogien B“ besteht aus 25 Items im Multiple-Choice-Format, bei denen die Schüler figural dargestellte Beziehungen verstehen und auf weitere Figuren übertragen müssen. Diese Skala erfasst nonverbale kognitive Fähigkeiten, die als relativ unabhängig von kulturellen und sozialen Rahmenbedingungen gelten (Holling, Preckel & Vock, 2004). Die interne Konsistenz beträgt 0.93 (Kuder-Richardson Formel 20), die Retestrelabilität liegt ebenso bei 0.93 (Heller & Perleth, 2000). Zusätzlich wurden verbale kognitive Fähigkeiten über die Subskala „Wortanalogien B“ erhoben. Vergleichbar mit dem oben genannten nonverbalen Test muss hier die Beziehung zwischen zwei Begriffen zunächst erfasst und dann auf einen weiteren Begriff übertragen werden. Für verbale kognitive Fähigkeiten spielen sprachbezogenes Wissen und somit das kulturelle Umfeld der Lernenden eine größere Rolle (Holling et al., 2004). Insofern erfasst diese Skala eine andere Facette kognitiver Fähigkeiten als die Subskala „Figurale Analogien“. Sowohl die interne Konsistenz (Kuder-Richardson Formel 20) als auch die Retestrelabilität liegen bei 0.81 (Heller & Perleth, 2000). Beispielitems für beide Subskalen finden sich in Abbildung 4-7.

Abbildung 4-7

Beispielitems für verbale und nonverbale kognitive Fähigkeiten (Heller & Perleth, 2000; S. 114 bzw. S. 121)

groß : riesig = klein : ?				
A Junge	B erwachsen	C winzig	D wenig	E mehr

	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding: 5px;">A</td> <td style="padding: 5px;">B</td> <td style="padding: 5px;">C</td> <td style="padding: 5px;">D</td> <td style="padding: 5px;">E</td> </tr> <tr> <td style="padding: 5px;"></td> <td style="padding: 5px;"></td> <td style="padding: 5px;"></td> <td style="padding: 5px;"></td> <td style="padding: 5px;"></td> </tr> </table>	A	B	C	D	E					
A	B	C	D	E							

#### ***4.2.6 Test zur Erfassung des Arbeitsgedächtnisses***

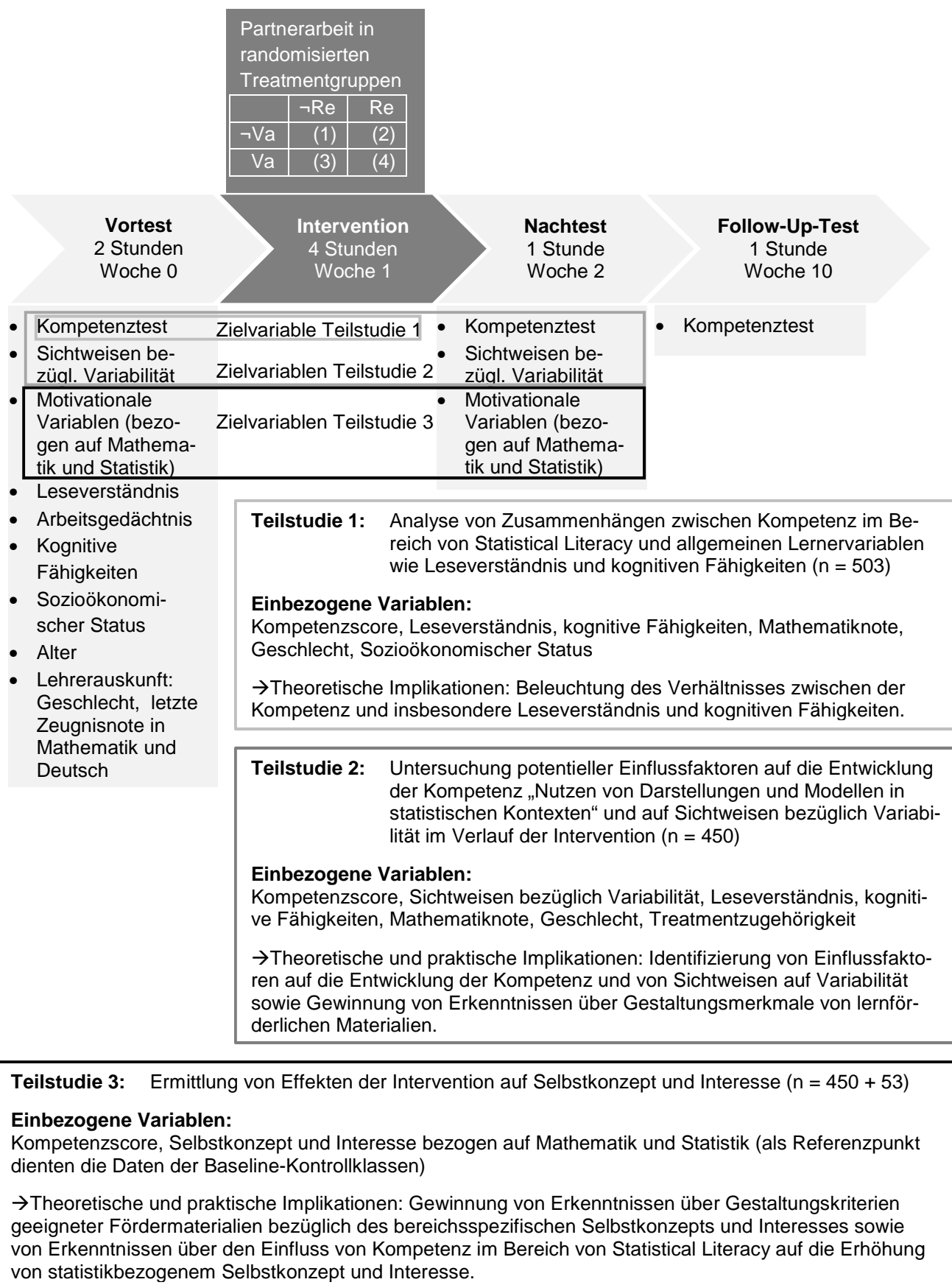
Zusätzlich zu verbalen und nonverbalen kognitiven Fähigkeiten wurde das Arbeitsgedächtnis der Schüler durch eine Adaption des „Verbalen Lern- und Merkfähigkeitstests VLMT“ (Helmstaedter, Lendt & Lux 2001) getestet. Dazu wurde den Lernenden eine Liste von 15 semantisch unabhängigen Wörtern langsam vorgelesen, von denen sie sich anschließend innerhalb von zwei Minuten an möglichst viele erinnern und diese notieren sollten. Die Anzahl der korrekt wiedergegebenen Worte kann als Indikator für die unmittelbare Gedächtnisleistung betrachtet werden, wobei aufgrund fehlender Normwerte individuelle Leistungen nur innerhalb der eigenen Stichprobe eingeordnet werden konnten.

#### ***4.2.7 Erfassung weiterer Kovariaten***

Neben den erläuterten Tests wurden zum ersten Messzeitpunkt in Interventions- und Baseline-Gruppe zusätzliche Variablen abgefragt: Die Teilnehmer machten Angaben zu ihrem Alter und zu den bei ihnen zu Hause verfügbaren Büchern. Auch wenn dies in erster Linie einen Aspekt des kulturellen Kapitals widerspiegelt, so wird dieses Merkmal dennoch auch häufig als relativ zuverlässiger Indikator für den sozioökonomischen Status genutzt (Paulus, 2009). Insofern wird diese Variable in der vorliegenden Arbeit als „sozioökonomischer Status“ bezeichnet. Das Geschlecht der Teilnehmer sowie in den Interventionsklassen die Mathematik- und Deutschnote des vergangenen siebten Schuljahres wurden anonymisiert über den jeweiligen Lehrer erhoben.

Abbildung 4-8

Überblick der eingesetzten Instrumente, des zeitlichen Ablaufs und der drei Teilstudien des Projekts



### **4.3 Übergreifende Bemerkung zu den statistischen Analysen**

Die Teilnehmer dieser Studie bilden aufgrund ihrer hierarchischen Nestung in Klassen und Schulen eine sogenannte Klumpenstichprobe, bei der nicht von unabhängigen Beobachtungen ausgegangen werden kann. Wenn sich Schüler einer Klasse bzw. Schule ähnlicher sind als unabhängige, zufällige Beobachtungen, können Standardfehler unterschätzt werden und Signifikanztests zu liberal ausfallen (Raudenbush & Bryk, 2002). Die drei Teilstudien dieses Projektes berücksichtigten die genestete Datenstruktur durch die Verwendung von geeigneten Analyseverfahren. So ermöglichte in Teilstudie 1 eine Mehrebenenregression mit HLM 7 (Raudenbush, Bryk, Cheong, Congdon & du Toit, 2011) auch den Einbezug von Prädiktoren auf der Klassenebene. Die bereits im Vortest geringen Klassenunterschiede (ICC der abhängigen Variable = 3,4%) nivellierten sich im Laufe der Intervention noch (ICC der abhängigen Variablen  $\leq 0,6\%$ ), so dass diese Auswertungsmethode für Teilstudie 2 und 3 nicht notwendig erschien. Um dennoch möglichen Verzerrungen vorzubeugen, wurden die entsprechenden Signifikanztests mit dem „type is complex“-Befehl in Mplus 7.1 (Muthén & Muthén, 1998 – 2012) durchgeführt, der die Datennestung durch Adjustierung der Standardfehler berücksichtigt.



## **5 Teilstudie 1**

### **Is statistical literacy interrelated with reading comprehension and general cognitive abilities?**

#### **A multilevel analysis**

This manuscript was submitted for publication to *Journal für Mathematikdidaktik*. It was integrated in this doctoral thesis with permission of the Editor-in-Chief Rolf Biehler.

Sproesser, U., Kuntze, S. & Engel, J. (submitted). Is statistical literacy interrelated to reading comprehension and cognitive abilities? A multilevel analysis.

#### **5.1 Abstract**

Which individual variables are connected to how statistically literate a person is? In this paper we investigate how general cognitive abilities and reading comprehension relate to the competency to use models and representations in statistical contexts, a core element of statistical literacy. We analyze data collected on 503 German 8th-graders based on a competency measure to indicate statistical literacy and standardized psychometric tests for reading comprehension and general cognitive abilities. Additionally, information provided by the teachers and students about grades in mathematics, gender and the socio-economic status are taken into account to control the examined relationship for these covariates. The results of the multilevel regression suggest moderate interdependencies between the competency construct and verbal and nonverbal cognitive abilities. The relationship between reading comprehension and the competency of using models and representations in statistical contexts turned out to be rather low, in particular when controlling for cognitive abilities. Furthermore, grades in mathematics and gender at the individual level as well as the socio-economic status at the class level were significantly related to the competency construct and hence shaped particularly the relationship between cognitive abilities and the competency construct. These findings separate the competency of using models and representations in statistical contexts from general learner variables such as cognitive abilities and reading comprehension. Knowledge about interdependencies between these learner variables and the competency in the domain of statistical literacy may help to identify students' needs concerning learning to deal with data and hence to develop adequate learning material.

*Keywords* statistical literacy, cognitive abilities, reading comprehension, multilevel analysis

## **5.2 Deutsche Zusammenfassung**

### **Hängt Statistical Literacy mit Leseverständnis und allgemeinen kognitiven Fähigkeiten zusammen?**

#### **Eine Mehrebenenanalyse**

Welche individuellen Variablen stehen in Verbindung zur Statistical Literacy einer Person? In diesem Artikel untersuchen wir wie allgemeine kognitive Fähigkeiten und das Leseverständnis in Bezug zu der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ stehen, einem Kernelement von Statistical Literacy. Wir analysieren Daten von 503 Achtklässlern aus 8 Schulen in Baden-Württemberg. Dabei benutzen wir einen Kompetenztest sowie standardisierte psychometrische Tests für das Leseverständnis und allgemeine kognitive Fähigkeiten. Außerdem werden die direkt von den Lernenden und ihren Lehrkräften erfragten Mathematiknoten, das Geschlecht und der sozio-ökonomische Status in die Analyse aufgenommen, um diese Kovariaten zu kontrollieren. Die Ergebnisse einer Mehrebenenanalyse legen Zusammenhänge zwischen dem Kompetenzkonstrukt und verbalen und nonverbalen kognitiven Fähigkeiten nahe. Die Beziehung zwischen Leseverständnis und dieser Kompetenz stellte sich als eher gering heraus, vor allem, wenn um die kognitiven Fähigkeiten kontrolliert wurde. Außerdem hingen die Mathematiknote und das Geschlecht auf der Individual- sowie der sozioökonomische Status auf der Klassenebene signifikant mit dem Kompetenzkonstrukt zusammen, wirken also insbesondere auf die Beziehung zwischen kognitiven Fähigkeiten und der Kompetenz. Die Ergebnisse grenzen die Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ von allgemeinen Lernervariablen wie kognitiven Fähigkeiten und dem Leseverständnis ab. Wissen über Zusammenhänge zwischen diesen Lernervariablen und der Kompetenz im Bereich von Statistical Literacy ermöglicht es, Bedürfnisse von Schülerinnen und Schülern bezüglich des Lernens von Statistik zu identifizieren und geeignete Lernmaterialien zu konzipieren.

*Keywords* Statistical Literacy, kognitive Fähigkeiten, Leseverständnis, Mehrebenenanalyse

### **5.3 Introduction**

Statistically literate individuals are able to make sense of data without relying on others for help. Hence they are able to autonomously pass judgments and take decisions in daily life situations. In this sense, Wallman (1993) points out that enhancing statistical literacy would lead to a substantial enrichment of our society. As statistical literacy is considered crucial to responsible citizenship in information societies, research in this domain has received growing interest over the past 30 years. However, little is known about the relationship between statistical literacy and more general cognitive abilities and reading comprehension.

Statistical literacy represents - similar to mathematical literacy (OECD, 2003) - a competency in a specific domain of expertise (Watson & Callingham, 2003). In the case of statistical literacy, the domain of expertise is related to statistical contexts. In line with Weinert (2001), we consider a competency to encompass not only specific knowledge and abilities but also the motivational, volitional and social dispositions to make use of this knowledge and to draw on these abilities in order to solve problems in the corresponding domain of expertise.

In this study, we refer to a competency that integrates core elements of statistical literacy (Kuntze, Lindmeier & Reiss, 2008a; b; see also Kuntze, 2013). We investigate the relevance of reading comprehension and general cognitive abilities for this competency. In order to get deeper insight into the relationship between the competency and these individual variables, we also include grades in mathematics, gender and the socio-economic status into the analysis to control for these learner variables that may potentially also be related to this competency. The present study contributes to clarify how strong this competency in the domain of statistical literacy relies on the learner variables taken into account. Consequently, this study helps to identify requirements of statistical literacy and hence the needs of individuals when learning to deal with data. Such knowledge may be useful to develop adequate learning material.

In the following section (4), we provide an overview of the theoretical background leading to the research questions in Section 5. Afterwards, we describe the methods and design of the study (Section 6). Results are presented in Section 7 and discussed in Section 8.

### **5.4 Theoretical Background**

#### ***5.4.1 Statistical literacy***

The increased attention to statistical literacy does not mean there is clear agreement on its definition. Several statistics educators have described statistical literacy as understanding the basic language of statistics (Garfield & delMas, 2010; Garfield et al., 2005). Others take a

broader perspective and claim that statistical literacy includes higher order abilities such as to interpret and critically evaluate reports in the media with statistical content. Accordingly, Wallman (1993) describes statistically literate individuals as being able “to understand and critically evaluate statistical results that permeate our daily lives – coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions” (1993, p. 1). Watson and colleagues (2003) have argued that statistical literacy involves three levels of increasing sophistication: basic understanding of statistical terminology; understanding of statistical language and concepts when they are embedded in the context of wider societal discussions; and the ability to apply a questioning attitude to statistical claims and arguments. According to Gal, “statistical literacy is portrayed as the ability to interpret, critically evaluate, and communicate about statistical information and messages“ (2002, p. 1). Gal’s concept of statistical literacy assumes that understanding, interpretation and reaction to statistical messages depend not only on statistical knowledge per se, but as much on the availability of literacy skills, mathematical knowledge, and general world knowledge. And it involves the ability to ask critical questions and to activate a critical stance with regard to statistical information (Gal & Murray, 2011).

Another approach does not attempt to give an overall definition but provides a list of ideas and abilities that describe statistical literacy, specified for particular target groups. In the US Guidelines for Assessment and Instruction in Statistics Education (GAISE) report for pre-K-12 school students, statistical literacy is described by a set of abilities including understanding polls, the behavior of random samples, interpreting a margin of error and understanding and question scientific findings (Franklin et al., 2007).

Beyond different nuances of various authors about what it specifically entails, statistical literacy - just like mathematical literacy (Baumert et al., 2001; Klieme, Neubrand & Lüdke, 2001; OECD, 2003) is a competency that relates to solving real-world problems. Problem solving is characterized by the fact that the learner initially does not know how to reach a solution: Applying routines or algorithms is not sufficient to solve a problem but the complexity of the problem has to be reduced by degrees until a solution can be achieved and verified in the perspective of the initial problem (Klieme, Funke, Leutner, Reimann & Wirth, 2001). Referring to mathematical contexts, (mathematical) modeling is indispensable for solving real world problems because modeling activities involve a change between the real situation and the mathematical model in order to a) simplify the real situation until the problem can be addressed by mathematical means (building up a situation model, a real model and a mathemati-

cal model) and to b) interpret the mathematical result in the light of the initial situation (see e.g. Blum & Leiss, 2005). Similarly, in statistics as the science of learning from and modeling data (Agresti & Franklin, 2013) learners have to go back and forth between the real situation and the statistical model when solving problems in a statistical context (Engel & Kuntze, 2011; Garfield, delMas & Zieffler, 2012; Lehrer, 2014). Like in any mathematical modeling, modeling activities in statistical contexts require the ability to understand, create, manipulate and investigate various representations of real world situations. Such (statistical) representations may be diagrams or tables used for example by the media or in professional and private settings. To get a deeper understanding of the data, statistical representations often have to be changed (Wild & Pfannkuch, 1999). Such a change of representations again includes the translation between distinct statistical models and their real-world meaning. Hence, dealing with representations and modeling both constitute fundamental elements of dealing with statistical data that hardly can be separated (e.g. Kuntze, 2013; Kuntze et al., 2008a, b).

In an earlier landmark paper Moore (1997) emphasized the omnipresence of variation as key characteristic of statistical data. For instance, individuals differ in their height or their cognitive abilities, but also measuring the same variable twice may not produce the same results. The origins of such differences can be systematic or nonsystematic, the latter being attributed to random (cf. Wild & Pfannkuch, 1999). Dealing adequately with statistical variation means to judge if such differences arise from systematic causes or can be attributed to random effects. Modeling variation may help to explain observed differences, to control the underlying processes and to predict further developments. Watson and Callingham (2003) highlight the relevance of dealing with variation for statistical literacy by identifying the understanding of variation as a central element of a competency model for statistical literacy.

Under the perspective of considering data analysis as an activity of extracting signals from noisy data (Konold & Pollatsek, 2002), dealing with variation often appears together with the reduction of data. In this sense, the relevant information - the *signal* - is obtained through data reduction steps while the random variation in the data - the *noise* - is omitted. Such a procedure involves trying to fit, reject or stay with a model for the data at hand. Kröpfl, Peschek and Schneider (2000) point out that meaningful overviews of data often can be obtained only through reduction steps, such as generating a simplified diagram showing e.g. percentages or calculating the mean and / or the spread. However, every reduction deletes information from the original data. Reflectively reducing data includes an awareness about the loss of information when, e.g. smoothing out random variation for instance through forming averages

(ibid.). A conscious choice of an appropriate form of data reduction therefore is more than a purely algorithmic problem. It evaluates carefully if the associated loss of information is acceptable in order to clarify the question under investigation in the corresponding context. In line with such theoretical expectations about the relevance of data reduction when dealing with statistical data, findings from a study by Reading (2002) indicate that competency differences in reducing data can be considered representative for the spectrum of what she calls *statistical understanding*.

Including also elements such as data reduction or the understanding of variation, several researchers have developed test instruments in order to assess learners' competency of dealing with statistical data. Most of these tests are used in post-secondary education and they do not focus particularly on statistical literacy as described above (e.g. delMas, Garfield, Ooms & Chance, 2007; Garfield et al., 2012; Ziegler, 2014; ...). In contrast to these tests at the tertiary level, the test instrument for statistical literacy conceived by Watson and Callingham (2003) was developed for students from grade 3 to grade 9. However, a certain drawback of their approach is the fact that the scores are based on a coding procedure which makes twice use of some of the students' answers and hence several items might not be independent from each other. Whereas the aspect of variation is dominant in Watson and Callingham's approach Kuntze and colleagues (2008a, b; Kuntze, 2013) complement this aspect with further relevant elements of statistical literacy. In the following, we introduce this approach based on the ideas described above to determine hierarchical levels of competency in the domain of statistical literacy.

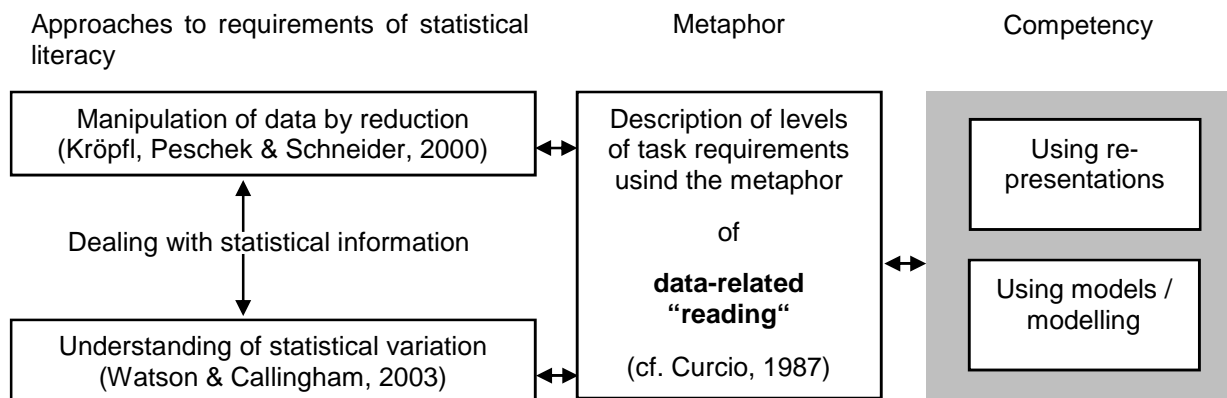
#### ***5.4.2 Statistical Literacy Competency Construct (SLCC) - a competency construct for using models and representations in statistical contexts***

As outlined in Section 4.1, data reduction as well as dealing with variation can be considered as central ingredients of statistical literacy. Both are often involved when making sense of statistical information, hence Kuntze and colleagues (e.g. 2008a, b; 2010; see also Kuntze, 2013) linked them with the metaphor of data-related reading. The idea of data-related reading is based on Curcio's (1987) model describing learner's understanding of statistical graphs along three hierarchical levels: reading the data (e.g. directly extracting information from a statistical representation), reading between the data (e.g. integrating several aspects of a given data representation) and reading beyond the data (e.g. making predictions of new data on the basis of the available data). Kuntze et al. (2008a, b; 2010) enlarged Curcio's model by observing that reading statistical information represented in tables, graphs or numerical summaries

usually involves the use of statistical models. Based on this broader concept of data-related reading, Kuntze and colleagues established the competency model of “using models and representations in statistical contexts”. The emphasis on the use of statistical representations and models also links this competency model to the German standards that include “using representations” and “modeling” as comprehensive competencies (KMK, 2003). Figure 5-1 displays an overview of the theoretical conceptualization of Kuntze et al.’s competency construct.

Figure 5-1

Elements of the competency “using models and representations in statistical contexts” (cf. Kuntze et al., 2008b, p. 2)



This competency model in the domain of statistical literacy comprises five hierarchical levels that primarily can be distinguished by the number of steps needed for solving the underlying problem (Kuntze et al., 2010). In the following, we describe the requirements of the competency levels and as illustration discuss several students’ answers to tasks of a corresponding test instrument. Further examples of tasks can be found in Kuntze et al. (2008a).

At the lowest competency level (I), only one-step activities of using statistical models and representations are required, such as directly extracting a given value from a diagram. This corresponds to the elementary use of a given model, as – for instance – a graphical element has to be linked to the scale given in a diagram in order to read a quantity from the diagram. Requirements at Level II involve a two- or multi-step use of representations, which is often required when e.g. having to change between two representations, or a change between two given models. An item at this level may for instance request to select and sum up several values from a diagram or to compare data from two diagrams.

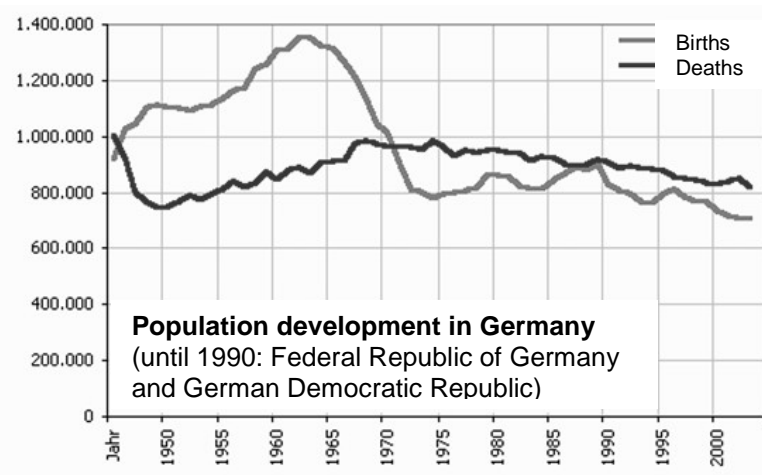
Whereas at the two lowest competency levels no own modeling activities are required, Level III demands a multi-step use of representations including active modeling steps, i.e. the use of a non-given model. Figure 5-2 displays exemplarily the answer of a student to a Level III task. In order to identify the year corresponding to the intersection of the curves, the answer appears to be based on a subtraction model integrating aspects of both curves: The current population change appears as the difference of births and deaths, hence the population decreases in years when more people die than babies are born. Of course, alternative non-given models can also lead to appropriate answers here, e.g. modeling the population as a “reservoir” with positive “incomings” (births) and negative “outgoings” (deaths).

Figure 5-2

Student’s answer to an item at Level III (see e.g. Lindmeier, Kuntze & Reiss, 2007, p. 4)

“The Germans don’t have enough children“, was a recent newspaper’s headline.

Consider the following diagram about the development of the German population:



From which year on has the population been decreasing?

Ca ab dem Jahre 1970, da sind weniger Menschen geboren, als gestorben.

Approximately from the year 1970 onwards because that time less people were born than have died.

On the basis of a multi-step use of statistical representations involving also own modeling activities, Level IV tasks focus additionally on dealing with variation. As an example, the task in Figure 5-3 asks to compare two distributions of recovery times for headache tablets which differ in mean and variance: Here, answering this question requires considering statistical variation in the modeling process in order to give suitable arguments. The use of sophisticated models such as spread or variance may lead to appropriate answers, but, alternatively, also basic models such as comparing frequencies in different time intervals. In any case, this task

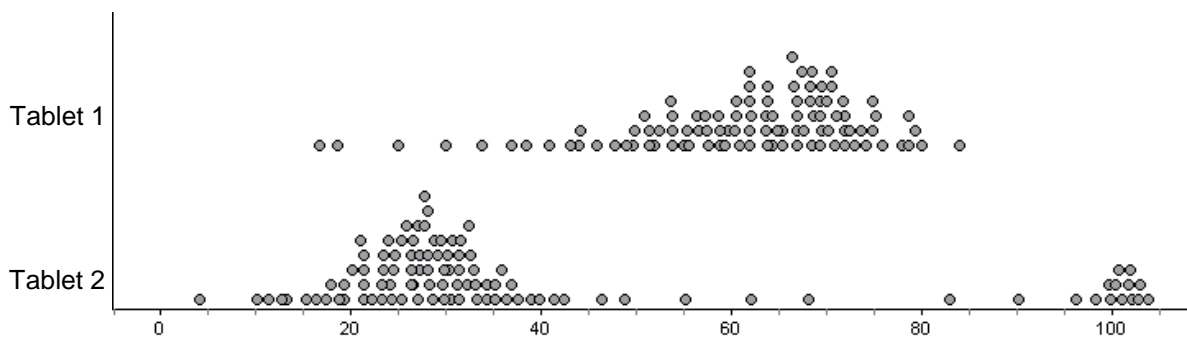


requires own modeling steps in order to deal appropriately with variation. The illustrated answer of a student appears to be based on a rather basic model: observing distinct time intervals may lead to the recognition that the recovery times for tablet 2 vary among the tested people. Hence, this answer indicates a distinction between two groups of people using tablet 2: the majority was cured within 40 minutes but a few people suffered longer from headache than all the people using tablet 1.

Figure 5-3

Student's answer to an item at Level IV (see also Kuntze et al., 2008b, p. 4)

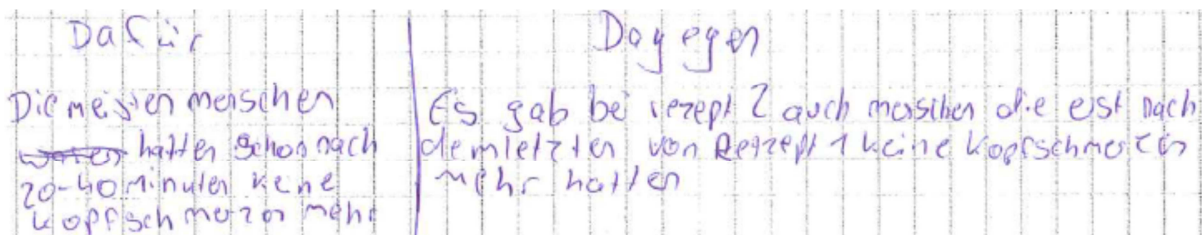
A company produces two sorts of headache tablets. Both sorts have been tested in a laboratory with respectively 100 persons suffering from headache. The diagram below shows, how long it took until the headache was over. Each point represents one test person.



Dr. Green says:

Tablet 2 is better than tablet 1!

Find an argument for and an argument against this statement!



For

Most people had no more headache after 20-40 minutes.

Against

There were also people using tablet 2 whose headache disappeared after the latest person using tablet 1.

The highest competency level (V) particularly requires the mental construction of (non-given) data that suit the context information. Based on a multi-step use of representations, an adequate answer calls again for own modeling activities. Here data reduction steps have to be inverted taking statistical variation into account. In the example displayed in Figure 5-4, students need to mentally generate and organize data points in accordance to a given average.

The student's answer illustrated in Figure 5-4 below may be based on a model of the average as a balance of upward and downward deviations. This answer also takes into account that there are less upward than downwards deviations. Additionally, the distribution respects statistical variation concerning the length and the order of the fish.

Figure 5-4

Student's answer to an item at Level V (Kuntze, 2013, p. 80)

A fisherman catches fish. He remarks that deviations above the average length (oversized fish) happen more rarely than deviations below the average (small-sized fish).

Give a typical distribution of fish lengths for an average length of 30cm that conforms with this description by completing the following diagram:

Length of fish  
(approximately)



The hierarchical structure of this competency has been verified empirically in various samples through Rasch analyses of the corresponding test instrument (Fröhlich, Kuntze & Lindmeier, 2007; Kuntze, Engel, Martignon & Gundlach, 2010; Kuntze et al., 2008a, b). In the present study, we used this test instrument to assess students' competencies in the domain of statistical literacy. To facilitate readability, in the following we refer to the competency of *using models and representations in statistical contexts*, which we consider as a core competency in statistical literacy, under the label of SLCC (Statistical Literacy Competency Construct).

Despite obvious overlap between SLCC and the overarching idea "data and chance" of the German standards (KMK, 2003), they cannot be considered as congruent. "Data and chance" represents a curricular category that summarizes contents, knowledge elements and abilities related to dealing with statistics (similar to the US report on Guidelines for Assessment and Instruction in Statistics Education, (Franklin et al., 2007)) whereas SLCC describes different levels of a specific competency in the domain of statistical literacy. Furthermore, SLCC and "data and chance" also differ concerning relevant elements: For instance probability, an important aspect of "data and chance", is not directly addressed in SLCC. On the other hand, dealing with variation, a central element of SLCC, is not explicitly listed in the "data and

chance” curriculum. Similar considerations apply to differences between SLCC and the PISA overarching idea “uncertainty” (cf. Blum et al., 2004). Insofar, findings from research focusing on “data and chance” or “uncertainty” cannot be directly transferred to SLCC. Hence our study, focusing on the relationship between SLCC and reading comprehension as well as general cognitive abilities, is not a replication of parts of PISA but provides new empirical evidence about how SLCC as a core competency in the domain of statistical literacy relates to certain learner variables.

### ***5.4.3 Theoretical considerations and empirical findings concerning variables potentially related to SLCC***

The domain-specificity of a competency such as SLCC implies that corresponding tests do not simply measure general learner variables that are unspecific to the underlying domain of expertise. For instance, reading comprehension and cognitive abilities should not determine SLCC. Otherwise, the competency construct would not be separable from such general variables, rendering the construct obsolete. This article focuses on clarifying the relationship between SLCC and reading comprehension as well as cognitive abilities. In this section, we present theoretical considerations and review empirical results from related studies that investigated the association between reading comprehension, general cognitive abilities and competencies potentially related to SLCC such as mathematical literacy. Moreover, we discuss arguments for including additional context variables as control such as mathematical knowledge, gender and the socio-economic status that may be interwoven into the examined relationship.

Reading is considered to be a basic prerequisite to acquire knowledge and abilities in various domains (Kamil, Mosenthal, Pearson & Barr, 2000). In this sense, Klieme, Neubrand et al. (2001) explain the significant relationship between reading competency and mathematical literacy in PISA 2000 (ibid, p. 184: path coefficient  $\beta = .55$ ,  $p < .05$ ) by the interdisciplinary character of reading competency that is also required when solving complex modeling tasks. Referring to our field of research, Gal (2002) emphasizes the importance of literacy skills such as reading for statistical literacy because statistics-related problems are often given in written formats. This may also apply for the format of the test instrument used in this study: High demands in reading comprehension may have an influence on the complexity of the written tasks and therefore a lack of reading comprehension could prevent students from succeeding in the SLCC test even if they are able to solve the tasks from the point of view of using statistical models and representations. Furthermore, data-related reading is a basic com-

ponent of SLCC and it might be expected that the processes behind data-related and text-oriented reading can be similar. For instance, several information units (of sentences or of statistical representations) have to be recognized as important, embedded in their context and interpreted in relation to others. As such process similarities between data-related reading and text-oriented reading comprehension are quite possible, we assessed reading comprehension with as little content overlap as possible and hence – in contrast to PISA (Baumert et al., 2001) - left out reading from statistical representations such as tables and diagrams. Instead, we approach reading comprehension from a purely text-oriented perspective (e.g. Schneider, Schlagmüller & Ennemoser, 2007) that allows researchers to investigate if the wording of the tasks is appropriate for learners or if it hinders their success.

Another variable that could be related to students' SLCC are cognitive abilities because cognitive processes are the basis of solving mathematical and statistical problems. Klieme, Neubrand et al. (2001) for instance emphasize the relevance of cognitive abilities such as understanding abstract symbolic systems and connecting numerical, figural and verbal information for mathematical literacy. These considerations may explain the significant relationship between general cognitive abilities and mathematical literacy found e.g. in PISA 2000 (ibid, p. 184: path coefficient  $\beta = .32$ ,  $p < .05$ ). Additionally to this direct relation, cognitive abilities were found in PISA to be mediated by reading competency. In this study, we investigate whether there is a similar relationship between SLCC and general cognitive abilities.

Further context variables potentially related to SLCC may shape the relationship between SLCC and reading comprehension as well as cognitive abilities. Therefore, in this study we include grades in mathematics, gender and the socio-economic status into the analysis in order to control for these covariates. In the following we present considerations suggesting an inter-relatedness between these variables and SLCC.

As mathematical concepts and procedures (e.g. average, percent, proportions) are often required when dealing with statistical problems, Gal (2002) points out that knowledge in the domain of mathematics is an important prerequisite for statistical literacy. Tent reports relatively high correlations between school grades and student assessments referring to the same domain (2001, S. 808:  $-.60 \leq r \leq -.70$ ). Brunner (2005) found a significant relationship between grades in mathematics and specific mathematical abilities beyond general cognitive abilities (see also Brunner, 2008). Hence, we consider grades in mathematics as adequate to indicate students' knowledge in the domain of mathematics and include this variable potentially linked to students' SLCC into the analysis.

Concerning the domain of mathematics, gender differences in favor of boys are reported by a vast number of studies. For instance Klieme (1997) found in a meta-analysis gender differences of an average effect size of  $d = .38$  (Cohen, 1988). According to his results, these differences are particularly large in mathematical problem solving. Other studies such as PISA (e.g. Brunner, 2005; Klieme, Neubrand et al., 2001) and TIMSS (e.g. Hosenfeld, Köller & Baumert, 1999) confirm such findings concerning gender differences from grade seven to grade nine. As statistics is by and large taught as part of mathematics education, gender differences are also possible for SLCC. Furthermore, Martignon (2010) states that girls are less open-minded with respect to statistical contents than boys. Hence, gender is included as a context variable in this study.

In several studies, the socio-economic status has shown to be related to students' achievement at the individual and at the class level. Concerning the individual level, for instance Sirin (2005) found mean correlations of  $r = .35$  between SES and mathematical ability. A significant relationship between these variables was also discovered in PISA (e.g. Ehmke, Hohensee, Siegle & Prenzel, 2006; Klieme, Neubrand et al., 2001; Watermann & Baumert, 2006). Bradley and Corwyn (2002) mention for instance a better access to social resources and higher educational aspirations for families with a high socio-economic status as possible explanations of this association. Other studies (e.g. Opdenakker & Van Damme, 2001; Stanat, Schwippert & Gröhlich, 2010; Thrupp, Lauder & Robinson, 2002) showed that the class average of socio-economic status is related to students' achievement. Willms (1992) claims that a rather high average of socio-economic status at the school or class level can provide advantages for students' achievement for instance because of "fewer disciplinary problems and an atmosphere conducive to learning" (p. 41). Opdenakker and Van Damme (2001) attribute the relationship between the average of socio-economic status and individual achievement to a significant correlation between mean ability and mean socio-economic status. As the mentioned considerations may also apply for SLCC, this study takes the socio-economic status at the individual and at the class level into account.

### **5.5 Research questions**

As highlighted above, there is substantial interest in examining connections of SLCC with reading comprehension and general cognitive abilities. Empirical insight is needed to be able to describe how specific this competency construct is for the corresponding domain of expertise as opposed to general aspects, such as cognitive abilities and reading comprehension. In

the area of statistical literacy and in the case of SLCC in particular, such empirical evidence is scarce. Hence, this study investigates the following research questions:

- 1) To which extent are reading comprehension and SLCC interdependent?
- 2) To which extent are general cognitive abilities and SLCC interconnected?

In order to gain a more complete picture about the relationship between SLCC and reading comprehension as well as cognitive abilities, we also consider further context variables. More particular, we include grades in mathematics (individual level), gender (individual level) and the socio-economic status (individual and class level) into the analysis. Hence, in this study we further examine the following research question:

- 3) How do grades in mathematics, gender and the socio-economic status affect the relationship between SLCC and reading comprehension as well as general cognitive abilities?

## **5.6 Methods**

### ***5.6.1 Sample and context of the study***

In this study, data from 503 German 8<sup>th</sup>-graders from technical-track secondary schools (“Realschule”; this type of school is for students with an average level of academic performance) was analyzed. The 263 boys and 240 girls aged between 12 and 15 (M 13.48, SD 0.62) were recruited from 25 classes from 8 schools in large cities but also rural areas across the state of Baden-Württemberg.

A paper-and-pencil test described in detail in the next section was administered to the participants during two consecutive periods of ordinary 45-minute classes. During the whole testing period, the regular teacher and the first author were present to monitor the implementation of the test and to document the classroom processes in a logbook.

### ***5.6.2 Measures and covariates***

#### *Test instrument for statistical literacy*

Participants completed a written competency test with tasks related to statistical literacy (e.g. Kuntze, 2013; Kuntze et al., 2010; Kuntze et al., 2008 a, b). According to the competency model (see Section 4.2 for sample items and a description of the competency model), this test comprises 15 items on five hierarchical levels. This hierarchy could also be observed in the

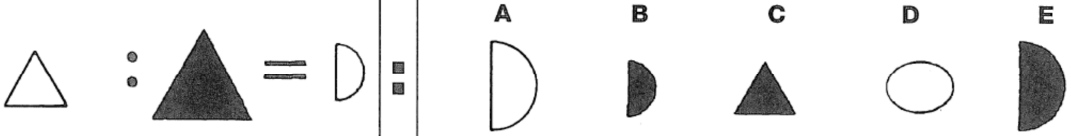
frequency of the empirical solutions within our sample. Analyses<sup>3</sup> and model comparisons in Mplus 7.1 (Muthén & Muthén, 1998 – 2012) referring to our sample confirmed prior research (e.g. Fröhlich et al., 2007; Kuntze et al., 2010; Lindmeier et al., 2007) indicating that the SLCC scores fit to a one-dimensional Rasch model (BIC Rasch: 4413, BIC Birnbaum: 4464; BIC one-dimensional: 4413, BIC two-dimensional: 4424, BIC three-dimensional: 4442). Further analyses in ConQuest 2.0 (Wu, Adams, Wilson, & Haldane, 2007) revealed adequate item fit indices (weighted mean squares within the confidence intervals ( $0.97 \leq \text{MSNQ} \leq 1.04$ ),  $t$ -values  $< 1.0$ ). Probably due to ceiling and bottom effects within our relatively homogeneous sample, EAP/PV reliability was rather moderate (0.47).

*Test instrument referring to general cognitive abilities*

To tap students’ nonverbal cognitive abilities we used the Figure Analogies subscale (N2B) from the Cognitive Ability Test 4-12+R (Heller & Perleth, 2000), a German version of the Cognitive Abilities Test by Thorndike and Hagen (1971). This scale comprises of 25 figural items in multiple-choice format. In order to tap verbal cognitive abilities, we consulted the Word Analogies test (V3B), also a subscale of the Cognitive Ability Test 4-12+R, consisting of 20 multiple-choice items. The internal consistency (Kuder-Richardson formula 20) was 0.93 for the nonverbal and 0.81 for the verbal subscale (Heller & Perleth, 2000, p. 19). Whereas the Figural Analogies subscale taps a highly  $g$ -loaded ability component, the Word Analogies subscale includes also language-related, hence culture-bound knowledge (Holling, Preckel & Volck, 2004). Consulting two different components of cognitive abilities enables to examine the relationship between SLCC and cognitive abilities in a relatively broad approach. Table 5-1 presents two sample items.

Table 5-1

Examples of items included in the cognitive abilities subscales used in this study (translated by the authors)

Sub-test	Example (Heller & Perleth, 2000, p. 121; p. 114)
N2B	

<sup>3</sup> As two tasks at the highest level of competency were only rarely solved by the students, we excluded them from the analysis of the test instrument.

V3B	tall : huge = small : ?				
	A boy	B adult	C tiny	D little	E more

*Test instrument referring to reading comprehension*

To tap students' reading comprehension, we used the LGVT 6-12 test (Schneider et al., 2007) that was developed as part of the PISA 2000 German extension test. Its basis is a cloze test, that is, a continuous text with several gaps which students have to fill with the most suitable word from a list of three. The total reading comprehension score is calculated by two points for each correct answer and one negative point for each incorrect answer. The example in Figure 5-5 reveals the structure of the items.

Figure 5-5

Example of an LGVT item (cf. Schneider et al., 2007, translation by Leiss, Schukajlow, Blum, Messner & Pekrun, 2010, p. 127)

The giraffe is one of the largest mammals in the world. It can be up to six meters tall. It has two or three horns on its head depending on whether it is from the north or the south. These [heads, spots, horns] are covered with a skin and end with a tuft.

As for KFT, precise instructions are provided to ensure the implementation and therefore the objectivity of its results. According to Schneider and colleagues (2007, p. 7) retest reliability amounts to  $r = .87$ .

In addition to the variables described above, data referring to gender and grades in mathematics was collected directly from the teachers involved in this study. Additionally, students were asked about the number of books they have at home, taken as an indicator for the socio-economic status of their family (cf. Paulus, 2009). In the following we refer to this variable by the label socio-economic status even if it rather represents a component of objectified cultural capital (Bourdieu, 1986).

### **5.6.3 Statistical Analyses**

As data involving students from different school classes tends to be clustered according to the context of the classrooms, the study was designed to allow for multilevel regression (see e.g. Ditton, 1998). Ordinary multiple regression presupposes cases to be independent, an assumption that may well be violated when data comes from students grouped into classes (Hox, 2010). Disregarding the structure of multilevel data could lead to inaccurate parameter esti-



mates as well as inappropriate conclusions regarding the level of interpretation (Eid, Gollwitzer & Schmitt, 2011).

Checking the data of this study for assumptions required by multilevel analysis, scatterplots and residual analyses did not indicate violations of linearity and homoscedasticity. The assumption of normal distribution could not be met for the variables nonverbal cognitive abilities or reading comprehension. Suitable logarithmic transformations were applied to get them closer to normal. However, as the untransformed data led to very similar results, and for ease of interpretation, in the following we report the results based on the original data only.

Due to missing values, 63 cases were excluded prior to the main analysis. As a control, the analyses were replicated with these missing data replaced by imputed values. Since both analyses produced similar results, we refer in the following to the dataset without imputations (n=503).

An important parameter in multilevel regression is the intraclass-correlation (ICC), which can be interpreted as a measure of class homogeneity. For our data we computed an ICC of 3.4% (p=0.009), which means that there were moderate but statistically significant differences concerning SLCC among the classes. These differences in the competency score were considered by allowing random intercepts in our regression model, that is, an overall intercept (see  $\gamma_{00}$  in Equation 5-1) for the SLCC score was supplemented by discrepancies (see  $u_{0j}$  in Equation 5-1) for each class. Variance of SLCC was hence divided into a class component (see  $u_{0j}$  in Equation 5-1) and an individual component (see  $r_{ij}$  in Equation 5-1). We found no random slopes for any of the explanatory variables, indicating that the relationships between SLCC and the predictors did not significantly differ from one class to another (Hox, 2010). Hence, simple multilevel equations such as displayed in Equation 5-1 were adequate to model the data.

#### Equation 5-1

Example of a simple multilevel equation

$$\text{SLCC\_Score} = \gamma_{00} + \gamma_{01} \cdot \text{class\_variable} + \gamma_{10} \cdot \text{individual\_variable} + \dots + u_{0j} + r_{ij}$$

Concerning the explanatory variables the ICCs were rather low (0.2% – 3.8%) except for the socio-economic status (6.3%) indicating that the classes involved in this study were relatively similar with respect to most of the observed variables. The differences in the socio-economic

status may be due to the fact that schools both from rural areas and from large cities were included in this study.

We based the multilevel analyses on the HLM statistics package with full maximum likelihood estimation that allows chi-square tests to compare two models (Hox, 2010). The reported results were estimated with robust standard errors. All variables were z-transformed to enable direct comparisons between the regression coefficients. Grades in mathematics were centered at the group mean to account for potential differences in teachers' gradings (see Enders & Tofighi, 2007 for further information). The other variables were centered at the grand mean.

## 5.7 Results

### 5.7.1 Correlations between the consulted variables

Table 5-2 shows bivariate correlations between the variables that were included into the analysis. Except for gender, all of the explanatory variables correlated significantly with the competency score. However, we found differences referring to the magnitude of these relationships. Most of the explanatory variables showed also significant intercorrelations. In the following analysis steps, we introduce these explanatory variables successively into multiple regression models to predict SLCC.

Table 5-2

Bivariate correlations between the consulted variables

	SLCC	RC	VCA	NCA	MG	SES
RC	.16 ***					
VCA	.35 ***	.16 ***				
NCA	.33 ***	.16 ***	.41 ***			
MG	-.33 ***	-.08	-.29 ***	-.27 ***		
SES	.19 ***	.14 **	.16 ***	.08	-.22 ***	
Gender	-.09	.08	.05	.07	.02	-.09

**Note:** RC ... Reading comprehension; VCA ... Verbal cognitive abilities; NCA ... Nonverbal cognitive abilities; MG ... Grades in mathematics; SES ... Socio-economic status; Gender: 1 ... female, 0 ... male)

\*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05

### 5.7.2 *The relationship between SLCC and reading comprehension as well as cognitive abilities*

According to research question 1, we focus first on potential interdependencies between reading comprehension and SLCC. As suggested from the correlation table, reading comprehension turned out to be a significant predictor for SLCC at the individual level (Level 1). Table 5-3 shows the precise parameter estimates of Model 1 which assumes a linear relationship between SLCC and reading comprehension only: The regression coefficient of 0.15, implies that the increase of one standard deviation of the explanatory variable reading comprehension leads to a raise of 0.15 standard deviations for the dependent variable SLCC. Pseudo-R<sup>2</sup> amounts to 2.5% of explained variance for the explanatory variable reading comprehension. Deviance indicating the quality of model fit (less deviance means a better model fit) is given by 1409.2. A highly significant chi-squared test on the difference of deviance between the intercept-only model without any predictor (Deviance  $\approx$  1420.6) and this model (degrees of freedom = 1 for the number of added parameters) underlines that reading comprehension is a significant predictor for SLCC.

Table 5-3

Model 1 including reading comprehension

Fixed effect	Coefficient	SE	p-value	Deviance	Pseudo-R <sup>2</sup>
Reading comprehension	$\approx$ 0.15	$\approx$ 0.04	<0.001	$\approx$ 1409.2	$\approx$ 2.5%

Next, we extended Model 1 by introducing cognitive abilities as another linear covariate. (Model 2, see Table 5-4). Both verbal and nonverbal cognitive abilities turned out to be significant predictors at a high level of significance. Comparing the regression coefficients of cognitive abilities to the regression coefficient of reading comprehension suggests that the relationship between SLCC and cognitive abilities is closer than between SLCC and reading comprehension. The inclusion of cognitive abilities leads to a significantly better model fit (Deviance  $\approx$  1329.0) and to about 17.0% of explained variance.

Table 5-4

Model 2 including reading comprehension, verbal and nonverbal cognitive abilities

Fixed effect	Coefficient	SE	p-value	Deviance	Pseudo-R <sup>2</sup>
Reading comprehension	≈0.08	≈0.04	0.027	≈1329.0	≈17.0%
Verbal cognitive abilities	≈0.26	≈0.04	<0.001		
Nonverbal cognitive abilities	≈0.20	≈0.04	<0.001		

As expected the inclusion of cognitive abilities reduced the regression coefficient of reading comprehension because these explanatory variables correlate significantly (see Table 5-2). However, Model 2 suggests that reading comprehension does not explain much variance of SLCC beyond cognitive abilities. Dropping reading comprehension from Model 2, i.e. considering SLCC with verbal and nonverbal cognitive abilities as only linear predictors (Model 3, see Table 5-5) leads to a similar model fit and explication of variance as Model 2. The regression parameters for cognitive abilities were almost the same in Model 3 as in Model 2. We compared model fit of Model 3 and Model 2 by means of a chi-square test performed on the deviances with degrees of freedom equal to the number of added parameters (Hox, 2010) resulting in  $\chi^2 = 3.6$ ,  $df = 1$ . This value is not significant at the 5% level indicating that Model 3 is not worse than Model 2. Hence, we prefer Model 3 for reasons of parsimony and exclude reading comprehension from the following analysis steps. Comparisons of the multilevel models presented in the following and the same models supplemented by reading comprehension also confirm that the more parsimonious models without reading comprehension fit the data better.

Table 5-5

Model 3 including verbal and nonverbal cognitive abilities

Fixed effect	Coefficient	SE	p-value	Deviance	Pseudo-R <sup>2</sup>
Verbal cognitive abilities	≈0.27	≈0.04	<0.001	≈1332.6	≈16.3%
Nonverbal cognitive abilities	≈0.21	≈0.04	<0.001		

### 5.7.3 *The relationship between SLCC and cognitive abilities controlled for grades in mathematics, gender and the socio-economic status*

In order to get a more accurate view of the relationship between SLCC and cognitive abilities, we introduced further context variables into Model 3. Additionally including students' grades in mathematics (Model 4) as an indicator for mathematical knowledge reduces the regression coefficients of verbal and nonverbal cognitive abilities slightly. Concerning its regression coefficient and p-value, the effect of grades in mathematics has a similar size as the effect of verbal and non-verbal cognitive abilities (see Table 5-6 for the precise parameter estimates). The negative algebraic sign of the regression coefficient of grades in mathematics is caused by the fact that the German grading system labels very good performances with 1 and insufficient performances with 6. Deviance decreased by 28.0 points from Model 3 to Model 4, indicating a significant improvement in model fit. The inclusion of grades in mathematics into the model not only reduces the relevance of cognitive abilities slightly but also contributes to further variance explanation: Pseudo-R<sup>2</sup> provides 20.7% of variance explanation by Model 4, a substantial increase compared to 16.3% by Model 3.

Table 5-6

Model 4 including verbal and nonverbal cognitive abilities as well as grades in mathematics

Fixed effect	Coefficient	SE	p-value	Deviance	Pseudo-R <sup>2</sup>
Verbal cognitive abilities	≈0.22	≈0.04	<0.001	≈1304.6	≈20.7%
Nonverbal cognitive abilities	≈0.17	≈0.04	<0.001		
Grades in mathematics	≈-0.23	≈0.05	<0.001		

In the next step, we added gender as a further covariate into Model 4 (see Model 5 in Table 5-7). As suggested by the non-significant correlations in Table 2, the inclusion of gender did not reduce the regression weights of the other variables and in particular did not affect the relationship between SLCC and cognitive abilities. However, gender appears to be a significant predictor for SLCC. The negative algebraic sign of the regression coefficient indicates that boys achieved on average higher SLCC scores than girls. Model 5 shows a significantly improved model fit and about 21.7% of variance explanation.

Table 5-7

Model 5 including cognitive abilities, grades in mathematics and gender

Fixed effect	Coefficient	SE	p-value	Deviance	Pseudo-R <sup>2</sup>
Verbal cognitive abilities	≈0.22	≈0.04	<0.001	≈1297.3	≈21.7%
Nonverbal cognitive abilities	≈0.17	≈0.04	<0.001		
Grades in mathematics	≈-0.23	≈0.05	<0.001		
Gender (0 ... ♂; 1 ... ♀)	≈-0.21	≈0.08	0.005		

The socio-economic status turned out to be non-significant at the individual level (parameters for the socio-economic status added to Model 5:  $\beta \approx 0.08$ ,  $SE \approx 0.04$ ;  $p \approx 0.060$ ). Hence, the socio-economic status does neither significantly shape the relationship between SLCC and cognitive abilities nor explain variance beyond the already included explanatory variables at the individual level.

However, the socio-economic status was a significant predictor at the class level (see Table 5-8 displaying Model 6). The regression coefficients of cognitive abilities remained almost the same as in Model 5 indicating that the socio-economic status did hardly affect their relationship to SLCC. As a relevant variable for SLCC, the class average of the socio-economic status was adequate for explaining the initial differences between the classes to such a degree (85.8%) that the remaining inter-class differences were no longer significant. A chi-squared test performed on the reduction of deviance indicates a significant improvement of model fit. The overall variance explanation increased to the value of  $Pseudo-R^2 = 24.2\%$  for Model 6. This final model includes beyond the significant predictors cognitive abilities, grades in mathematics, gender and the average of socio-economic status also the non-significant socio-economic status at the individual level to ensure that this class effect would be adjusted for the individual effect.

Table 5-8

Model 6 including cognitive abilities, grades in mathematics, gender and the socio-economic status

Fixed effect	Coefficient	SE	p-value	Deviance	Pseudo-R <sup>2</sup>
Verbal cognitive abilities	≈0.21	≈0.04	<0.001	≈1285.1	≈24.2%
Nonverbal cognitive abilities	≈0.17	≈0.04	<0.001		
Grades in mathematics	≈-0.22	≈0.05	<0.001		
Gender (0 ... ♂; 1 ... ♀)	≈-0.21	≈0.07	0.005		
Socio-economic status	≈0.04	≈0.04	0.330		
∅ Socio-economic status	≈0.42	≈0.12	0.003		

## 5.8 Discussion and conclusion

In this section, we summarize results and discuss implications for theory and for instructional practice. Our findings should be interpreted with care, given that the sample is not representative for all German students. However, the research questions could be answered, as the analyses reported in Section 5.6.3 indicate that the statistical prerequisites necessary to conducting quantitative analyses were met. We discuss in the following our findings in the order of the research questions posed and then turn to implications for theory and for instructional practice.

The first research question focused on interdependencies between SLCC and reading comprehension. The correlation between these variables turned out to be rather moderate. As including general cognitive abilities into the model to predict SLCC by reading comprehension cut its regression coefficient almost in half, a relatively large part of the relationship between SLCC and reading comprehension can be attributed to cognitive abilities. Furthermore, the rather low association between SLCC and reading comprehension may be due to our measure for reading comprehension that explicitly excluded reading from statistical formats such as tables and diagrams. This finding also suggests that a low level of reading comprehension did not necessarily prevent students from reaching high scores on the SLCC test. This indicates further that the wording of the SLCC test seems to be appropriate for the participants of the study.

Concerning research question 2, we found for both verbal and nonverbal cognitive abilities a significant relationship to SLCC with greater emphasis on verbal abilities. This may be because some of the tasks on the SLCC test were presented in a verbal format and had to be completed using written words or sentences. Even though an association between verbal and nonverbal cognitive abilities and SLCC was discovered, this relationship is rather moderate. These findings confirm SLCC is a domain-specific construct in its own right and does not measure only learners' general cognitive abilities.

Within the scope of research question 1 and 2, these findings suggest several implications on a theoretical level: Even though the reading comprehension test LGVT was conceived in the context of the PISA study, there is a significant difference concerning content and representation formats between reading comprehension on the LGVT and reading competency on the PISA tests. Our results reflect this difference empirically by showing a rather weak relationship between SLCC and text-oriented reading comprehension as measured by the LGVT test: Compared to corresponding findings from the PISA study (Baumert, 2001; OECD, 2003), the findings from our study show less interrelatedness between the SLCC score and reading comprehension as well as general cognitive abilities. We offer two possible explanations: First, the high correlations between mathematical competency and reading competency as well as between mathematical competency and general cognitive abilities in PISA have to be seen against the background of the characteristics of the PISA sample. As the students assessed in the PISA study ranged from the lowest to the highest level of education within Germany, predicted and explanatory variables showed more total variance and therefore the correlation values may have been higher. Second, the different conceptualizations of reading comprehension in the sense of LGVT vs. reading competency in the PISA study that includes also reading from tables and diagrams may explain the observed difference. Based on the design of the present study, the results suggest that text-oriented reading comprehension is only sparsely connected to statistics-related reading. This is in line with results of Leiss et al. (2010), who found no significant correlations between text-oriented reading comprehension in the sense of LGVT and students' (mathematical) modeling abilities.

For classroom practice, the findings suggest that relatively poor reading comprehension can be compensated for with increased cognitive abilities. Hence, also appropriate text-based learning environments may be adequate to enhance students' statistical literacy.

The third research question took further learner variables into account and hence provided a deeper insight into the relationship between SLCC and cognitive abilities. Controlling for



grades in mathematics slightly reduced the regression coefficient of cognitive abilities. Hence when comparing learners with the same grades in mathematics, their cognitive abilities are less predictive for SLCC which means that grades in mathematics relativize at least slightly the relevance of cognitive abilities. Relatively strong learners of the mathematics classroom may therefore compensate to a certain degree weaknesses of their general cognitive abilities. Even if gender and the socioeconomic status turned out to be significantly related to SLCC, they did hardly influence the relationship between SLCC and cognitive abilities.

Regarding research question 3 our data also provided evidence about significant relationships of SLCC and learner variables at different levels. At the individual level and beyond cognitive abilities, grades in mathematics seem to account for a substantial part of the SLCC score variance. As grades in mathematics might reflect mathematical competency, this interdependency may indicate that students draw on domain-specific abilities and content-related mathematical knowledge when dealing with statistical tasks. This is also in line with Kuntze et al. (2010) who found significant correlations between SLCC and conceptual knowledge in the domains of probability and functions. Furthermore, this result is congruent with findings from Klieme, Neubrand et al. (2001) and Brunner (2005) indicating that domain-specific mathematical thinking processes or mathematical abilities obviously do not completely coincide with general cognitive abilities. Hence, our findings suggest that statistical literacy requires specific mathematics-related knowledge in addition to general cognitive abilities and therefore confirm the corresponding part of Gal's (2002) model of prerequisites for statistical literacy.

Even though the bivariate correlation between gender and SLCC was not significant (see Table 5-2) this variable turned out to be significantly related to SLCC at least when controlling for general cognitive abilities and grades in mathematics. The initially non-significant correlation may be due to the fact that there are large differences of SLCC within the subsample of boys and the subsample of girls. Similar to various findings within mathematics education in particular concerning modeling tasks, on average boys seem to hold higher SLCC scores than girls. According to the logbook kept during testing, there were no obvious differences between girls and boys in respect to their concentration or motivation while solving the test, hence, we do not know the reasons for this finding. However, this result is a reason for concern because statistical literacy is needed in everyday situations by girls just as much as by boys.

For classroom practice, this finding implies that girls need to be particularly supported concerning their statistical literacy. Carmichael and Hay (2009) identified statistical contexts that are especially interesting for girls. Martignon (2010) states that girls make a greater effort when (mathematical) contents are personally relevant for them. Hence, it is possible that dealing with specific (statistical) problems may foster girls' statistical literacy. This expectation should be investigated by follow-up research. Furthermore, possible reasons for girls' lower achievement such as a weaker self-concept or gender stereotypes concerning mathematics (and statistics) education should be analyzed.

Our findings concerning the relationship between SLCC and the socio-economic status should be interpreted carefully as we assessed this variable by only the number of books at students' homes. Whereas in other studies a close relationship was found between the individual socio-economic status and competencies such as mathematical literacy (e.g. Klieme, Neubrand et al., 2001), in this study, the socio-economic status did not show to be significantly connected to students' SLCC at the individual level. However, the socio-economic status turned out to be a significant predictor at the class level. As suggested by Willms (1992), a rather high average of socio-economic status in the classes may enable teachers (including those from other school subjects) to teach at a higher level of complexity (e.g., integrating real world data and statistical representations in the classroom) because they may encounter less disciplinary problems and a more positive learning atmosphere. Another possible explanation for this relationship may be – as Opdenakker and Van Damme (2001) describe – a significant correlation between classes' mean ability and the average of socio-economic status. Hence, the association found here may also be due to further variables that were not assessed by this study. To get insight into the reasons for the relationship between SLCC and the average of socioeconomic status, further (qualitative) research is needed. In any case, a high class average of socio-economic status appears even to compensate for weaknesses at the individual level. As a large proportion of students from families with high socio-economic status achieve higher SLCC scores, students in classes with less socio-economic status might be in need of more specific support in statistics education.

The ReVa-Stat research project (“Developing concepts of data-related reduction and statistical variation for building up statistical literacy”) is supported by research funds from Ludwigsburg University of Education.

Ute Sproesser is a member of the “Cooperative Research Training Group” at the University of Tübingen and the University of Education, Ludwigsburg, which is supported by the Ministry of Science, Research and the Arts in Baden-Wuerttemberg, Germany.

## 5.9 References

- Agresti, A. & Franklin, C. (2013). *Statistics: The art and science of learning from data*. Pearson: Boston.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.J. & Weiß, M. (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Blum, W. & Leiss, D. (2005). Modellieren im Unterricht mit der „Tanken“-Aufgabe. *mathematik lehren*, 128, 18-21.
- Blum, W., Neubrand, M., Ehmke, T., Senkbeil, M., Jordan, A., Ulfig, F. & Carstensen, C.H. (2004). Mathematische Kompetenz. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff & U. Schiefele, *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 47-92). Münster: Waxmann.
- Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (241-258). New York: Greenwood.
- Bradley, R.H. & Corwyn, R.F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, 53, 371–399.
- Breiman, L., Friedman, J., Olshon, R. Stone, C. (1984). *Classification and regression trees*. Wadsworth: Belmont, CA.
- Brunner, M. (2008). No g in education? *Learning and individual differences* 18 (2008), 152-165.
- Brunner, M. (2005). *Mathematische Schülerleistung: Struktur, Schulformunterschiede und Validität*. Humboldt Universität zu Berlin. <http://edoc.hu-berlin.de/dissertationen/brunner-martin-2006-02-08/PDF/brunner.pdf> [20.05.2014].
- Carmichael, C. & Hay, I. (2009). Gender differences in middle school students' interest in a statistical literacy context. In R. Hunter, B. Bicknell, & T. Burgess (Eds.), *Crossing divides: Proceedings of the 32nd annual conference of the Mathematics Education Research Group of Australasia* (Vol. 1). Palmerston North, NZ: MERGA.
- Cohen, J. (1988): *Statistical power analysis for the behavioral sciences*, 2. Edition, Hillsdale: Lawrence Erlbaum Associates.

- Curcio, F.R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for research in mathematics education*, 18 (5), 382 – 393.
- Ditton, H. (1998). *Mehrebenenanalysen. Grundlagen und Anwendungen des Hierarchisch Linearen Modells*. Weinheim: Juventa.
- delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, 6(2), 28-58. <http://www.stat.auckland.ac.nz/serj> [28.8.2014].
- Ehmke, T., Hohensee, F., Siegle, T. & Prenzel, M. (2006). Soziale Herkunft, elterliche Unterstützungsprozesse und Kompetenzentwicklung. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost & U. Schiefele (Eds.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (225-248). Münster: Waxmann.
- Eid, M., Gollwitzer, M. & Schmitt, M. (2011). *Statistik und Forschungsmethoden*. Weinheim: Beltz.
- Enders, C.K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12(2), 121-138.
- Engel, J. & Kuntze, S. (2011). From data to functions: Connecting modelling competencies and statistical literacy. In: G. Kaiser et al. (Eds.), *Trends in teaching and learning of mathematical modelling* (S. 397-406) Heidelberg: Springer.
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) project: A pre-k-12 curriculum framework*. <http://www.amstat.org/education/gaise/> [04.09.2014].
- Fröhlich, A., Kuntze, S. & Lindmeier, A. (2007). Testentwicklung und -evaluation im Bereich von „Statistical Literacy“. *Beiträge zum Mathematikunterricht 2007* (S. 783-786). Hildesheim: Franzbecker.
- Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review* 70, 1-51.
- Gal, I., & Murray, S. (2011). Users' statistical literacy and information needs: Institutional and educational implications. *Statistical Journal of the IAOS*, 27(S. 3-4), 185-195.

- Garfield, J., Aliaga, M., Cobb, G., Cuff, C., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, R., Utts, J., Velleman, P., & Witmer, J. (2005). *Guidelines for assessment and instruction in statistics education (GAISE) project: College report*. <http://www.amstat.org/education/gaise> [04.09.2014].
- Garfield, J., & delMas, R. (2010). A web site that provides resources for assessing students' statistical literacy, reasoning and thinking. *Teaching Statistics*, 32(1), 1-7.
- Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM – The International Journal on Mathematics Education*, 44(7), 883-898.
- Hannover, B. (2010). Lernen Mädchen anders? In M. Matzner & I. Wyrobnik (Eds.), *Handbuch Mädchen-Pädagogik* (S. 95-107). Weinheim: Beltz.
- Heller, A. K. & Perleth, C. (2000). *KFT 4-12+R. Manual*. Göttingen: Hogrefe.
- Holling, H., Preckel, F. & Vock, M. (2004) *Intelligenzdiagnostik. Kompendien Psychologische Diagnostik, Bd. 6*. Göttingen: Hogrefe.
- Hosenfeld, I., Köller, O. & Baumert, J. (1999). Why sex differences in mathematics achievement disappear in German secondary schools: A reanalysis of the German TIMSS-data. *Studies in Educational Evaluation*, 25, 143-162.
- Hox, J. (2010). *Multilevel analysis. techniques and applications*. New York: Routledge.
- Kamil, M. Mosenthal, P., Pearson, P. & Barr, R. (2000). *Handbook of reading research, III*. New York: Routledge.
- Klieme, E. (1997). Gender-related differences in mathematical abilities: Effect-size, spatial mediation, and item content. *Paper presented at the 7<sup>th</sup> Conference of the European Association for Research in Learning and Instruction (EARLI)*, Athens, Greece.
- Klieme E., Funke, J., Leutner, D, Reimann, P. & Wirth, J (2001). Problemlösen als fächerübergreifende Kompetenz – Konzeption und erste Resultate aus einer Schulleistungstudie. *Zeitschrift für Pädagogik* 47 (2001), 179–200.
- Klieme, E., Neubrand, M. & Lüdtke, O. (2001). Mathematische Grundbildung: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K. J. Tillmann & M. Weiß (Eds.), *PISA 2000. Basiskompe-*

- tenzen von Schülerinnen und Schülern im internationalen Vergleich (S. 141-191). Opladen: Leske + Budrich.
- KMK (Kultusministerkonferenz) (2003). *Bildungsstandards im Fach Mathematik für den mittleren Schulabschluss*. München: Wolters Kluwer.
- Konold, C. & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education* 33(4), 259-289.
- Kröpfl, B., Peschek, W., Schneider, E. (2000). Stochastik in der Schule: Globale Ideen, lokale Bedeutungen, zentrale Tätigkeiten. *mathematica didactica*, 23, 25 – 57.
- Kuntze, S. (2013). Modellieren beim Nutzen von Darstellungen in statistischen Kontexten. Hierarchische Beschreibung und Bedingungsvariablen eines Aspekts mathematischer Kompetenz. In R. Borromeo Ferri, G. Greefrath & G. Kaiser (Eds.), *Mathematisches Modellieren in Schule und Hochschule* (S. 71-94). Wiesbaden: Springer Spektrum.
- Kuntze, S., Engel, J., Martignon, L. & Gundlach, M. (2010). Aspects of statistical literacy between competency measures and indicators for conceptual knowledge – Empirical research in the project RIKO-STAT. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. [www.stat.auckland.ac.nz/~iase/publications.php](http://www.stat.auckland.ac.nz/~iase/publications.php) [14.6.2013].
- Kuntze, S., Lindmeier, A. & Reiss, K. (2008a). „Daten und Zufall“ als Leitidee für ein Kompetenzstufenmodell zum „Nutzen von Darstellungen und Modellen“ als Teilkomponente von Statistical Literacy. *Anregungen zum Stochastikunterricht*, 4 (S. 111-122). Hildesheim: Franzbecker.
- Kuntze, S., Lindmeier, A. & Reiss, K. (2008b). “Using models and representations in statistical contexts” as a sub-competency of statistical literacy – Results from three empirical studies. *Proceedings of the 11th International Congress on Mathematical Education Monterrey (Mexico)*. <http://tsg.icme11.org/document/get/474> [22.12.11].
- Lehrer, R. (2014): Data Modeling supports development of statistical reasoning. [http://peabody.vanderbilt.edu/research/pro/about\\_peabody\\_research/funded\\_projects/data\\_modeling\\_\\_home.php](http://peabody.vanderbilt.edu/research/pro/about_peabody_research/funded_projects/data_modeling__home.php) [05.09.2014].

- Leiss, D., Schukajlow, S., Blum, W., Messner, R. & Pekrun, R. (2010). The role of the situation model in mathematical modelling – Task analyses, student competencies, and teacher interventions. *Journal für Mathematik-Didaktik*, 31(1), 119-141.
- Lindmeier, A., Kuntze, S. & Reiss, K. (2007). *Representations of data and manipulations through reduction – competencies of German secondary students*. In B. Philips & L. Weldon (Eds.), *Proceedings of the IASE/ISI Satellite Conference on Statistical Education, Guimarães, Portugal, 19-21 August 2007*. Voorburg, NL: International Statistical Institute. [http://www.stat.auckland.ac.nz/~iase/publications/sat07/Lindmeier\\_et\\_al.pdf](http://www.stat.auckland.ac.nz/~iase/publications/sat07/Lindmeier_et_al.pdf) [16.11.11].
- Martignon, L. (2010). Mädchen und Mathematik. In M. Matzner & I. Wyrobnik (Eds.), *Handbuch Mädchen-Pädagogik* (S. 220-232). Weinheim: Beltz.
- Moore, D. (1997): New pedagogy and new content: The case of statistics, *International Statistical Review*, 65(2), 123-165.
- Muthén, L.K. and Muthén, B.O. (1998-2012). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.
- OECD (2003). *The PISA 2003 assessment framework. mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD.
- Opdenakker, M.C. & Van Damme, J. (2001), Relationship between school composition and characteristics of school process and their effect on mathematics achievement. *British Educational Research Journal*, 27, 407–432.
- Paulus, C. (2009). *Die „Bücheraufgabe“ zur Bestimmung des kulturellen Kapitals bei Grundschulern*. [http://bildungswissenschaften.uni-saarland.de/personal/paulus/Artikel/BA\\_Artikel.pdf](http://bildungswissenschaften.uni-saarland.de/personal/paulus/Artikel/BA_Artikel.pdf) [15.06.2013].
- Reading, C. (2002). Profile for statistical understanding. In B. Philips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics: Developing a statistically literate society, Cape Town, South Africa*. Voorburg, the Netherlands: International Statistics Institute.
- Schneider, W., Schlagmüller, M. & Ennemoser, M. (2007). *LGVT 6-12. Manual*. Göttingen: Hogrefe.



- Sirin, S.R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75, 417-453.
- Stanat, P., Schwippert, K. & Gröhlich, C. (2010). Der Einfluss des Migrantenanteils in Schulklassen auf den Kompetenzerwerb. Längsschnittliche Überprüfung eines umstrittenen Effekts. In C. Allemann-Ghionda, C., P. Stanat, K. Göbel & C. Röhner (Eds.), *Migration, Identität, Sprache und Bildungserfolg* (147-164). Weinheim: Beltz.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Allyn and Bacon.
- Tent, L. (2001). Zensuren. In D.H. Rost (Eds.), *Handwörterbuch Pädagogische Psychologie* (S. 805-811). Weinheim: Psychologie Verlags Union.
- Therneau, T., Atkinson, E. (1997). *An introduction into recursive partitioning using the RPART routines*. <http://www.mayo.edu/hsr/techrpt/61.pdf> [20.02.2012].
- Thorndike, R.L. & Hagen, E. (1971). *Cognitive Abilities Test (CAT). Examiner's manual*. Boston: Multi-Level Edition.
- Thrupp, M., Lauder, H. & Robinson, T. (2002). School composition and peer effects. *International Journal of Educational Research* 37(2002), 483-504.
- Wallman, K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, 88(421), 1-8.
- Watermann, R. & Baumert, J. (2006). Entwicklung eines Strukturmodells zum Zusammenhang zwischen sozialer Herkunft und fachlichen und überfachlichen Kompetenzen: Befunde national und international vergleichender Analysen. In J. Baumert, P. Stanat & R. Watermann (Eds.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit* (S. 61-94). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Watson, J.M. & Callingham, R.A. (2003). Statistical literacy: a complex hierarchical construct. *Statistics Education Research Journal* 2, 3-46.
- Watson, J.M., Kelly, B.A., Callingham, R.A. & Shaughnessy, J.M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology*, Vol. 34, 1-29.

- Weinert, F. E. (2001): Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Ed.), *Leistungsmessung in Schulen* (S. 17-31). Weinheim: Beltz.
- Wild, C.J. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review* 67, 223-265.
- Willms, J. D. (1992). *Monitoring school performance*. London: Falmer Press.
- Wu, M. L., Adams, R. J., Wilson, M. R. & Haldane, S.A. (2007). *ACERConQuest Version 2: Generalised item response modelling software*. Camberwell: Australian Council for Educational Research.
- Ziegler, L.A. (2014). Reconceptualizing statistical literacy: developing an assessment for the modern introductory statistics course. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9<sup>th</sup> International Conference on Teaching Statistics*, Flagstaff, Arizona, USA. Voorburg, The Netherlands: International Statistical Institute. [http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9\\_7F1\\_ZIEGLER.pdf](http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_7F1_ZIEGLER.pdf) [16.12.2015].

## 5.10 Additional analyses of the data

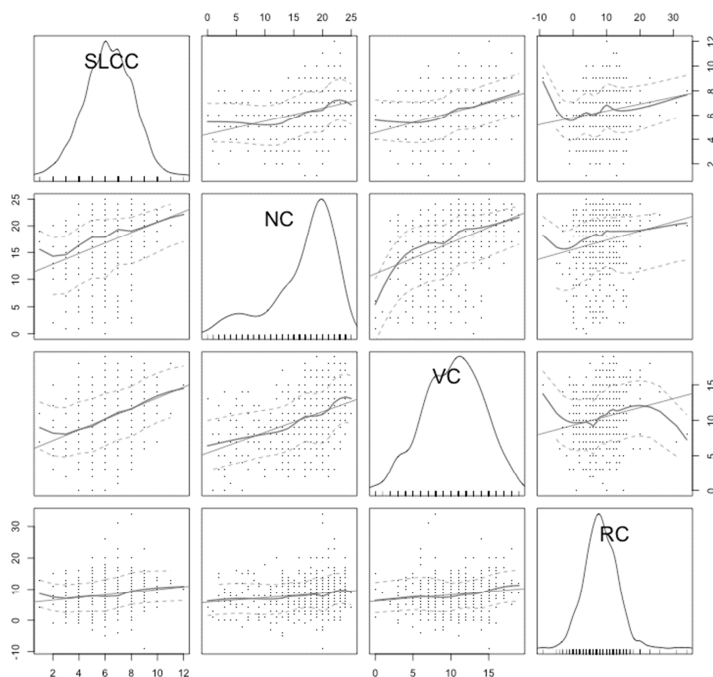
### 5.10.1 Data cleaning and checking of assumptions for linear models

A rich multivariate dataset requires careful scrutiny to check for any violations of basic assumptions. Data collected on verbal and nonverbal cognitive abilities and reading comprehension as well as the SLCC scores underwent a thorough process of data cleaning including outlier analysis, checking for normal distribution and homoscedasticity across classes. This section summarizes several additional analyses not included in the submitted paper because it exceeded the length limit of the article.

Homogeneity of variance between the 25 classes was tested based on  $F_{\max}$  (Tabachnick and Fidell, 2007). All variance ratios were within the acceptable limits. The Levene test for heteroscedasticity showed non-significant p-values between 23.89% and 92.97% for the main variables of the study. Figure 5-6 displays a scatterplot matrix for the variables reading comprehension (RC), verbal cognitive abilities (VC) and nonverbal cognitive abilities (NC) abilities as well as for the SLCC scores. On the diagonal, a nonparametric density estimate for these variables is shown, while all the pairwise scatterplots include a linear regression line as well as a nonparametric smoother with confidence bounds. Inspecting these scatterplots did not indicate violations of linearity or homoscedasticity.

Figure 5-6

Scatterplot Matrix of Density Estimates, Linear Regression Lines and Scatterplot Smoothers with Confidence Bands



Only very few cases with extreme z-scores were identified as univariate outlier. As they were not unconnected from the rest of the sample, these cases were left in the data set (cf. Tabachnick & Fidell, 2007). Mahalanobis distance ( $p < 0,001\%$ ) revealed no multivariate outlier within the data.

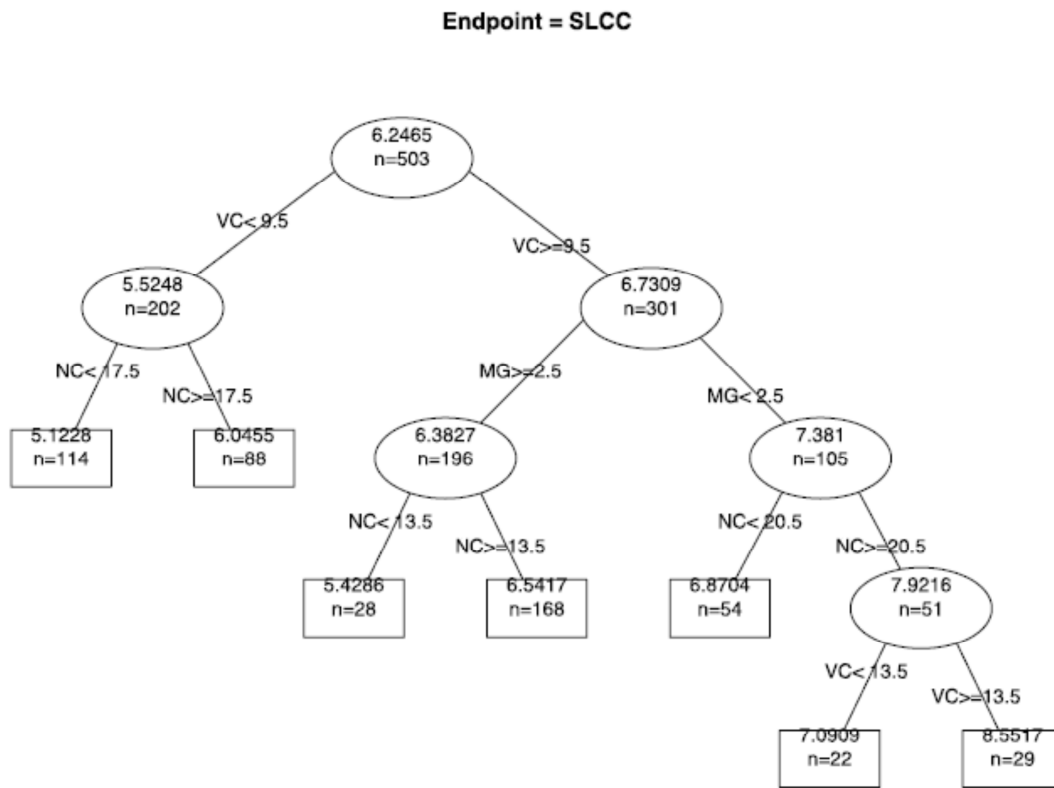
### ***5.10.2 Supplementary analyses of the relationship between the variables: CART***

We supplemented our multilevel regression analysis with a robust exploratory technique based on binary regression trees. These trees are an insightful method for statistical analysis because they are intuitive and allow us to explore the structure of a multivariate dataset while developing easy-to-visualize decision rules for predicting a categorical (classification tree) or continuous (regression tree) outcome. The algorithm for classification and regression trees (short: CART) by Breiman, Friedman, Olshen and Stone (1984) is based on a generic binary decision tree which proceeds by recursively subdividing the sample space into more homogeneous subsets and calculating locally an average. We used the implementation of the R package “rpart” (Therneau & Atkinson, 1997) leading to a regression tree for SLCC based on a model which included the following predictors: verbal and nonverbal cognitive abilities, reading comprehension, socio-economic status, gender and grades in mathematics.

The result in Figure 5-7 is a tree that has only verbal cognitive abilities (VC), nonverbal cognitive abilities (NC) and grades in mathematics (MG) as relevant variables. This outcome confirms the multilevel analysis that also identified grades in mathematics, verbal and nonverbal cognitive abilities as main predictors for statistical literacy (see Model 3 and 4). This is noteworthy because the CART algorithm does not require any of the restrictive assumptions that linear multilevel analysis are based on, such as linearity or normality.

Figure 5-7

Regression tree including grades in mathematics, verbal and nonverbal cognitive abilities



## **6 Teilstudie 2**

### **Einflussfaktoren auf die Entwicklung von Statistical Literacy und von Sichtweisen auf zufallsbedingte Variabilität in der achten Realschulklasse**

Dieser Artikel wurde bei der Zeitschrift *Lernen und Lernstörungen* veröffentlicht. Er wurde mit Genehmigung des Verlags Hans Huber im Rahmen dieser Dissertation abgedruckt.

Sproesser, U., Engel, J. & Kuntze, S. (2015). Einflussfaktoren auf die Entwicklung von Statistical Literacy und von Sichtweisen auf zufallsbedingte Variabilität in der achten Realschulklasse. *Lernen und Lernstörungen* 4, 115-129.

#### **6.1 Zusammenfassung deutsch**

Obschon grundlegende Kompetenzen des Verstehens und Interpretierens von Daten in unserer Informationsgesellschaft inzwischen als unerlässlich gelten, existieren jedoch bislang nur wenige Erkenntnisse darüber, welche Variablen die Entwicklung von Statistical Literacy begünstigen. Diese Studie untersuchte daher in einer Stichprobe von 450 Schülerinnen und Schülern der achten Realschulklasse wesentliche Variable bezüglich Statistical Literacy im Verlauf einer vierstündigen Intervention. Insbesondere wurde in den Blick genommen, inwieweit Leseverständnis, kognitive Fähigkeiten, mathematische Schulleistung und das Geschlecht dazu beitragen, Kompetenz im Bereich von Statistical Literacy und Sichtweisen auf Variabilität zu entwickeln. Während nur geringe Unterschiede in der Entwicklung von Statistical Literacy zwischen verschiedenen Treatments der Intervention festgestellt werden konnten, stellten sich kognitive Fähigkeiten, mathematische Schulleistung und Geschlecht als bedeutsam heraus. Für die Entwicklung von Sichtweisen auf zufallsbedingte Variabilität dagegen spielte ausschließlich die Treatmentzugehörigkeit eine Rolle. Die vorliegenden Ergebnisse ermöglichen für die fachdidaktische Theoriebildung bedeutsame Erkenntnisse über die Ausprägung und Entwicklung von Statistical Literacy in der achten Realschulklasse sowie Einblicke in Zusammenhänge mit individuellen Voraussetzungen der Lernenden. Im Hinblick auf die Unterrichtspraxis können die Ergebnisse einen Beitrag zu einer evidenzbasierten Einschätzung darüber liefern, inwiefern Statistical Literacy durch die entwickelten Lernmaterialien gefördert werden kann.

**Schlagwörter** *Statistical Literacy, Interventionsstudie, Leseverständnis, kognitive Fähigkeiten, Geschlecht*

## **6.2 Ausführliche englische Zusammenfassung**

### **Factors that influence the development of statistical literacy and of views on variation in grade eight**

#### ***6.2.1 Background***

One of today's challenges for citizens is to understand data in the news, brochures and advertisements in order to be able to make good decisions under uncertainty (Wallman, 1993). Data are, in fact, almost omnipresent (Ben-Zvi & Garfield, 2005). Statistical literacy involves an understanding and a critical evaluation of data in everyday life as well as the awareness that statistical thinking enhances decision-making (Wallman, 1993). The recent development of many national curricula testifies how the relevance of statistical literacy is being acknowledged.

Far from being limited to mathematical exercises only, statistical literacy aims at using statistical models to address “realworld problems” (Ben-Zvi & Garfield, 2005; Wallman, 1993). As is the case for many instances of mathematical modeling (cf. Blum & Leiss, 2005), statistical literacy requires the ability to understand, create, manipulate and investigate useful and informative representations of real world situations. Correspondingly, dealing with representations cannot be separated from modeling in statistical contexts. One phenomenon that often requires statistical modeling is the variation in data (Wild & Pfannkuch, 1999). The analysis of variation in data may draw on our ability to quantify uncertainty or to focus on determining which component of variation can be attributed to randomness and which to systematic effects. Under random variation in data we understand fluctuations seen subjectively as negligible that cannot be anticipated, detected, identified or eliminated. In a more technical perspective random variation is seen as noise to be separated from the systematic signal contained in the data (Konold & Pollatsek, 2002). Recognizing this difference is crucial for explanation and prediction data as well as for controlling the underlying processes (Wild & Pfannkuch, 1999). Modeling statistical variation often requires data reduction steps because meaningful overviews cannot consider the whole amount of complex information (Kröpfl et al., 2000). Thus a good understanding of data reduction requires an awareness of the loss of information in such overviews provided by graphical or numerical summaries. The abilities of dealing with variation and reducing data have shown to be important indicators of statistical literacy (Reading, 2002; Watson & Callingham, 2003). Therefore, Kuntze and colleagues (2008) have integrated these ideas into the metaphor of data-related reading when establishing a competency model in the domain of statistical literacy. The competency “using models and repre-

sentations in statistical contexts” describes students’ competencies at five levels. Rasch analyses of a corresponding test instrument have confirmed the hierarchical structure of this competency (Kuntze, Engel, Martignon & Gundlach, 2010; Kuntze, Lindmeier & Reiss, 2008).

This study analyzed how factors such as reading comprehension, general cognitive abilities, grades in mathematics and gender interdepend with the development of students’ competency “using models and representations in statistical contexts” and of their views on variation in the course of a statistics-related intervention. These factors have been detected as relevant for achievement in mathematical and statistical contexts (e. g. Brunner, 2005; Hosenfeld et al., 1999; Klieme et al., 2001; Sproesser, Kuntze & Engel, submitted). The intervention trained students according to a 2×2-design composed by four treatments: treatment 1 dealt with tables, diagrams and numerical parameters in statistical contexts. The remaining three treatments were enriched by reflections on reduction (treatment 2), variation (treatment 3) or both (treatment 4). The impact of these treatments on the development of students’ competency and of their views on variation was also measured.

### **6.2.2 Aims**

This study investigated how competent students of grade eight are and become concerning “using models and representations in statistical contexts” in the course of a statistics-related intervention. Furthermore, we examined how reading comprehension, cognitive abilities, grades in mathematics and gender influence the development of this competency. We also evaluated how these individual factors affect the development of students’ views on variation. In addition, we measured the impact of four different treatments on students’ competency and their views on variation.

### **6.2.3 Methods**

This study analyzed data from 450 eighth-graders who took part in a four lesson intervention. In each of the 25 classes, students were assigned to four treatment groups according to a mixed 2×2-design: The (1) basic training involved dealing with tables, diagrams and characteristic values in a rather algorithmic way. The treatments (2) “reduction”, (3) “variation” and (4) “reduction and variation” also addressed tasks encouraging learners to reflect about data reduction or/and statistical variation rather than dealing only algorithmically with data. Before and after the intervention, students completed a paper and pencil test measuring the competency “using models and representations in statistical contexts” (Kuntze, 2013; Kuntze et al., 2008). In four other items students were requested to display their understanding of a typical random distribution of objects (e. g. each three of four) by marking figures out of a given se-



quence (see Fig. 6-2). Reading comprehension was measured by the test LGVT 6-12 (Schneider et al., 2007). For cognitive abilities, we used a verbal and a nonverbal subscale of the test KFT 4-12+R (Heller & Perleth, 2000). Teachers informed us about gender and grades in mathematics. Multiple regression analyses were conducted via the “type is complex” procedure in Mplus 7.1 (Muthén & Muthén, 1998 – 2012), which adjusts deflated standard errors due to nested data in school classes.

#### **6.2.4 Results**

Within our sample, we found an average competency of “using models and representations in statistical contexts” of 5.8 points (SD 1.6) in the pre- and 6.8 points (SD 1.3) in the post-test (each out of 14 points). The solution frequencies show that especially items which required dealing with variation uncovered corresponding deficits of the students. With respect of potential impact factors cognitive abilities ( $\beta = .168$ ,  $p < 0.001$ ), grades in mathematics ( $\beta = -.107$ ,  $p = 0.001$ ) as well as gender ( $\beta = -.111$ ,  $p = 0.012$ ) turned out to be significant predictors for the competency development in the domain of “using models and representations in statistical contexts”. The influence of reading comprehension was not significant. The treatment groups “variation” and “reduction and variation” did not significantly differ from the basic training. However, students in the treatment “reduction” showed a smaller increase of competency than their peers in the basic training ( $\beta = -.097$ ,  $p = 0.022$ ).

Reading comprehension, cognitive abilities, grades in mathematics and gender did not significantly impact on the development of students' views on variation. Collating the students of the treatments “reduction”, “variation” and “reduction and variation” to one reflection-oriented treatment group indicated a significantly higher increase ( $\beta = .061$ ,  $p = 0.016$ ) in these students' views on variation than compared to their peers in the basic training ( $d = 0.25$ ).

#### **6.2.5 Discussion**

A central finding of this research is that the eighth-graders of our sample showed deficits related to the competency of “using models and representations in statistical contexts” especially with regard to dealing with variation. Whereas potentially low of reading comprehension values did not seem to prevent learners from developing higher competency in the domain of statistical literacy, cognitive abilities and grades in mathematics turned out to be significant predictors for the development of this competency. Boys appeared to benefit more from our intervention. This could mean that girls may need more specific support in the domain of statistics to counteract a systematic gender effect. Gender differences of this kind have been reported by several mathematics-related achievement studies. Moreover, possible reasons for

this gender effect such as the mediation of self-concept (e. g. Hannover, 2010) should be investigated by follow-up research. From the analysis of the four treatments we deduced: Reflecting about reduction or/and variation in the course of the intervention did not enable students to achieve higher scores on the posttest than their peers who only had participated in treatment 1, i. e. the basic training. A possible explanation is a bottom effect: the tasks dealing with variation (see Tab. 6-1) were so demanding for eighth-graders that many of them failed in them, even after the intervention. However, reflecting about reduction or/and variation (treatments 2, 3, 4) turned out to positively influence the other outcome variable, namely students' views on random variation.

**Keywords** *Statistical Literacy, intervention study, reading comprehension, cognitive abilities, gender*

## 6.3 Einleitung

### 6.3.1 *Statistical Literacy*

Der sinnstiftende Umgang mit Daten beispielsweise aus den Medien ist unumgänglich, um an gesellschaftlichen Prozessen teilzuhaben, datenbasierte Entscheidungen treffen zu können und somit reflektiert handeln zu können (Kuntze, Lindmeier & Reiss, 2008; Wallman, 1993). Dies reicht vom Verständnis einfacher statistischer Darstellungen bis hin zum eigenständigen Interpretieren statistischer Information. Grundlegende Kompetenzen des Umgangs mit statistischen Daten werden oft mit dem Begriff *Statistical Literacy* bezeichnet. Wallman (1993) definiert *Statistical Literacy* als das Verstehen und kritische Überprüfen von alltagsrelevanten Daten auch in Hinblick auf das Treffen von Entscheidungen.

Ähnlich wie *Mathematical Literacy* zielt *Statistical Literacy* nicht auf die Beherrschung isolierter Regeln, Verfahren oder Fertigkeiten ab; vielmehr sollen in dieser Weise gebildete Individuen zum Lösen vielfältiger Probleme fähig sein (Ben-Zvi & Garfield, 2005; Klieme, Neubrand & Lüdtke, 2001; Wallman, 1993). Ähnlich wie bei *Mathematical Literacy* (Klieme et al., 2001) stellt Modellierungsfähigkeit auch beim Lösen von Problemen in statistischen Kontexten ein zentrales Element von *Statistical Literacy* dar (Kuntze et al., 2008; Wild & Pfannkuch, 1999). Modellieren (Blum & Leiss, 2005) erfordert den verständnisvollen Wechsel zwischen der konkreten Sachsituation und einem zur Lösung genutzten Modell. Da Daten oft nur über statistische Darstellungen wie Listen, Tabellen oder Graphiken zugänglich sind, lassen sich Modellierungsaktivitäten in statistischen Kontexten überdies kaum vom Nutzen von Darstellungen trennen. So muss beispielsweise beim Wechseln zwischen verschiedenen Darstellungen von Daten, aber auch bei deren Interpretation zwischen Sachsituation und statistischem Modell gewechselt, also modelliert werden. In diesem Sinn stellen das Nutzen von Darstellungen und das Modellieren grundlegende und kaum voneinander zu trennende Aktivitäten des Umgangs mit Daten dar (z. B. Kuntze, 2013; Kuntze et al., 2008).

Die Notwendigkeit von Modellierungsaktivitäten in statistischen Kontexten wird insbesondere deutlich, wenn man sich bewusst macht, dass statistische Daten selten schwankungsfrei sind oder exakt regulären Mustern folgen. Variabilität zeigt sich bei vielen Beobachtungen und Messungen. Beispielsweise unterscheiden sich Lernende hinsichtlich ihrer Körpergröße, ihres IQs oder ihrer Mathematiknote. Dabei kann Variabilität systematische Ursachen haben oder von unregelmäßigen, unberechenbaren oder subjektiv als unwichtig erachteten Schwankungen hervorgerufen sein, die in der Praxis oft nicht vorhergesehen, identifiziert oder eliminiert werden können. Von einer technischen Perspektive aus kann letztere Form der Variabili-

tät als ein zufallsbedingtes Rauschen angesehen werden, das von dem zu erkennenden Signal nach Möglichkeit zu trennen ist (Konold & Pollatsek, 2002). Watson und Callingham (2003) identifizierten das Verständnis für Variabilität als zentrales Element eines Kompetenzmodells für Statistical Literacy.

Eine Beispielaufgabe zum Umgang mit Variabilität findet sich in Abbildung 6-1: Es zeigen sich zwischen fast allen Monaten Unterschiede im Kaffee-Konsum von Herrn Arnold. Allerdings liegt im Juli ein deutlicher Ausschlag nach oben vor, während der Konsum in den anderen Monaten nur geringfügige Unterschiede aufweist. Sinnvoll mit Variabilität umgehen zu können, könnte hier bedeuten, die Variabilität des Kaffeekonsums zwischen den Monaten Februar bis Juni dem Zufall zu attribuieren – im Unterschied zu der erheblicheren Abweichung im Monat Juli. Je nachdem, ob die Unterschiede als zufällig oder als signifikant eingeschätzt werden, können sie ignoriert oder müssen speziell beachtet werden. Die situationsangemessene Beurteilung von Variabilität ermöglicht gegebenenfalls die Erklärung der Unterschiede, Vorhersagen zukünftiger Entwicklungen und die Steuerung von zu Grunde liegenden Prozessen (Wild & Pfannkuch, 1999). In der vorliegenden Beispielaufgabe bedeutet dies, dass der deutlich höhere Kaffee-Konsum im Juli als solcher erkannt wird und somit nicht die alleinige Grundlage eines monatlichen Pauschalpreises sein sollte.

Abbildung 6-1

Beispielaufgabe „Variabilität“

Monat	Kaffee-Verbrauch Arnold
Juli	☞ ☞ ☞ ☞ ☞
Juni	☞ III
Mai	☞ I
April	☞ IIII
März	☞ II
Februar	☞ III

Ausgehend vom Kaffee-Verbrauch im Juli soll für jeden Mitarbeiter ein monatlicher Pauschalpreis ermittelt werden.  
Herr Arnold findet das nicht gerecht.

Dem teilweise überwältigende Umfang von durch Variabilität geprägten Daten kann oft nur durch die Reduzierung dieser Daten (z. B. Auslassen von Details), beispielsweise zu einem numerischen Kennwert, begegnet werden (Kröpfl, Peschek & Schneider, 2000). In einer Studie von Reading (2002) konnte empirisch gezeigt werden, dass sich das Spektrum des statistischen Verständnisses von Lernenden durch Fähigkeiten im Bereich der Reduktion abbilden lässt. Dabei ist zu beachten, dass Fähigkeiten des Reduzierens von Daten nicht auf die algo-

rithmische Bildung einer numerischen oder grafischen Zusammenfassung beschränkt sind, sondern Wissen, was bei solchen Reduktionsprozessen mit den Daten geschieht (z. B. über den Verlust an Information), miteinschließen. Die Entscheidung für eine geeignete Art der Reduzierung ist oft komplex und kann in Anbetracht statistischer Variabilität mit der Metapher des Suchens nach dem Signal im Rauschen (Konold & Pollatsek, 2002) beschrieben werden. Dies bedeutet, zufällige Unterschiede, d. h. das Rauschen, nach Möglichkeit auszublenken, um bedeutsame Information, d. h. ein Signal, zu identifizieren. Dabei muss darauf geachtet werden, dass die bedeutsame Information nicht durch eine zu starke Reduzierung verloren geht. Bezogen auf die oben abgebildete Aufgabe könnte nach der Identifizierung des abweichenden Verbrauchs im Juli beispielsweise ein Durchschnittswert für die Monate Februar bis Juni gebildet werden, um einen monatlichen Pauschalpreis festzulegen, sofern es sich im Juli nur um eine erklärbare einmalige Abweichung handelt.

### ***6.3.2 „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ – ein Kompetenzmodell im Bereich von Statistical Literacy***

Um Kompetenz im Bereich von Statistical Literacy im Sinne eines Kompetenzmodells hierarchisch beschreiben zu können, verbinden Kuntze und Kollegen (2008; Kuntze, Engel, Martignon & Gundlach, 2010 bzw. Kuntze, 2013) grundlegende Elemente von Statistical Literacy durch die Metapher des datenbezogenen Lesens (vgl. Curcio, 1987): Curcio hatte ursprünglich ein dreistufiges Modell bezüglich des Lesens von Daten aus einem Graphen konzipiert, das vom einfachen Ablesen eines Datenpunktes bis hin zu Prognosen über die gegebenen Daten hinaus reicht. Kuntze et al. (2008; 2010; siehe auch Kuntze, 2013) erweitern dieses Modell insbesondere um die Erkenntnis, dass beim Lesen von Daten der erfolgreiche Umgang mit statistischen Darstellungen auch das Nutzen von Modellen miteinbezieht. In ihrem fünfstufigen Modell der innerhalb von Statistical Literacy angesiedelten Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ (siehe Tab. 6-1 für eine Übersicht) ist außerdem der Umgang mit Variabilität und Reduktion von Bedeutung.

Mit Hilfe eines entsprechenden Rasch-skalierbaren Testinstruments konnte die hierarchische Struktur dieses Kompetenzkonstrukts entlang einer Dimension empirisch bestätigt werden (Kuntze et al., 2008). Das beschriebene Kompetenzmodell liegt dieser Studie zu Grunde. In Sproesser et al. (eingereicht) erläutern wir die Kompetenzstufen des Modells auch anhand von Aufgaben, die in der vorliegenden Studie verwendet wurden.

Tabelle 6-1

Kompetenzmodell zum Nutzen von Darstellungen und Modellen in statistischen Kontexten (Kuntze, 2013, S. 7)

Kompetenz-niveau I	Einschrittiges Nutzen einer Darstellung oder Arbeiten innerhalb eines gegebenen Modells (z.B. Ablesen eines gegebenen Werts aus einem Diagramm, Vervollständigen eines gegebenen Diagramms bei gegebener Datentabelle)
Kompetenz-niveau II	Zwei- oder mehrschrittiges Nutzen von Darstellungen oder Wechsel zwischen zwei gegebenen Modellen (z.B. Vergleichen von Daten unter Einschluss eines transformierenden Schrittes oder von Begriffswissen)
Kompetenz-niveau III	Mehrschrittiges Nutzen von Darstellungen einschließlich der Nutzung eines nicht gegebenen Modells (z.B. eigene Modellierungsaktivitäten zur Unterstützung einer kumulativen Interpretation von in Diagrammen gegebenen Daten)
Kompetenz-niveau IV	Mehrschrittiges Nutzen von Darstellungen bzw. Nutzung eines nicht gegebenen Modells einschließlich des adäquaten Umgangs mit statistischer Variabilität (z.B. eigene Modellierungsaktivitäten auf der Basis von Diagrammdarstellungen, die den Umgang mit statistischer Variabilität mit einschließen)
Kompetenz-niveau V (Projekt RIKO- STAT)	Mehrschrittiges Nutzen von Darstellungen bzw. Nutzung eines nicht gegebenen Modells einschließlich des adäquaten Umgangs mit statistischer Variabilität und des Modellierens/mentalen Konstruierens nicht gegebener Daten (z.B. mögliche zugrunde liegende Daten modellieren beim Beurteilen oder Interpretieren gegebener Aussagen)

### 6.3.3 Sichtweisen bezüglich zufallsbedingter Variabilität

Wie bereits beschrieben, spielt der Umgang mit Variabilität eine wesentliche Rolle für Statistical Literacy. So sind adäquate Vorstellungen über statistische Variabilität gefordert, wenn Lernende vorliegende Daten situationsangemessen erklären sollen oder wenn Entscheidungen auf der Grundlage von Daten getroffen werden müssen. Wild und Pfannkuch (1999) weisen darauf hin, dass Menschen dazu tendieren, auch dann nach kausalen Erklärungen für Abweichungen zu suchen, wenn Beobachtungen unter Berücksichtigung von Zufallsvariabilität noch gut innerhalb der Grenzen des zu Erwartenden liegen. Engel und Sedlmeier (2005) rechnen aufgrund solcher Sichtweisen mit Schwierigkeiten bei der Entwicklung eines angemessenen Verständnisses für zufallsbedingte Variabilität und somit beim Aufbau von Statistical Literacy. Empirische Studien zu Sichtweisen bezüglich Variabilität bestätigen die Vorliebe von Schülerinnen und Schülern für regelmäßige, gleichsam „variabilitätslose“ Strukturen in Daten (Engel & Sedlmeier, 2005; Green, 1982, 1986, 1990; Piaget und Inhelder, 1975). Gründe dafür sehen Engel und Sedlmeier (2005) wie auch Green (1982) und Fischbein (1975) im Einfluss der Schule. Hier werde den Lernenden häufig ein deterministisches Denken vermittelt,

das zufallsbedingte Variabilität kaum thematisiere. Diese Einschätzung legt es nahe, Schülerinnen und Schüler durch die unterrichtliche Auseinandersetzung mit zufälligen Erscheinungen für die Berücksichtigung von Variabilität zu sensibilisieren.

#### ***6.3.4 Lernchancen durch die Auseinandersetzung mit zufallsbedingter Variabilität und datenbezogener Reduktion***

Die vorliegende Studie sah Reflexionsprompts bezüglich Variabilität oder/und Reduktion im Sinne eines 2×2-Designs vor, um differenzielle Effekte dieser Reflexionsanlässe erfassen zu können. Neben einem Basis-Training, das auf einen eher algorithmischen Umgang mit Daten abzielte, regte je ein Treatment das Nachdenken über Reduktion beziehungsweise Variabilität an. Zusätzlich enthielten die Materialien eines weiteren kombinierten Treatments Reflexionsanlässe bezüglich beider Konzepte. Die Anregung dieser spezifischen Reflexionen zielte auf die Veränderung von Sichtweisen bezüglich zufallsbedingter Variabilität sowie auf die Entwicklung der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ ab. Neben der Wirkung der vier Treatments (1) „Basis-Training“, (2) „Reduktion“, (3) „Variabilität“ sowie (4) „Reduktion und Variabilität“ nahm die vorliegende Studie einige individuelle Lernvoraussetzungen<sup>4</sup> in den Blick, die die Entwicklung dieser Variablen im Laufe der Intervention begünstigen könnten.

#### ***6.3.5 Empirische Befunde zu möglichen Einflussfaktoren auf kognitive Variablen im Bereich von Statistical Literacy***

Es ist zu erwarten, dass im Verlauf der Intervention nicht nur die vier Treatments, sondern auch individuelle Voraussetzungen der Schülerinnen und Schüler dazu führen, dass sich deren Sichtweisen auf Variabilität sowie die Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ unterschiedlich entwickeln. Insofern werden in diesem Abschnitt empirische Befunde früherer – auch mathematikdidaktischer Studien – vorgestellt, die den Einbezug der genannten Variablen in diese Studie begründen.

Beispielsweise wurde in der PISA-Studie (Klieme et al., 2001) ein Zusammenhang zwischen Lesekompetenz (im Sinne einer „aktiven (Re-)Konstruktion der Textbedeutung“, S. 71) und mathematischer Kompetenz (Pfadkoeffizient  $\beta = .55$ ,  $p < 0.05$ ) gefunden. Ebenso bestätigte PISA den erwarteten Zusammenhang zwischen kognitiven Grundfähigkeiten und mathemati-

---

<sup>4</sup> Im folgenden Text wird zugunsten der Lesbarkeit darauf verzichtet, beide Geschlechter in zusammengesetzten Formulierungen wie „Lernvoraussetzungen“ oder „schülerzentriert“ explizit zu benennen. Selbstverständlich sind darunter beide Geschlechter zu verstehen.

scher Kompetenz (Pfadkoeffizient  $\beta = .32$ ,  $p < 0.05$ ) empirisch. Leserverständnis und kognitive Grundfähigkeiten wurden aus diesem Grund in diese Studie einbezogen.

Zensuren wie die Mathematiknote stellen fachspezifische Indikatoren für individuelle Eingangsvoraussetzungen der Lernenden dar, die als solche die Entwicklung von Kompetenzen in benachbarten Bereichen beeinflussen könnten. Die meist relativ starken Korrelationen zwischen Schulleistungstests und Fachnoten (z. B. Tent, 2001, S. 808:  $-.60 \leq r \leq -.70$ ) deuten auf eine angemessene Aussagekraft dieser Indikatoren hin. Dies legt den Einbezug der Mathematiknote in die vorliegende Studie nahe.

Im Zusammenhang mit mathematischen Kompetenzen werden in der Literatur durchgängig Geschlechterunterschiede zu Gunsten der Jungen berichtet; diese fallen besonders stark beim mathematischen Problemlösen aus (Brunner, 2005; Hosenfeld, Köller & Baumert, 1999; Klieme, 1997; Klieme et al., 2001). Insofern wurde auch das Geschlecht der Teilnehmenden in die Auswertung aufgenommen.

#### **6.4 Forschungsfragen**

Trotz der Bedeutung von Statistical Literacy im täglichen Leben sind Interventionsstudien, die auf deren Förderung abzielen, selten. Insofern fokussierte die vorliegende Studie auf ein weitgehend unbekanntes Forschungsfeld, indem sie die Entwicklung von zwei für Statistical Literacy relevante Zielvariablen während einer Intervention in den Blick nahm: Zunächst wurde ermittelt, auf welchem Niveau sich die Kompetenz „Nutzen von Darstellungen und Kompetenzen in statistischen Kontexten“ der Teilnehmenden im Durchschnitt verorten ließ. Anschließend wurde untersucht, inwieweit die verschiedenen Treatments und einige individuelle Lernervoraussetzungen diese Kompetenz beeinflussten. Die Forschungsfragen bezüglich der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ lauten demnach:

- 1) Über welche Kompetenz des „Nutzens von Darstellungen und Modellen in statistischen Kontexten“ verfügten die Lernenden?
- 2) Inwieweit begünstigten die Lernervoraussetzungen Leseverständnis, kognitive Fähigkeiten, die Mathematiknote und das Geschlecht die Entwicklung dieser Kompetenz?
- 3) Welche Treatments wirkten besonders förderlich auf die Kompetenz?

Obgleich der oben geschilderte Kompetenztest Aufgaben enthält, die den Umgang mit Variabilität erfordern, lässt deren Bearbeitung keine direkten Rückschlüsse auf diesbezügliche



Sichtweisen zu. In einem zweiten Schritt wurde deshalb die Entwicklung von Sichtweisen auf Variabilität im Verlauf der Intervention als separate Zielvariable analysiert. Es wurden folgende Forschungsfragen untersucht:

- 4) Inwieweit beeinflussten die Lernervoraussetzungen Leseverständnis, kognitive Fähigkeiten, die Mathematiknote und das Geschlecht die Entwicklung von Sichtweisen auf Variabilität?
- 5) Welche Treatments wirkten besonders förderlich auf die Entwicklung von Sichtweisen auf Variabilität?

## **6.5 Methodik**

### ***6.5.1 Forschungsdesign und Stichprobe***

Den hier berichteten Ergebnissen liegen Daten von 450 Schülerinnen und Schülern (davon 212 weiblich) aus 25 achten Klassen zu Grunde. Die Lernenden im Alter zwischen 12 und 15 Jahren ( $M = 13.50$ ;  $SD = 0.62$ ) stammten aus 8 öffentlichen Realschulen im Umfeld von Ludwigsburg. Bei der Auswahl dieser Schulen wurde auf ein ausgewogenes Verhältnis zwischen städtischem und ländlichem Einzugsgebiet geachtet.

Im Zuge dieser Studie nahmen die Lernenden an einer vierstündigen statistikbezogenen Intervention teil. Dabei wurden die Schülerinnen und Schüler jeder Klasse parallelisiert nach ihren Vortestleistungen entsprechend eines mixed  $2 \times 2$ -Designs in die vier Treatments (1) „Basis-Training“ ( $n = 117$ ), (2) „Reduktion“ ( $n = 102$ ), (3) „Variabilität“ ( $n = 120$ ), sowie (4) „Reduktion und Variabilität“ ( $n = 111$ ) eingeteilt. Die Aufgaben des Basis-Trainings ( $n = 117$ ) bezogen sich im Wesentlichen auf grundlegende Fähigkeiten des Umgangs mit statistischen Darstellungen wie Tabellen und Diagrammen oder der Bestimmung von Kennwerten. Die anderen drei Interventionstypen enthielten eine geringere Anzahl dieser Grundaufgaben zugunsten zusätzlicher Reflexionsanlässe, so dass die Lernzeiten der vier Treatmentgruppen identisch waren. In Sproesser, Engel und Kuntze (angenommen) sind Aufgaben der verschiedenen Treatments abgebildet. Die zeitliche Struktur des Forschungsdesigns wird im elektronischen Supplement besser beschrieben.

Zu Beginn der Intervention hatten die beteiligten Lehrerinnen und Lehrer feste Zweierteams innerhalb der einzelnen Treatmentgruppen gebildet, die laut ihrer Erfahrung gut miteinander arbeiten konnten. Die lernerzentrierten Materialien der Intervention wurden insofern durchgängig in Partnerarbeit bearbeitet. Während der Intervention hatte die jeweilige Lehrperson

die Aufgabe, die Klassen zu beaufsichtigen und bei Disziplinproblemen einzuschreiten. Die Schülerinnen und Schüler konnten jederzeit auf standardisierte Rückmeldung in Form von Hilfekarten und Musterlösungen zugreifen. Außerdem konnten sie bei Verständnisfragen auf die Erstautorin zukommen. Nach Bearbeitung eines bestimmten Umfangs an Aufgaben erhielten sie individuelles Feedback zur Qualität der Bearbeitungen und wurden gegebenenfalls dazu aufgefordert, bestimmte Teilaufgaben nochmals zu überarbeiten.

### **6.5.2 Messinstrumente**

Die Teilnehmenden bearbeiteten vor und nach der Intervention einen schriftlichen Kompetenztest zum „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ (Kuntze, 2013; Kuntze et al., 2008; Kuntze et al., 2010). Vor- und Nachtest enthielten je 5 identische Aufgaben und 10 Parallelaufgaben, somit insgesamt jeweils 15 Aufgaben. Diese können den fünf hierarchischen Niveaustufen aus Tabelle 1 zugeordnet werden. Die empirischen Lösungshäufigkeiten bestätigten diese Hierarchie in unserer Stichprobe. Die technischen Details der Analyse des Testinstruments werden in Sproesser et al. (eingereicht) dargestellt.

Zusätzlich zum beschriebenen Kompetenztest sollten die Versuchspersonen ihr Sichtweisen bezüglich zufälliger Vorgänge darstellen. Hierzu wurden ihnen vier Aufgaben vorgelegt, in denen sie eine zufällige Verteilung von Objekten, beispielsweise von Zahnspangen unter Kindern, markieren sollten (vgl. Engel, 2011; Engel & Sedlmeier, 2005; Kuntze, 2012). Der Typ der Fragestellung zielte nicht auf eine Bewertung der Antworten gemäß der Kategorien richtig/falsch ab, sondern die Antworten können als Ausdruck davon gesehen werden, wie stark das Denken der Schülerinnen und Schüler Erscheinungsformen von Variabilität einbezieht. Im Beispiel-Item (siehe Abb. 6-2) wurden die Lernenden dazu aufgefordert, eine aus ihrer Sicht typische Verteilung von Zahnspangen in den 16 linear hintereinander abgebildeten Figuren zu kennzeichnen. Diese Figuren waren in vier Vierergruppen arrangiert, so dass die Antworten in drei Kategorien eingeteilt werden konnten. Somit war es möglich, Abstufungen in den Sichtweisen auf Variabilität zu ermitteln (vgl. Kuntze, 2012): Bearbeitungen, die in jeder Vierergruppe in regelmäßiger Struktur ein Individuum ausließen und somit keine Berücksichtigung von Variabilität in den Sichtweisen der Lernenden widerspiegeln, wurden mit dem Code 0 versehen. Dagegen wurden Schülerdokumente mit 1 kodiert, wenn sie in jeder Vierergruppe genau drei Markierungen in unregelmäßiger Reihenfolge aufwiesen und somit Sichtweisen dokumentierten, die Variabilität in den einzelnen Vierergruppen einbezogen. Für Verteilungen, die in der Gesamtgruppe kein regelmäßiges Muster, sondern Variabilität aufwiesen, wurde der Code 2 vergeben. Bei Code 2 war es unerheblich, ob die Lernenden entsprechend

dem durchschnittlichen Verhältnis (3 von 4 Kindern im Allgemeinen, das entspricht 12 von 16 abgebildeten Figuren) genau 12 Individuen markierten oder ob sie davon abwichen. Um einen groben Indikator für die Sichtweisen der Schülerinnen und Schüler auf Variabilität zu erhalten, wurde in dieser Studie der einfache Summenwert der vier Aufgaben gebildet. Dabei drücken höhere Summenwerte eine stärkere Berücksichtigung von Variabilität in den Verteilungen aus.

Abbildung 6-2

Beispiel-Item zur Erhebung von Sichtweisen auf Variabilität

### Zahnspangen

Karin weiß, dass Kieferorthopäde Dr. Müller **durchschnittlich drei von vier Patienten** eine Zahnspange verpasst. Sie befragt an einem Tag vor seiner Praxis die 16 nacheinander herauskommenden Patienten.

Zeichne in die abgebildeten Smileys die Spangen so ein, wie du es in der Reihenfolge der herauskommenden Patienten für typisch halten würdest.



Das Leseverständnis der Lernenden wurde mit dem Lesegeschwindigkeits- und -verständnistest LGTV 6-12 (Schneider, Schlagmüller & Ennemoser, 2007) erhoben, der im Rahmen von PISA 2000 entwickelt wurde. Es handelt sich dabei um einen kontinuierlichen Text, in dem in regelmäßigen Abständen ein Wort fehlt. Diese Lücken müssen die Schülerinnen und Schüler mit dem jeweils sinnvollsten von drei zur Auswahl gestellten Wörtern füllen. Da die Bearbeitungszeit lediglich vier Minuten beträgt, muss nicht nur sinnentnehmend, sondern auch schnell gelesen werden. Für jede richtige Ergänzung erhalten die Lernenden zwei Punkte, für jede falsche Markierung wird ein Punkt abgezogen. Der Gesamtscore berechnet sich als Summenwert. Schneider und Kollegen (ibid.) geben die Retestreliaibilität mit  $r = 0.87$  an.

Mit Hilfe zweier Subskalen des Kognitiven Fähigkeitstests KFT 4-12+R (Heller & Perleth, 2000) wurden allgemeine kognitive Fähigkeiten der Teilnehmenden erfasst. Die Subskala „Figurale Analogien B“ umfasst 25 Items im Multiple-Choice-Format, bei denen die Schülerinnen und Schüler figural dargestellte Beziehungen zunächst erfassen und dann auf weitere Figuren übertragen müssen. Die interne Konsistenz sowie die Retestreliaibilität wird von Heller und Perleth (ibid.) mit 0.93 angegeben. Zusätzlich wurde die Subskala „Wortanalogien B“ eingesetzt, bei der ein ähnliches Vorgehen wie bei „Figuralen Analogien“ gefordert ist; ledig-

lich die Darstellungsformate der beiden Testteile unterscheiden sich. Die interne Konsistenz sowie die Retestrelabilität liegen bei jeweils 0.81. Da in dieser Studie lediglich ein Indikator für kognitive Fähigkeiten benötigt wurde, genügte die Zusammenfassung beider Subskalen zu einem einfachen Summenwert.

Das Geschlecht der Teilnehmenden und die Mathematiknote des vergangenen siebten Schuljahres wurden anonymisiert über die jeweilige Lehrperson erhoben.

## **6.6 Ergebnisse**

### ***6.6.1 Der Einfluss von individuellen Voraussetzungen auf die Entwicklung der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“***

In dieser Studie kam es zu Decken- und Bodeneffekten beim Kompetenztest: Die Mehrheit der Probanden hatte kaum Schwierigkeiten, Aufgaben auf der Niveaustufe I und II korrekt zu beantworten, während nur sehr wenige Schülerinnen und Schüler die Aufgaben der Stufen IV und V lösen konnten. Im Verlauf der Intervention erhöhten sich die Gesamtmittelwerte im Kompetenztest von 5.8 ( $SD \approx 1.6$ ) auf 6.8 ( $SD \approx 1.3$ ) Punkte. Dabei wurden über alle Treatmentgruppen hinweg die stärksten Punktezuwächse in Aufgaben des zweiten und dritten Kompetenzniveaus erreicht, während sich die Lösungsraten der Aufgaben auf den höchsten beiden Kompetenzniveaus zwischen Vor- und Nachtest kaum veränderten (pro Aufgabe maximal 5.8 % Vortest, maximal 6.4 % im Nachtest). Im Durchschnitt waren die Kompetenzausprägungen der Lernenden dieser Stichprobe in Vor- und Nachtest demnach hauptsächlich auf den Kompetenzniveaus II und III anzusiedeln.

Bezüglich der Zusammenhänge mit Lernervariablen wurde in einem ersten Analyseschritt der Einfluss der individuellen Lernervoraussetzungen auf die Entwicklung der Kompetenz untersucht. Die genannten unabhängigen Variablen wurden nacheinander als Prädiktoren in Regressionsmodelle zur Vorhersage des Nachtestwertes aufgenommen. Um modellieren zu können, wie die Prädiktoren auf die Kompetenzentwicklung wirkten, wurde in diesen Modellen um den signifikanten Vortestwert kontrolliert (Ausgangsmodell zur Vorhersage des Nachtests durch den Vortest:  $\beta = .556$ ,  $p < 0.001$ ,  $R^2 = 30.9\%$ ).

Zunächst untersuchten wir den Einfluss des Leseverständnisses auf die Entwicklung der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“. Hier zeigte sich, dass ein hohes Leseverständnis nicht signifikant zur positiven Entwicklung der Kompetenz beitrug ( $\beta = .045$ ,  $p = 0.329$ ). Die entsprechende Regressionsgleichung findet sich zur Illustration in der folgenden Formel.

## Formel 6-1

Beispiel einer Regressionsgleichung zur Vorhersage des Nachtestwerts

$$\text{Kompetenz\_Nachtest} = \beta_0 + \beta_1 \cdot \text{Kompetenz\_Vortest} + \beta_2 \cdot \text{Leseverständnis} + \varepsilon$$

Dagegen wirkten kognitive Fähigkeiten ( $\beta = .166$ ;  $SE = .019$ ;  $p < 0.001$ ) signifikant auf die Entwicklung der Kompetenz: Höhere kognitive Fähigkeiten gingen mit einer überdurchschnittlichen Kompetenzentwicklung einher. Die Aufnahme von kognitiven Fähigkeiten in das Regressionsmodell erhöhte die Varianzaufklärung um  $\Delta R^2 = 2.9\%$  auf insgesamt  $33.8\%$ .

Zusätzlich zu kognitiven Fähigkeiten konnte ein signifikanter Zusammenhang zwischen der Entwicklung von Kompetenz und der Mathematiknote ( $\beta = -.102^5$ ;  $SE = .033$ ;  $p = 0.002$ ;  $\Delta R^2 = 0.8\%$ ) sowie dem Geschlecht ( $\beta = -.108$ ;  $SE = .042$ ;  $p = 0.010$ ;  $\Delta R^2 = 1.2\%$ ) gefunden werden. Schülerinnen und Schüler mit besseren Mathematiknoten verbesserten sich im Laufe der Intervention eher als ihre schwächeren Mitschüler – und dies auch über die kognitiven Fähigkeiten hinaus. Das negative Regressionsgewicht des Geschlechts drückt aus, dass sich Jungen in der Gesamtstichprobe durch die Intervention stärker verbesserten als Mädchen. Die zusätzliche Aufnahme der Mathematiknote und des Geschlechts erhöhte die Varianzaufklärung durch das Modell auf  $35.8\%$ .

### ***6.6.2 Der Einfluss der vier Treatments auf die Entwicklung der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“***

In einem weiteren Schritt wurde untersucht, inwieweit die Zugehörigkeit zu den jeweiligen Treatments über die bereits einbezogenen Variablen hinaus zur Entwicklung der Kompetenz beitrug. Da die Zuweisung zu den vier Gruppen „Reduktion“, „Variabilität“, „Reduktion und Variabilität“ sowie „Basis-Training“ parallelisiert nach Vortestleistung, sonst aber randomisiert erfolgt war, war nicht mit systematischen Zusammenhängen mit individuellen Voraussetzungen zu rechnen. Für die in die Analyse einbezogenen unabhängigen Variablen wurde dies auch durch korrelative Auswertungen bestätigt.

Die Leistungen der Treatmentgruppen „Reduktion“, „Variabilität“ und „Reduktion und Variabilität“ wurden jeweils mit denen des Treatments „Basis-Training“ verglichen. Es zeigten sich keine signifikanten Unterschiede zwischen dieser Referenzgruppe und den Gruppen „Variabilität“ und „Reduktion und Variabilität“. Dagegen verbesserten sich die Schülerinnen und

---

<sup>5</sup> Das negative Regressionsgewicht der Mathematiknote resultiert aus der Tatsache, dass im deutschen Notensystem eine niedrige Ausprägung (Note 1) der besten Leistung und eine hohe Ausprägung (Note 6) der schlechtesten Leistung zugeordnet werden.

Schüler der Treatmentgruppe „Reduktion“ signifikant weniger als diejenigen, die am Basis-Training teilnahmen (genaue Ergebnisse siehe Tab. 6-2). Durch Aufnahme der Treatmentzugehörigkeit in das bestehende Regressionsmodell erhöhte sich die Varianzaufklärung um  $\Delta R^2 = 0.6\%$  auf  $36.4\%$ .

Tabelle 6-2

Der Einfluss von individuellen Voraussetzungen und Treatmentvariablen auf die Entwicklung der Kompetenz

Prädiktor	Standardisiertes Regressionsgewicht	Standardfehler	p-Wert	$\Delta$ Varianzaufklärung
Vortest-Wert	.448	.045	< 0.001	30.9%
Kognitive Fähigkeiten	.168	.017	< 0.001	2.9%
Mathematiknote	-.107	.032	0.001	0.8%
Geschlecht	-.111	.044	0.012	1.2%
Treatment „Reduktion“	-.097	.042	0.022	0.6%
Treatment „Variabilität“	-.066	.054	0.223	/
Treatment „Reduktion und Variabilität“	-.048	.068	0.479	/

### ***6.6.3 Der Einfluss von individuellen Voraussetzungen und Treatmentvariablen auf Sichtweisen bezüglich zufallsbedingter Variabilität***

Neben der Wirkung von individuellen Voraussetzungen und Treatmentvariablen auf die Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ wurde ebenso untersucht, inwieweit diese Variablen die Sichtweisen auf zufallsbedingte Variabilität im Verlauf der Intervention beeinflussten. Auch hier wurde der Vortestwert bereits in das Ausgangsmodell zur Vorhersage des Nachtests aufgenommen, um durch anschließend hinzugefügte unabhängige Variablen Veränderungen der Zielvariablen über die Intervention hinweg erklären zu können. Außerdem ermöglichte diese Vorgehensweise die Kontrolle unterschiedlicher Ausgangsleistungen zwischen den vier Gruppen. Das Vortestergebnis klärte in diesem Ausgangsmodell ( $\beta = .632$ ,  $p < 0.001$ ) bereits  $39.9\%$  der Varianz des Nachtests auf. Die zusätzlich in das Regressionsmodell aufgenommenen individuellen Lernervariablen Leseverständnis, kognitive Fähigkeiten, Mathematiknote und Geschlecht stellten sich als nicht signifikant heraus.

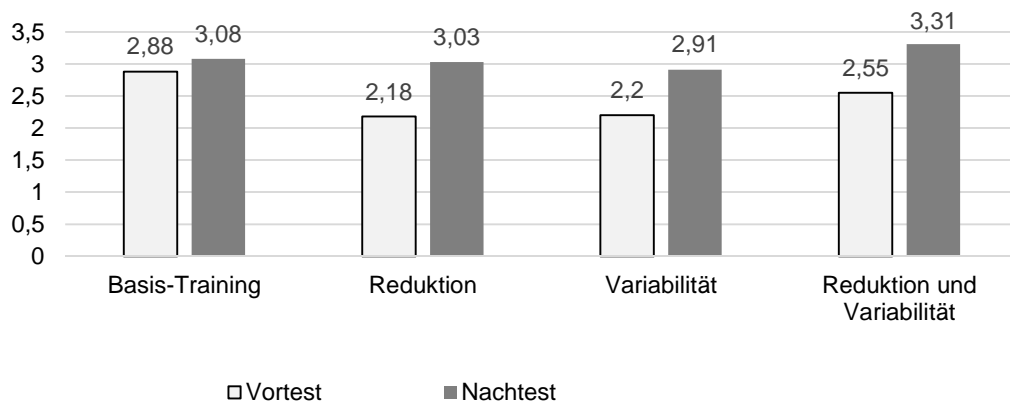
Im nächsten Schritt untersuchten wir den Einfluss der Treatments auf die Veränderung von Sichtweisen bezüglich zufallsbedingter Variabilität. Aus der deskriptiven Statistik in Abbil-

dung 6-3 ist ersichtlich, dass sich die Gruppenmittelwerte der drei Treatments „Reduktion“, „Variabilität“ und „Reduktion und Variabilität“ sowohl in den absoluten Ausprägungen als auch im Zuwachs untereinander ähnlicher sind als im Vergleich zur Gruppe „Basis-Training“.

Regressionsanalysen, die die vier Treatments einzeln miteinander verglichen, ergaben keine signifikanten Unterschiede unter diesen. Fasste man jedoch die reflexionsorientierten Treatments „Reduktion“, „Variabilität“ und „Reduktion und Variabilität“ zu einer Gruppe zusammen und verglich diese mit dem Basis-Training, so zeigte sich ein signifikanter Unterschied zugunsten der reflexionsorientierten Gruppe ( $\beta = .061$ ,  $p = 0.016$ ). Trotz der geringen Erhöhung der Varianzaufklärung um  $\Delta R^2 = 0.4\%$  auf  $40.3\%$  beträgt die Effektstärke  $d = 0.25$ , was einem kleinen Effekt (Cohen, 1988) entspricht.

Abbildung 6-3

Entwicklung von Sichtweisen auf Variabilität nach Treatmentgruppen getrennt



## 6.7 Diskussion

Ein zentrales Ergebnis dieser Studie ist die Erkenntnis, dass die Kompetenz der Lernenden aus der untersuchten Stichprobe auf Defizite beim Umgang mit statistischer Variabilität schließen lässt. Auf den Kompetenzniveaus IV und V waren die Lösungsraten gering, während viele Schülerinnen und Schüler auf den Niveaus I bis III die meisten Aufgaben richtig lösen konnten. Die Lösungsraten auf dem zweiten und dritten Niveau konnten durch die Intervention am deutlichsten erhöht werden. Die Aufgaben des ersten Niveaus wurden bereits weitgehend im Vortest richtig gelöst, wohingegen sich die Lösungshäufigkeiten der Aufgaben auf den höchsten beiden Niveaus im Verlauf der Intervention kaum veränderten. Dies kann so gedeutet werden, dass die Kompetenz der meisten Lernenden dieser Stichprobe auch nach erfolgter Intervention im Durchschnitt auf den Kompetenzniveaus II und III angesiedelt war. In einer Studie von Kuntze, Vargas, Martignon und Engel (2014) wurde gezeigt, dass Aufga-

ben von Niveau IV zu großen Teilen in der neunten Klasse gelöst werden können. Dies legt den Schluss nahe, dass sich innerhalb eines Schuljahres eine wesentliche Entwicklung der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ vollzieht und sich diese Kompetenz tatsächlich „schrittweise“ zu entwickeln scheint.

Bei der Untersuchung des Einflusses von individuellen Voraussetzungen auf die Entwicklung der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ zeigte sich, dass das Leseverständnis keinen signifikanten Einfluss hatte. Auch wenn dieses Ergebnis Auswertungen des Vortests dieser Studie (Sproesser et al., eingereicht) über Zusammenhänge zwischen Leseverständnis im Sinne des verwendeten Messinstruments und der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ sinngemäß bestätigt, so war es doch in Anbetracht der textgestützten Anlage der Lernmaterialien nicht unbedingt erwartungskonform. Wenngleich das verständnisvolle Lesen der Unterrichtsmaterialien eine notwendige Bedingung für diesbezüglichen Lernerfolg darstellt, so waren hier vermutlich andere Lernervoraussetzungen für die Entwicklung der Kompetenz entscheidend.

Signifikante Zusammenhänge mit der Kompetenzentwicklung ergaben sich dagegen für kognitive Fähigkeiten und die Mathematiknote. Insgesamt scheint der Zusammenhang mit der Mathematiknote leicht hinter dem mit kognitiven Fähigkeiten zurückzubleiben. Hierbei gilt es aber zu beachten, dass die relativ geringe zusätzliche Varianzaufklärung durch den Einschluss der Mathematiknote auch darauf zurückzuführen sein mag, dass die Mathematiknote signifikant mit den zuvor bereits ins Modell aufgenommenen kognitiven Fähigkeiten korreliert. Gemeinsame Varianzaufklärung wird bei dem gewählten Vorgehen dem zuerst einbezogenen Prädiktor zugeschrieben. Beide Variablen waren bereits mit dem Vortest korreliert und zeigten sich hier als über diesen hinaus bedeutsam für die Entwicklung der Kompetenz. Schülerinnen und Schüler mit hohen kognitiven Fähigkeiten beziehungsweise guten Mathematiknoten profitierten also etwas stärker von der Intervention als ihre schwächeren Peers. Auch wenn diese Zusammenhänge eher moderat ausfielen, sind sie dennoch aufgrund der Kürze der Intervention zu beachten. Auf eine längere Unterrichtseinheit übertragen könnte dies bedeuten, dass sich Leistungsunterschiede im Bereich von Statistical Literacy unter Lernenden mit unterschiedlichen Ausprägungen von kognitiven Fähigkeiten bzw. von Mathematikleistungen innerhalb eines Schuljahres deutlich verstärken könnten.

Des Weiteren zeigte sich, dass Jungen stärker von der vierstündigen Intervention profitierten als Mädchen. Jungen verfügten bereits vor der Intervention über höhere durchschnittliche Kompetenzscores (siehe Sproesser et al., eingereicht) und konnten die statistikbezogenen



Lernmaterialien auch stärker nutzen als Mädchen. Dieser Befund ist konsistent mit Ergebnissen aus einer Vielzahl an Mathematikleistungsstudien, die eine Verstärkung von Geschlechterunterschieden mit zunehmendem Alter bzw. mit zunehmender Beschulung feststellen (z. B. Klieme, 1997). Ob das Selbstkonzept bzw. Geschlechterstereotypen bezüglich des Unterrichtsfachs Mathematik hinter dem schlechteren Abschneiden der Mädchen stecken (vgl. z. B. Hannover, 2010), sollte in Anschlussforschung untersucht werden. Aufgrund der unterschiedlich starken Entwicklung von Kompetenz unter Mädchen und Jungen in dieser recht kurzen Zeit sowie aufgrund der Bedeutung von Kompetenz im Bereich von Statistical Literacy im Alltag ergibt sich ein besonderer Förderbedarf für Mädchen. Eine spezielle Förderung von Mädchen könnte beispielsweise durch Inhalte, die besonders für Mädchen interessant sind (vgl. z. B. Carmichael & Hay, 2009), erfolgen.

Vergleiche zwischen den Interventionsgruppen ergaben, dass die Treatments „Reduktion“, „Variabilität“ und „Reduktion und Variabilität“, die komplexe Reflexionen über wesentliche Aspekte von Statistical Literacy anregen sollten, zu keiner stärkeren Entwicklung von Kompetenz führten als das Basis-Training, das eher auf algorithmische Fähigkeiten des Umgangs mit statistikbezogenen Darstellungen und Modellen fokussierte. Die Treatmentgruppe „Reduktion“ wies sogar eine signifikant niedrigere Entwicklung von Kompetenz auf als das Basis-Training. Bei differenzierter Betrachtung der im Nachtest erzielten Punkte konnte festgestellt werden, dass nur wenige Schülerinnen und Schüler auf den beiden höchsten Kompetenzniveaus Punkte erreicht hatten. Da speziell die Aufgaben dieser Kompetenzniveaus ein Verständnis für Variabilität bzw. Fähigkeiten des datenbezogenen Reduzierens erforderten, scheint die spezielle Berücksichtigung dieser Aspekte in den reflexionsorientierten Treatments kaum zu höheren Lösungsraten entsprechender Items geführt zu haben. Gründe dafür könnten darin vermutet werden, dass die Intervention zu kurz war oder dass die Aufgaben der beiden höchsten Kompetenzniveaus für Lernende der achten Klassenstufe an Realschulen insgesamt zu anspruchsvoll beziehungsweise zu ungewohnt waren.

Durch die Beschäftigung mit den spezifischen Reflexionsaufgaben hatten die Lernenden der reflexionsorientierten Treatments vermutlich vergleichsweise weniger Zeit für eher grundlegende Aufgaben, wie sie auch auf den unteren drei Kompetenzniveaus angesiedelt waren. Zumindest für die Treatments „Variabilität“ sowie „Reduktion und Variabilität“ ergab sich aus diesen kürzeren Lernzeiten bezüglich grundlegender Aufgaben kein negativer Effekt. Gründe dafür, warum nur die Gruppe „Reduktion“ schlechter abschnitt, konnten aus den vorliegenden Daten nicht ermittelt werden. Insgesamt bleibt hervorzuheben, dass die niedrigen

Regressionsgewichte sowie der geringe Anteil an Varianzaufklärung eine im Vergleich zu individuellen Voraussetzungen eher geringe Bedeutung der Treatmentzugehörigkeit für den Kompetenzaufbau anzeigten.

Für die Entwicklung von Sichtweisen auf Variabilität spielten individuelle Voraussetzungen offenbar keine Rolle. Dagegen bewirkte die Zugehörigkeit zu einem der reflexionsorientierten Treatments „Reduktion“, „Variabilität“ oder „Reduktion und Variabilität“, dass zufallsbedingte Variabilität von den Lernenden nach der Intervention signifikant stärker berücksichtigt wurde. Die eher geringe zusätzliche Varianzaufklärung ( $\Delta R^2 \approx 0.4\%$ ) angesichts der Effektstärke von  $d \approx 0.25$  mag dadurch erklärt werden, dass der Vortestwert die Varianz des Nachtests in sehr viel stärkerem Maß erklärt als die Zugehörigkeit zu den Treatments. Auch innerhalb der reflexionsorientierten Treatments ist der Vortest stark prädiktiv für den Nachtest. Über den Einfluss des Vortests hinaus verzeichneten aber die Lernenden der reflexionsorientierten Treatments eine signifikant stärkere Veränderung in ihren Sichtweisen auf Variabilität. Offenbar sind die grundlegenden Ideen Reduktion und Variabilität inhaltlich so miteinander verwandt, dass alle reflexionsorientierten Treatments ähnliche Veränderungen bewirken konnten. Die Sichtweisen der Lernenden waren jedoch auch nach erfolgter Intervention noch relativ wenig von Variabilität geprägt. Dennoch zeichnet sich durch die Effekte der kurzen Intervention die Möglichkeit einer gezielten Förderung ab. Die reflexionsorientierten Treatments bewirkten also, dass die Sichtweisen der Schülerinnen und Schüler in diesen Gruppen nach der Intervention im Durchschnitt eine stärkere Berücksichtigung von zufallsbedingter Variabilität widerspiegeln.

Theoretische Überlegungen sprechen dafür, dass die hier eingesetzten Lernmaterialien einen Kompetenzzuwachs begünstigen können. In diesem Zusammenhang wäre es interessant zu erfahren, ob die geschilderte Intervention unter Lernenden, die eine gewisse Verständnisschwelle bereits überwunden haben (wie beispielsweise die Neuntklässler der Studie von Kuntze et al., 2014), differenziertere Effekte bewirken kann. Da außerdem die grundsätzliche Bedeutung des Umgangs mit Variabilität sowie von Fähigkeiten der Datenreduktion für Statistical Literacy unbestritten ist, sollten Unterrichtskonzepte, die eine längere beziehungsweise regelmäßig wiederkehrende Beschäftigung mit diesen Aspekten beinhalten, empirisch erforscht werden.

## **6.8 Infobox Forschungsmethoden**

### ***6.8.1 Zuweisung zu den Treatmentgruppen***

Das Design der vorliegenden Studie sah die Einteilung der Schülerinnen und Schüler jeder Klasse (insgesamt  $n = 450$ ) in vier Treatmentgruppen vor. Dadurch sollte überprüft werden, ob Reflexionsanlässe bezüglich datenbezogener Reduktion oder/und statistischer Variabilität die Entwicklung von Kompetenz im Bereich Statistical Literacy positiver beeinflussten als ein Basis-Training. Dieses beschränkte sich auf eher grundlegende Aufgaben des Umgangs mit Tabellen, Diagrammen und Kennwerten, wie sie auch in Schulbüchern zu finden sind. Die Einteilung der Lernenden in die Treatments erfolgte parallelisiert nach deren Kompetenzscore zum ersten Messzeitpunkt, sonst aber randomisiert: Zunächst wurden die im Vortest besten vier Schülerinnen und Schüler jeder Klasse ermittelt und zufällig auf die vier Treatments verteilt. Dieser Prozess wurde so lange wiederholt, bis alle Probanden einem Treatment zugeordnet waren. Alternativ hätten auch ganze Klassen an je einem Training teilnehmen können, was die Möglichkeit gemeinsamer Lernphasen geboten, aber auch Klasseneffekte verstärkt hätte. Um insbesondere den Einfluss der Lehrperson zu kontrollieren, erschien die Unterteilung jeder Klasse in vier Treatmentgruppen am sinnvollsten zu sein.

### ***6.8.2 Korrelations- und Regressionsanalysen***

Durch den Korrelationskoeffizienten  $r$  kann die Stärke und Richtung ( $-1 \leq r \leq 1$ ) eines statistischen Zusammenhangs zwischen zwei numerischen Variablen getestet werden. Werden zwei Variablen im Streudiagramm gegeneinander abgetragen, so kann die Größe des Zusammenhangs im Sinne einer Korrelation daran festgemacht werden, wie eng die einzelnen Datenpunkte um eine Gerade streuen. Dabei wird zunächst von der Nullhypothese ausgegangen, die besagt, dass es keinen systematischen Zusammenhang zwischen den untersuchten Variablen gibt ( $r = 0$ ). Der  $p$ -Wert gibt nun an, wie wahrscheinlich es unter der Nullhypothese ist, den beobachteten oder einen noch stärkeren Zusammenhang zufällig zu finden. Laut Konvention muss die Irrtumswahrscheinlichkeit für das Verwerfen der Nullhypothese, also der  $p$ -Wert, kleiner als 5 % sein, um einen Zusammenhang als statistisch signifikant einzustufen.

Eine Regression liefert weitere Informationen über den Zusammenhang zwischen zwei numerischen Variablen, denn sie beziffert die Steigung einer (Regressions-)Geraden in der Punktwolke. Am Regressionskoeffizienten  $\beta$  kann abgelesen werden, wie stark sich die abhängige Variable verändert, wenn die unabhängige Variable um eine Einheit erhöht wird. Zusätzlich liefert der Standardfehler Informationen über die Genauigkeit dieser Schätzung, wobei kleinere Standardfehler genauere Parameterschätzungen anzeigen. Da große Stichproben zu präzise-

ren Parameterschätzungen führen, fallen Standardfehler bei steigender Stichprobengröße geringer aus. Analog zur Korrelationsanalyse gibt der p-Wert an, ob der gefundene Zusammenhang als statistisch signifikant angesehen werden kann oder nicht.

Multiple Regressionen beziehen mehrere unabhängige Variablen zur Vorhersage einer abhängigen Variablen ein. Dabei gewährleistet der Ceteris-Paribus-Effekt, dass jeder Regressionskoeffizient um die anderen einbezogenen Variablen kontrolliert wird. In der vorliegenden Studie können folglich durch den Einbezug des Vortestwerts die Unterschiede in den Ausgangswerten zwischen den Treatmentgruppen bereinigt werden, was eine Kovarianzanalyse mit Messwiederholung als alternative Auswertungsmethode nicht hätte leisten können. Zudem sind Aussagen folgender Art möglich (bezüglich Tab. 6-2): „Je zwei Lernende mit gleichen Vortestwerten und gleicher Mathematiknote eines Geschlechts aus der gleichen Treatmentgruppe, die sich um eine Einheit in ihren kognitiven Fähigkeiten unterscheiden, weisen Unterschiede von 0.168 Einheiten im Nachtestwert auf.“ So kann für jede unabhängige Variable einzeln angegeben werden, in welcher Größenordnung sie zur Entwicklung von Kompetenz beiträgt.

In der vorliegenden Studie wurde eine besondere Art der Regression verwendet, da die Teilnehmenden in Schulklassen und Schulen hierarchisch genestet sind. Diese Strukturen können zu Verzerrungen der Auswertungen führen, wenn sich Lernende einer Klasse ähnlicher sind als unabhängige, zufällige Beobachtungen. Standardfehler werden dann unterschätzt, was zu liberale Signifikanztests nach sich ziehen kann (Raudenbush & Bryk, 2002). Da jedoch die systematischen Unterschiede zwischen den Klassen, die Intraclass-Korrelation mit 0.4 % für den Kompetenztest und 0.6 % für Sichtweisen bezüglich Variabilität (jeweils Testzeitpunkt 2) sehr gering ausfielen und die Fragestellung keine Variablen auf der Klassenebene einbezog, erschien die Auswertung über Mehrebenenregressionen als nicht zielführend. Das Statistikprogramm Mplus 7.1 (Muthén & Muthén, 1998 – 2012) ermöglicht es, die Abhängigkeitsstruktur der Daten durch Adjustierung der Standardfehler zu berücksichtigen. Zusätzlich können mit Hilfe des hier verwendeten robusten Maximum-Likelihood-Schätzers auch Variablen einbezogen werden, die keine exakte Normalverteilung aufweisen. Fehlende Werte der einbezogenen Variablen (hier maximal 1.3 %) können mit dem Full-Information-Maximum-Likelihood-Verfahren geschätzt und so der gesamte Datensatz genutzt werden.

## **6.9 Infobox Implikationen für die Praxis**

Eine Vielzahl theoretischer Arbeiten betont die Bedeutung des situationsangemessenen Einschätzens von Variabilität für Statistical Literacy (siehe Theorieteil). Beispielsweise muss

beim Treffen von Entscheidungen auf Grundlage von durch Daten gegebener Information erkannt werden, ob Unterschiede in diesen Daten zufälligen Schwankungen geschuldet sind oder systematische Ursachen haben, die insofern bei der Entscheidungsfindung beachtet werden sollten. Zur kritischen Beurteilung fremder Entscheidungen oder Daten-Interpretationen ist ebenso ein Verständnis für Variabilität erforderlich. Empirische Befunde zeigen jedoch, dass viele Menschen eine Vorliebe für regelmäßige Datenstrukturen und deterministische Erklärungen haben, die der Entwicklung eines Verständnisses für Variabilität und somit auch der Entwicklung von Statistical Literacy im Wege stehen können. Verschiedene Forscher sehen die Ursache dafür im klassischen Unterricht, der Zufall und Variabilität zu wenig berücksichtigt und den Lernenden vermittele, dass alles einen genau zu ermittelnden Grund habe. In dieser Studie wurden Schülerinnen und Schüler im Rahmen der Bearbeitung von Problemstellungen in statistischen Kontexten (Basis-Training) teilweise dazu aufgefordert, direkt (Treatments „Variabilität“ sowie „Reduktion und Variabilität“) oder indirekt (Treatment „Reduktion“) über Variabilität zu reflektieren.

Im Verlauf der vierstündigen Intervention veränderten sich die Sichtweisen der Lernenden in den reflexionsorientierten Treatments im Vergleich zu den Lernenden des Basis-Trainings signifikant. Die Lernenden der Treatments „Reduktion“, „Variabilität“ sowie „Reduktion und Variabilität“ entwickelten dabei Sichtweisen, die zufallsbedingte Variabilität bei der Einschätzung zufälliger Erscheinungen stärker berücksichtigten. Der hier in kurzer Zeit erzielte Effekt legt es nahe, dass die regelmäßige Auseinandersetzung mit Zufall und Variabilität Schülerinnen und Schülern zu situationsangemessenen Einschätzungen und Sichtweisen bezüglich zufälliger Phänomene verhelfen kann.

#### **6.10 Danksagung und Deklaration von Interessenkonflikten**

Ute Sproesser ist Mitglied des Kooperativen Promotionskollegs „Effektive Lehr-Lernarrangements“ der Pädagogischen Hochschule Ludwigsburg und der Universität Tübingen, das vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg gefördert wird. Zusätzlich wurde die vorliegende Studie mit Forschungsgeldern der PH Ludwigsburg unterstützt.

## 6.11 Literatur

- Ben-Zvi, D. & Garfield, J. (2005). Statistical literacy, reasoning and thinking: Goals, definitions and challenges. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking*. New York: Kluwer.
- Blum, W. & Leiss, D. (2005). Modellieren im Unterricht mit der „Tanken“-Aufgabe. *mathematik lehren*, 128, 18-21.
- Brunner, M. (2005). *Mathematische Schülerleistung: Struktur, Schulformunterschiede und Validität*. Humboldt Universität zu Berlin. Zugriff am 20.05.2014. Verfügbar unter <http://edoc.hu-berlin.de/dissertationen/brunner-martin-2006-02-08/PDF/brunner.pdf>.
- Carmichael, C. & Hay, I. (2009). Gender differences in middle school students' interest in a statistical literacy context. In R. Hunter, B. Bicknell, & T. Burgess (Eds.), *Crossing divides: Proceedings of the 32nd annual conference of the Mathematics Education Research Group of Australasia* (Vol. 1). Palmerston North, NZ: MERGA.
- Cohen, J. (1988): *Statistical power analysis for the behavioral sciences*, 2. Edition, Hillsdale: Lawrence Erlbaum Associates.
- Curcio, F.R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for research in mathematics education*, 18(5), 382 – 393.
- Engel, J. (2011). Über ikonische Repräsentationen von zufallsbedingter Variabilität. In R. Haug & L. Holzäpfel, *Beiträge zum Mathematikunterricht 2011* (S. 235-238). Münster: WTM.
- Engel, J. & Sedlmeier, P. (2005). On middle-school students' comprehension of randomness and chance variability in data. *Zentralblatt für Didaktik der Mathematik*, 37(3), 168-177.
- Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Reidel: Dordrecht.
- Fröhlich, A., Kuntze, S. & Lindmeier, A. (2007). Testentwicklung und -evaluation im Bereich von „Statistical Literacy“. *Beiträge zum Mathematikunterricht 2007* (S. 783-786). Hildesheim: Franzbecker.
- Green, D.R. (1990). *A longitudinal study of pupils' probability concepts*. Loughborough: Loughborough University.

- Green, D.R. (1986). Children's understanding of randomness. In R. Davidson & J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics* (pp. 287-291). Victoria, British Columbia: University of Victoria.
- Green, D.R. (1982). A survey of probabilistic concepts in 3000 students aged 11-16 years. In D.R. Green (Ed.), *Proceedings of the First International Conference on Teaching Statistics* (pp. 766-783). University of Sheffield: Teaching Statistics Trust.
- Gundlach, M., Kuntze, S., Engel, J., & Martignon, L. (2010). Motivation and self-efficacy related to probability and statistics: Task-specific motivation and proficiency. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Zugriff am 14.06.2013. Verfügbar unter <http://www.stat.auckland.ac.nz/~iase/publications.php>.
- Hannover, B. (2010). Lernen Mädchen anders? In M. Matzner & I. Wyrobnik (Eds.), *Handbuch Mädchen-Pädagogik* (S. 95-107). Weinheim: Beltz.
- Heller, A. K. & Perleth, C. (2000). *KFT 4-12+R. Manual*. Göttingen: Hogrefe.
- Hosenfeld, I., Köller, O. & Baumert, J. (1999). Why sex differences in mathematics achievement disappear in German secondary schools: A reanalysis of the German TIMSS-data. *Studies in Educational Evaluation*, 25, 143-162.
- Klieme, E. (1997). Gender-related differences in mathematical abilities: Effect-size, spatial mediation, and item content. *Paper presented at the 7<sup>th</sup> Conference of the European Association for Research in Learning and Instruction (EARLI)*, Athens, Greece.
- Klieme, E., Neubrand, M. & Lüdtke, O. (2001). Mathematische Grundbildung: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K. J. Tillmann & M. Weiß (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 141-191). Opladen: Leske + Budrich.
- Konold, C. & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education* 33(4), 259-289.
- Kröpfl, B., Peschek, W., Schneider, E. (2000). Stochastik in der Schule: Globale Ideen, lokale Bedeutungen, zentrale Tätigkeiten. *mathematica didactica*, 23, 25 – 57.

- Kuntze, S. (2013). Modellieren beim Nutzen von Darstellungen in statistischen Kontexten. Hierarchische Beschreibung und Bedingungsvariablen eines Aspekts mathematischer Kompetenz. In R. Borromeo Ferri, G. Greefrath & G. Kaiser (Hrsg.), *Mathematisches Modellieren in Schule und Hochschule* (S. 71-94). Wiesbaden: Springer Spektrum.
- Kuntze, S. (2012). Sichtweisen von Lernenden zu statistischer Variabilität – Vorstellungen von Grundschüler(inne)n, Realschüler(inne)n und Studierenden. In M. Ludwig & M. Kleine (Hrsg.), *Beiträge zum Mathematikunterricht 2012* (S. 505-508). Münster: WTM.
- Kuntze, S., Engel, J., Martignon, L. & Gundlach, M. (2010). Aspects of statistical literacy between competency measures and indicators for conceptual knowledge – Empirical research in the project RIKO-STAT. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Zugriff am 14.06.2013. Verfügbar unter [www.stat.auckland.ac.nz/~iase/publications.php](http://www.stat.auckland.ac.nz/~iase/publications.php).
- Kuntze, S., Lindmeier, A. & Reiss, K. (2008). “Using models and representations in statistical contexts” as a sub-competency of statistical literacy – Results from three empirical studies. *Proceedings of the 11th International Congress on Mathematical Education*. Monterrey (Mexico). Zugriff am 31.07.2014. Verfügbar unter <http://tsg.icme11.org/document/get/474>.
- Kuntze, S., Vargas, F., Martignon, L. & Engel, J. (2014). Competencies in understanding statistical information in primary and secondary school levels: An inter-cultural empirical study with German and Colombian students. *Avances de Investigación en Educación Matemática (AIEM)* 7, 5-25.
- Muthén, L.K. and Muthén, B.O. (1998-2012). *Mplus user's guide*. Seventh Edition. Los Angeles, CA: Muthén & Muthén.
- Piaget, J. & Inhelder, B. (1975). *The origin of the idea of chance in children*. London: Routledge & Kegan Paul.
- Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical linear models* (2<sup>nd</sup> ed.). Thousand Oaks: Sage Publications.



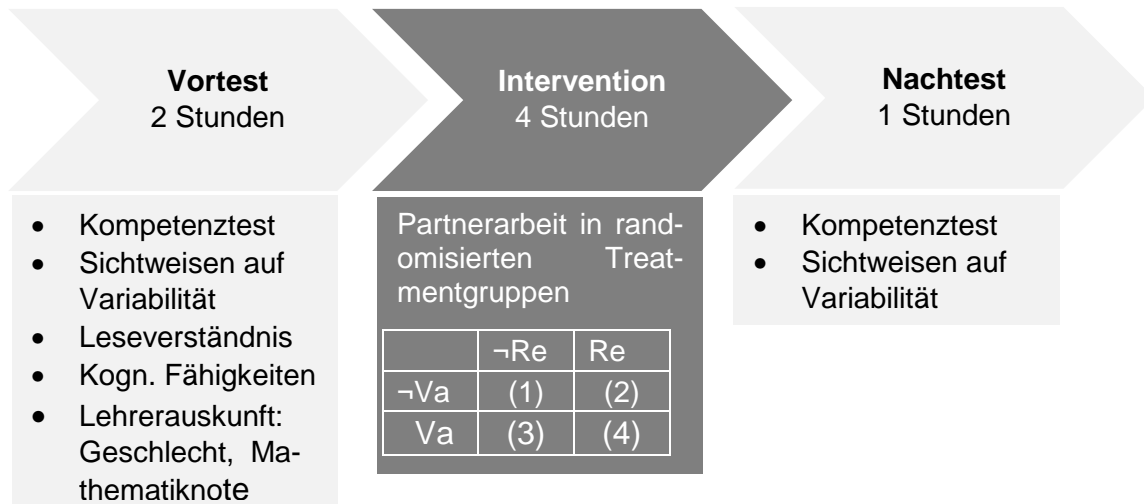
- Reading, C. (2002). Profile for statistical understanding. In B. Philips (Ed.), *Proceedings of the 6<sup>th</sup> Int.l Conference on Teaching Statistics: Developing a statistically literate society, Cape Town, South Africa*. Voorburg, the Netherlands: International Statistical Institute.
- Schneider, W., Schlagmüller, M. & Ennemoser, M. (2007). *LGVT 6-12. Manual*. Göttingen: Hogrefe.
- Sproesser, U., Engel, J. & Kuntze, S. (angenommen). Fostering self-concept and interest for statistics through statistics-specific learning environments. *Statistics Education Research Journal*..
- Sproesser, U., Kuntze, S. & Engel, J. (eingereicht). Is statistical literacy interrelated to reading comprehension and cognitive abilities? A multilevel analysis.
- Tent, L. (2001). Zensuren. In D.H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 805-811). Weinheim: Psychologie Verlags Union.
- Wallman, K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, 88(421), 1-8.
- Watson, J. M. & Callingham, R. A (2003). Statistical literacy: a complex hierarchical construct. *Statistics Education Research Journal* 2, 3-46.
- Wild, C.J. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review* 67, 223-265.

## 6.12 Elektronisches Supplement

Das elektronische Supplement ist mit der Online-Version dieses Artikels verfügbar unter <http://dx.doi.org/10.1024/2235-0977/a000099>.

Abbildung 6-4

Entwicklung von ESM 1. Word-Dokument. Zeitlicher Ablauf der Studie.



## 7 Teilstudie 3

### Fostering self-concept and interest for statistics through statistics-specific learning environments

A revised version of this article was accepted for publication by the Statistics Education Research Journal. This manuscript was integrated in this doctoral thesis with permission of the Co-Editor Maxine Pfannkuch.

Sproesser, U., Engel, J. & Kuntze, S. (accepted). Fostering Self-concept and Interest for Statistics Through Statistics-Specific Learning environments. *Statistics Education Research Journal*.

#### 7.1 Abstract

Supporting motivational variables such as self-concept or interest is an important goal of schooling because they relate to learning and achievement. In this study, we investigated if specific interest and self-concept related to the domains of statistics and mathematics could be influenced through a four-lesson intervention focusing on statistics. Data about these motivational variables and competency related to statistics were gathered from 450 eighth graders before and after the intervention. Our results indicate that students perceive mathematics and statistics differently in respect to their self-concept and interest. Moreover, statistics-related self-concept and interest could be fostered through the domain-specific intervention, whereby a greater increase was found among students with higher prior competency in the domain of statistics.

**Keywords** *statistics education research, self-concept, interest, intervention study*

## **7.2 Introduction**

Whereas affective and motivational variables were rather neglected for a long time in overall mathematics education (McLeod, 1992), more and more attention has been paid to them within the last decades. Motivational variables such as self-concept and interest influence learners' willingness to engage in and to maintain learning activities (Nagy, Trautwein, Baumert, Köller, & Garrett, 2006) and they are now considered to be important determinants of learning and achievement (Pekrun & Zirngibl, 2004).

Concerning the domain of statistics, Gal, Ginsburg and Schau (1997) underline the importance of motivational variables because of their impact on students' learning processes and their willingness to use statistics in their lives as well as on their educational choices. However, within statistics education the relevance attributed to affective and motivational variables is still rather low and only a few studies have investigated these variables related to this domain compared to mathematics education (Bond, Perkins & Ramirez, 2012; Ramirez, Schau & Emmioglou, 2012). Although Hood, Creed and Neumann (2012) emphasize the necessity to foster such motivational variables in particular for learners with rather poor competency in the domain of statistics, to our knowledge, there is hardly any quantitative empirical evidence about the support of secondary students' self-concept and interest in the domain of statistics through specific learning environments.

Consequently, this study examines whether self-concept and interest in mathematics and in statistics can be supported by a statistics-oriented and learner-centered intervention among eighth graders. As students often associate statistics with mathematics (Gal et al., 1997), we first investigated if middle school students perceived these two domains as separate with respect to their interest and self-concept. Then we explored whether self-concept and interest increased during the course of the domain-specific intervention. To get further insight about which characteristics of learners particularly favor a positive development of these motivational variables, we additionally assessed if previous statistics-related competency was related to the development of self-concept and interest in the corresponding domain. We introduce the theoretical background leading to the research questions in the next section.

## **7.3 Theoretical background**

Motivational variables such as self-concept and interest accompany the learning process and influence learners' developmental achievement at school and beyond (cf. Eccles & Wigfield, 2002; Helmke & Weinert, 1997; Jacobs, Lanza, Osgood). For instance, confidence in one's own competencies determine if learning activities are performed even beyond external obliga-

tions. In this regard, it is not surprising that there are various reciprocal effects between competency and motivational variables and that - besides fostering achievement - the development of students' motivational variables is considered to be an essential goal of schooling (O'Mara, Marsh, Craven & Debus, 2006; Pekrun & Zirngibl, 2004; Schiefele, 1991). Such motivational variables particularly mediate learning and achievement when they relate to a specific domain. They were therefore assessed domain-specifically in studies such as PISA or TIMSS (e.g. Artelt, Baumert, Julius-McElvany & Jules Peschar, 2003; Baumert, Bos & Lehman, 2000).

The present paper focuses on fostering statistics-related self-concept and interest. Therefore in the following sections, self-concept and interest will be introduced and described in particular with respect to the specific domains of mathematics and statistics.

### ***7.3.1 Self-concept***

Within the field of self-competency beliefs that represent an individual's conviction "about his or her attributes and abilities as a person" (Valentine, DuBois & Cooper, 2004, p. 112), self-concepts usually refer to distinct academic or non-academic domains such as school subjects or physical abilities. In the case of mathematics, (academic) self-concept expresses a person's confidence in his or her competencies in this particular domain. Whereas general self-concept integrates all academic and non-academic forms of self-concepts (Shavelson, Hubner & Stanton, 1976), self-efficacy, which is the conviction to master particular tasks (Bandura, 1977), is related to even more specific contents than domain-specific academic self-concept. These constructs mainly differ in the specificity of the domain they are related to.

Self-competency beliefs such as domain-specific academic self-concepts are considered to be positively related to motivation and achievement (Eccles & Wigfield, 2002; Jacobs et al., 2002). However, the level of specificity is crucial for the magnitude of the correlation between self-concept and achievement. That is why Marsh and Craven (1997) advise researchers to focus on self-concept related to the domains of their particular research. Hansford and Hattie (1982) for instance report findings of a meta-analysis indicating that domain-specific self-concept correlates at an average of  $r = .42$  with related achievement compared to  $r \leq .22$  for global self-concept.

Valentine and colleagues (2004) mention several reasons for the link between (domain-specific) self-competency beliefs and achievement: Learners with a more positive academic

self-concept are more likely to engage in and to hold on to success-oriented behaviors. For instance, they spend more time practicing and therefore get more corrective feedback and they choose more challenging tasks and develop better ways of coping with failure leading to more persistence and therefore better school performance. In contrast, individuals with a less positive academic self-concept are likely to develop avoidance strategies. Helmke and Weinert (1997) add that a negative self-concept can lead to negative attributions, deficient learning behaviors, consequently having a negative impact on achievement.

Because of the relationship between self-concept and learning, several studies have aimed at fostering this motivational variable. O'Mara and colleagues (2006) found in an extensive meta-analysis that self-concept interventions are particularly effective when they refer to the same domains that are related to the measured self-concept. In this case, the increase remained relatively stable over time. Interventions focusing particularly on praise and feedback but also on developing domain-specific skills have been successful in fostering academic self-concepts.

Furthermore, as academic self-concept is related to interest in the corresponding domains (Valentine et al., 2004; Pekrun & Zirngibl, 2004), the current study also investigated the development of interest over the course of the intervention. Hence, the next section outlines the theoretical background related to interest.

### **7.3.2 Interest**

Interest is a disposition with both trait and state characteristics. Whereas situational interest refers to particular situations and is characterized by being relatively fluctuating, individual interest – for instance related to school subjects like mathematics – is considered to be relatively stable over time (Frenzel, Goetz, Pekrun & Watt, 2010)., Schiefele (1991) defines individual interest as a “relatively enduring preference for certain topics, subject areas, or activities”. This quotation also indicates that interest represents a specific relationship between an individual and a particular content, hence interest is a domain-specific variable.

Schiefele (1992) identifies two central aspects of interest: First, individuals attribute a personal importance to their objects of interest (value-related valences). Furthermore, interest is typically accompanied by involvement and enjoyment (feeling-related valences) and therefore can be considered as a trigger for intrinsic motivation (Hidi & Renninger, 2006; Schiefele, 1991). In particular, both interest and intrinsic motivation have no external incentive such as

striving to succeed on an exam. An interested or intrinsically motivated individual is involved in an activity or topic for its own sake and therefore strives to learn about it.

As interest leads to engagement and mostly to learning activities, it may also enhance achievement (Schiefele, 1991). Helmke and Weinert (1997) found a mean correlation between interest and school achievement of  $r = .41$  within a meta-analysis. Schiefele gives several reasons for the positive relationship between interest and achievement: Highly interested students are more likely to engage in intense and meaning-oriented learning activities. Furthermore, they use more elaborated learning strategies and invest more time and effort on learning. In the end, these students report a higher quality of their learning experience and corresponding positive emotions.

The relationship between interest and achievement underlines the importance of paying attention to and even fostering students' interest by specific learning activities. Hidi and Renninger (2006) suggest that interest for a specific topic may even grow within classrooms, for example by dealing with the topic. Teachers can support students' interest, for instance, by providing helpful feedback, choosing challenging and interesting tasks, enabling social learning and promoting a sense of autonomy.

In the next section, we will outline the theoretical background related to these motivational variables in the specific domains of mathematics and statistics.

### ***7.3.3 Self-concept and interest: the case of mathematics and statistics***

Similar to many other countries, statistics in German schools is by and large taught as part of mathematics education. Therefore, it is often considered to be a pure subdomain of mathematics. According to Gal and colleagues (1997), students may transfer their perceptions of mathematics to the domain of statistics. However, students may not necessarily perceive statistics to be an exclusive part of mathematics because of the inherent interdisciplinary nature of statistics and the dominant role of the context therein (Cobb & Moore, 1997). From this point of view, statistics and mathematics may rather represent two distinct domains with a particular overlap. In this study, we assume that students of secondary school can hold different perceptions of mathematics and statistics from a motivational point of view. In order to take into account that assessing motivational variables related to statistics should first discriminate between students' perceptions of this domain and the domain of mathematics (Gal et al., 1997), we will thus check if this assumption is reflected by the empirical data. In a first analysis step, we will investigate by confirmatory factor analysis if academic self-concept and in-

terest related to the domains of mathematics and statistics are separable factors. Moreover, if students actually perceive mathematics and statistics differently as far as academic self-concept and interest is concerned, the statistics-related intervention may - according to the theory - primarily foster motivational variables referring to the domain of statistics.

Motivational variables concerning mathematics have been investigated by a large number of studies including PISA and TIMSS. According to Watt (2004), self-concept and interest referring to mathematics are correlated ( $r = .55$ ). Klieme, Neubrand and Lüdke (2001) even report a correlation of  $r = .7$  within the PISA sample. Baumert and Köller (2000) found high correlations between achievement and self-concept ( $r = .73$ ) as well as between achievement and interest ( $r = .69$ ) in the TIMSS sample. A study by Marsh, Trautwein, Lüdtke, Köller and Baumert (2006) confirms this relationship between mathematics-related self-concept and achievement ( $.51 \leq r \leq .77$ ) whereas correlation coefficients near zero ( $-.03 \leq r \leq .05$ ) were found between general or non-academic self-concept and achievement score. Even if self-concept and interest are relatively stable over time (Möller & Trautwein, 2009; Schiefele, 1991), several studies (cf. Frenzel et al., 2010; Jacobs et al., 2002; Watt, 2004) report a significant decline over the course of schooling for both variables related to mathematics.

Although several researchers (Bond et al. 2012; Carmichael, Callingham, Watson & Hay, 2009; Gal et al., 1997; Hood, Creed & Neumann, 2012; Ramirez et al., 2012) consider paying attention to or fostering motivational variables in the domain of statistics to be important goals of statistics education, only a few studies have investigated these variables in regard to statistics (Bond et al., 2012; Carmichael et al., 2009; Ramirez et al., 2012). Emmioglu and Capa-Aydin (2012) found a positive relationship between achievement and motivational variables in the domain of statistics in a meta-analysis among post-secondary students. A study by Schau and Emmioglu (2012) indicated that students' statistics-related self-concept and interest remained stable or became less positive in the course of post-secondary introductory statistics courses (see also Bond et al., 2012). Concerning secondary students, Carmichael, Callingham, Hay and Watson (2010) investigated learners' statistics-related motivational variables from grade 7 to grade 9. Within their sample, they determined a decline with increasing age for interest. Furthermore, they found significant relationships between self-efficacy, interest and prior mathematics achievement that positively especially influenced self-efficacy but also interest in the domain of statistics.

Even if the mentioned studies examined (secondary) students' motivational variables related to statistics, they do not provide empirical evidence about the support of domain-specific self-



concept and interest. Furthermore, there is no insight into whether learners' prior achievement is related to the development of these motivational variables over time. As this study focused on fostering students' domain-specific self-concept and interest through a statistics-related intervention, we outline the concept of this intervention in the next section.

#### ***7.3.4 The statistics-related intervention of the current study***

According to the German standards referring to mathematics instruction (KMK, 2003: “overarching idea”: data and chance), students ought to learn about data collection, data representation, data reduction, data interpretation and probability (cf. Holmes, 1980). Concerning these topics, textbooks often provide tasks requiring rather algorithmic activities related to tables, bar charts or characteristic values such as computing an average. However, statistical literacy, defined as “the ability to interpret, critically evaluate and communicate about statistical information and messages” (Gal, 2002, p. 1), encompasses more. One important reason why statistics is widely needed and used is the inherent variability or lack of uniformity among all observable objects (e.g. Wild & Pfannkuch, 1999). Dealing adequately with this statistical variation constitutes one key aspect within statistical literacy (Watson & Callingham, 2003). It implies separating systematic from non-systematic variability, i.e. recognizing if observed differences are due to random or if they are caused by a systematic effect (Wild & Pfannkuch, *ibid.*). Complementary to considerations of variability is the efficient reduction of information of data to a numerical or graphical summary. Such data reduction omits several details of the original data that often contain a large magnitude of relevant and irrelevant information and therefore reduced data mostly provide better overviews (Kröpfl, Peschek & Schneider, 2000). Abilities in the domain of data reduction include an awareness about the associated loss of information, (e.g. smoothing out random variability through forming averages (Kröpfl. et al., *ibid.*)).

These key aspects of statistical literacy were included in the intervention of the present study. Students were assigned to one of four treatments that involved (1) exclusively basic elements of statistics such as dealing with tables, bar charts and characteristic values or going beyond such basic elements and (2) considerations of variation or (3) of data reduction or (4) of both (2) and (3). These treatments were primarily designed to have an impact on students' competency of using representations and models in statistical contexts (e.g. Kuntze et al., 2008), which we refer to as “statistics-related competency”. However, it is possible that these four focuses of the learning environments may also cause differences in developing students' domain-specific self-concept and interest. Therefore, beyond the influence of the statistics-

related intervention as a whole, the impact of the four different treatments on domain-specific self-concept and interest was investigated in this study.

All four treatments provided learner-centered and hands-on activities in contexts relevant for eighth graders. During the whole intervention, students worked with a partner on statistics-related problems and had the possibility to get help from flash cards as well as to verify their answers with sample solutions. Upon finishing a topic, the students were given additional individual feedback. The intervention was implemented as such because several studies have shown that student-centered learning environments that provide feedback and autonomy in cooperative learning situations can promote motivational variables (Gläser-Zikuda, Fuß, Laukenmann, Metz, & Randler, 2005; Hattie, Biggs, & Purdie, 1996; Hidi & Renninger, 2006; O'Mara et al., 2006; Schukajlow et al, 2012; Schunk & Zimmermann, 2003).

#### **7.4 Research questions**

The previous sections emphasized the general relevance of self-concept and interest for learning and achievement. However, little is known about these motivational variables related to the domain of statistics - especially, whether they can be fostered through a learning intervention.

Before investigating motivational variables related to statistics, Gal and colleagues (1997) recommend that researchers discriminate between students' perceptions of this domain and the domain of mathematics. Consequently, after having explored whether students perceive statistics and mathematics differently from a motivational point of view, we investigated whether the intervention was adequate to support students' mathematics- and statistics-specific self-concept and interest. As the design of the study provided four contrasting treatment groups, the potential impact of these different treatments could be analyzed.

Therefore, we investigated the following research questions:

- 1) To what extent are mathematics- and statistics-related self-concept and interest empirically separable?
- 2) Can students' domain-specific self-concept and interest be fostered by a statistics-oriented intervention? If so, are there differences with respect to the effect of the four treatments?

As domain-specific self-concept and interest are generally related to competencies in the corresponding domain, students' prior statistics-related competency might determine to what

extent these motivational variables may develop during the intervention. In this sense, the third research question is:

- 3) Which role does learners' prior competency of using models and representations in statistical contexts play for the development of their statistics-related self-concept and interest?

## **7.5 Methods**

### ***7.5.1 Design and sample***

The present study is part of the project ReVa-Stat ("Developing concepts of data-related reduction and statistical variation as a support for building up statistical literacy"). Data from 450 German eighth graders (212 female, 238 male), who completed pre- and posttests and participated in the intervention were the basis of the analysis. The students were between 12 and 15 years old (M 13.50; SD 0.62) and were recruited from 25 classes in eight German technical-track secondary schools (so-called "Realschule"). To investigate research question 2, a baseline group consisting of 53 students (32 male, 21 female, between the ages of 13 and 15 years) from four additional classes who did not take part in the intervention was included in the study. An analysis of pretest scores revealed no significant differences between baseline and intervention group regarding pretest competency scores and statistics-related self-concept and interest. Over a period of four 45-minute classroom lessons, the students participating in the intervention worked on learner-centered material dealing with different statistics-relevant topics. In each class, students were split up into four subgroups according to a 2x2-design. Assignment to the four different subgroups was randomized and parallelized regarding their pretest competency score. Students in group 1 (n = 117) were assigned basic statistical problems such as dealing with tables, bar charts and characteristic values. On the basis of these basic problems, students in the other three groups were additionally asked to respond to prompts that required them to reflect about (2) statistical variation (n = 120) or (3) data reduction (n = 102) or both, variation and reduction (4) (n = 111). All four types of the student-centered treatments provided hands-on activities as well as standardized and individual feedback.

To gather data about motivational dispositions and statistics-related competency, students were asked to complete a paper and pencil test and questionnaire (45 minutes) before and after working on the learning material. During the testing and intervention period, the regular teacher as well as the first author monitored the implementation in the classes.

### **7.5.2 Measures**

Motivational dispositions referring to mathematics and statistics were measured in a multiple choice format using a 5-point Likert scale (1 = not true at all .... 5 = completely true). Mathematics-related self-concept (6 items) as well as a scale including intrinsic motivation and interest (6 items; in the following, we will refer to this scale under the label of “interest”) were based on scales established by Pekrun and colleagues (2002; 2003). In order to tap students’ statistics-related self-concept (3 items) as well as their interest and intrinsic motivation (3 items; in the following referred to under the label of “interest”), we used the instruments conceived by Gundlach, Kuntze, Engel and Martignon (2010) as parallel scales referring to statistics. These statistics-related scales were shown by a factor analysis to be empirically separable from the mathematics-related ones (see Table 7-2 in Section 6.1 for sample items). However, prior to further analysis steps, we checked by confirmatory factor analysis whether statistics- and mathematics-related self-concept and interest were also empirically separable variables within our sample.

Additionally, as an alternative measure for students’ self-concept (relatively close to the way self-efficacy can be measured) and interest in the domain of statistics, we asked participants how confident they felt about solving specific tasks and their interest connected to these specific tasks. In this format, three tasks referring to dealing with tables, bar charts and characteristic values (see Figure 7-1 for a sample item) were presented to the learners (cf. Zirngibl, Götz, Pekrun, vom Hofe, & Kleine, 2005). Students’ ratings of these tasks were summarized to a scale for self-concept, i.e. the confidence to solve these specific tasks, and a scale for interest in these specific tasks. In the following, we will refer to these scales under the label of “task-specific self-concept” and “task-specific interest”. Comparisons between the two task-specific and non-task-specific formats could give insight about whether students associated the term statistics with dealing with statistical contents such as the ones presented in the tasks.

Figure 7-1

Sample item for task-specific self-concept and interest

<b>What do you think about this task?</b>						
The table displays the grades in mathematics for class 8d from the last academic year. Compute the mean of these grades!						
grade	1	2	3	4	5	6
frequency	1	7	12	5	2	0

	Completely true.	Largely true.	Slightly true.	Hardly true.	Not true at all.
I am certain that I can solve this problem correctly.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I am interested in dealing with this task.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Additionally, to gather data about students' statistics-related competency, participants completed a competency test related to statistical literacy (Kuntze, Engel, Martignon & Gundlach, 2010; Kuntze et al., 2008) comprising of 15 tasks at five competency levels. Within our sample, the corresponding test instrument has shown to fit to a one-dimensional Rasch-model. EAP/PV reliability that is comparable to Cronbach's alpha (Rost, 2004) amounted to the rather moderate value of 0.47.

### 7.5.3 Analyses

Confirmatory factor analyses using maximum likelihood solutions were conducted with the software AMOS 19.0 (Arbuckle, 1995-2010). For each factor, the item with the highest factor loading was fixed to 1.0. As measures of fit, the comparative fit index (CFI; recommended  $\geq 0.95$  by Hu & Bentler, 1999) and the root-mean-square error of approximation (RMSEA; recommended  $\leq 0.06$  by Hu & Bentler, 1999) were assessed. Additionally, the  $\chi^2/df$ -ratio was calculated indicating that the model fits the data reasonably well when not exceeding 5 (Bollen & Long, 1993).

For the 503 cases, missing values for the items of the scales were below 4.7%. For significance tests, we used the Full Information Maximum Likelihood (FIML) procedure with the software Mplus 7.1 (Muthén & Muthén, 1998 – 2012) that allowed us to use the whole dataset despite single missing values on several variables. Additionally, the robust maximum likelihood estimator in Mplus corrects for non-normality in the measures. Hence, we could analyze our data despite slight violations of the assumption of normality for some of the variables.

As the participants in our sample were clustered in school classes, cases could not be assumed to be independent. Disregarding this dependency structure could lead to inaccurate parameter estimates because of deflated standard errors (Raudenbush & Bryk, 2002). Intraclass-correlations within our sample were rather low (0.0% - 0.5%), therefore, indicating no need for multilevel analysis (Snijders & Bosker, 1999). However, in order to respect the clustering of the data we conducted a design-based correction of standard errors via the “type is complex” procedure in Mplus.

To investigate if the means of domain-specific self-concept and interest significantly increased due to the intervention (research question 2), we specified multiple regression models predicting self-concept resp. interest in the posttest by the corresponding pretest values as well as by the group assignment (see Equation (7-1) for the structure of a regression model predicting self-concept (sc)).

#### Equation 7-1

Example of an equation to investigate research question 2

$$posttest(sc) = \beta_0 + \beta_1 \cdot pretest(sc) + \beta_2 \cdot group^6 + \varepsilon$$

This procedure allowed us to investigate differences in the development of self-concept resp. interest between the different groups. Here, the baseline group was used as a safeguard against effects that were not due to the intervention. We included the pretest score in the above regression model to control for the differences in this variable.

In order to evaluate if students’ pretest competency referring to statistical literacy influenced the development of their statistics-related self-concept resp. interest in the course of the intervention (research question 3), we specified regression models for self-concept resp. interest for the students of the intervention group. Equation (7-2) displays the structure of a regression model predicting statistics-related self-concept (sc) in the posttest by the pretest values of self-concept and pretest competency in the domain of statistical literacy (sl).

#### Equation 7-2

Example of an equation to investigate research question 3

$$posttest(sc) = \beta_0 + \beta_1 \cdot pretest(sc) + \beta_2 \cdot pretest(sl) + \varepsilon$$

---

<sup>6</sup> The groups were dummy coded, hence each treatment group was tested against the other ones.

Taking the pretest value of self-concept (sc) resp. interest as covariate into account allows to determine how strong competency in the domain of statistical literacy (sl) affects the development of self-concept.

## 7.6 Results

### 7.6.1 Empirical distinction of mathematics- and statistics-related self-concept and interest

As recommended by Gal et al. (1997), we first investigated whether the mathematics- and statistics-related scales used in this study represent empirically separable factors within our sample. Therefore, we included all 18 items of the questionnaire as indicators for the four scales *mathematics-related self-concept* and *mathematics-related interest* as well as *statistics-related self-concept* and *statistics-related interest* into distinct confirmatory factor analyses for the pretest and for the posttest. We compared global model fit (see Table 7-1) of this full model (Model D) with the fit of two reduced models: Model B includes one factor for mathematics-related self-concept and interest and another factor for statistics-related self-concept and interest; Model C includes one factor for mathematics- and statistics-related self-concept and another factor for mathematics- and statistics-related interest. We also compared the full model with a single factor model: Model A integrates the items of all four scales. Global fit indices displayed in Table 7-1 indicate that the full model (Model D) fits best the empirical data. Model comparisons by means of chi-square tests performed on the AICs confirms that the full model is significantly better than the other ones. Model fit indices for the posttest were very similar or even better than those reported for the pretest.

Table 7-1

Global fit indices for the tested models (pretest)

Model	$\chi^2/df$	CFI	RMSEA	Df	AIC
Model A (1 factor): sc & int for math & stat	14.20	0.69	0.17	135	2025.5
Model B (2 factors): sc & int math vs. sc & int stat	8.34	0.83	0.13	134	1227.6
Model C (2 factors): sc math & stat vs. int math & stat	9.71	0.80	0.14	134	1420.8
Model D (4 factors): sc math vs. int math vs. sc stat vs. int stat	2.33	0.97	0.05	129	421.0

**Note:** sc ... self-concept; int ... interest; math ... mathematics; stat ... statistics.  
See Section 5.3. for recommendations of the displayed model fit indices.

Concerning the full model, local fit indices showed that the latent variables were reliably measured by the corresponding items and that they were each distinguishable from the other variables (e.g., indicator reliabilities  $\geq 0.3$ ; statistically significant factor loadings; factor reliabilities  $\geq 0.85$ , see Table 7-2; average variance extracted  $\geq 0.66$ ; Fornell-Larcker-Criterion satisfactory).

Table 7-2

Reliabilities and sample-items of mathematics- and statistics-related self-concept and interest

Construct	No. of Items	Sample item	Reliability Pre / Post	Retest reliability (baseline)
Self-concept (Mathematics)	6	I am good at solving mathematical problems.	0.92 / 0.94	0.93
Interest (Mathematics)	6	I am interested in mathematics.	0.92 / 0.93	0.92
Self-concept (Statistics)	3	I am good at solving statistical problems.	0.85 / 0.86	0.79
Interest (Statistics)	3	I am interested in statistics.	0.85 / 0.87	0.78

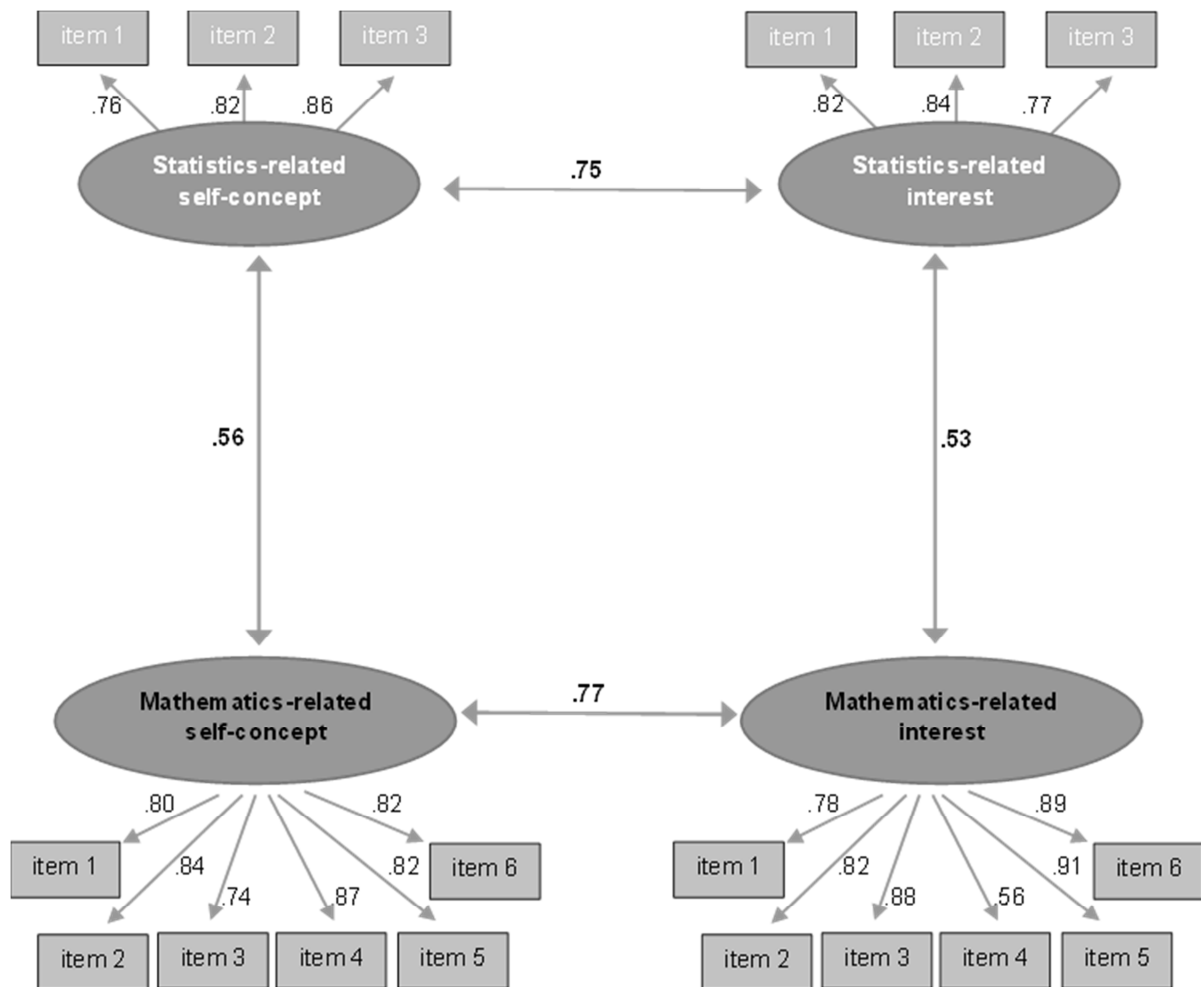
**Note:** Scales adapted from Pekrun et al. (2002; 2003) as used by Gundlach et al. (2010). Factor reliabilities estimated by AMOS 19.0.

Precise parameter estimates as well as the structure of the confirmatory factor analysis for the full model (pretest) can be found in Figure 7-2. Similar factor loadings and correlations were found for the posttest. Hence, the four scales mathematics- and statistics-related self-concept and interest turned out to be separable factors despite latent intercorrelations of .53 to .77. Retest-reliability of the baseline group was satisfactory (see also Table 7-2) for the four scales further indicating that the same constructs were measured in pre- and posttests.



Figure 7-2

Confirmatory factor analysis for the pretest



### 7.6.2 Fostering domain-specific self-concept and interest through a statistics-oriented intervention

In a first step, we investigated potential differences in the development of mathematics-related, statistics-related and task-specific self-concept and interest between the four treatment groups. For this purpose, we specified several regression models such as displayed in Equation (7-1) to test each of the four treatments against the other three. Except for one case, the regression weights for the group assignment were all non-significant. This means that there were (mostly) no significant differences at the  $\alpha = 0.05\%$  level concerning mathematics-related, statistics-related and task-specific self-concept resp. interest among the four treatment groups.

In a second step, we examined if there were differences between the treatment groups and the baseline group with respect to domain-specific self-concept and interest. Hence, we combined

the four treatment groups to one intervention group. We again used regression models such as illustrated in Equation (7-1) to investigate if the development of this intervention group was significant compared to the baseline group. Table 7-3 displays means and standard deviations to indicate how students' self-concept and interest related to mathematics and statistics developed from pre- to posttest. Within the intervention group, the means of the scales referring to mathematics barely differed between the two time points of testing. Although the difference concerning mathematics-related interest was significant for the intervention group compared to the baseline, the effect sizes for both mathematics-related scales were negligible. Concerning the domain of statistics, both students' self-concept and interest significantly increased from pre- to posttest in the intervention group. The effect sizes as measured by Cohen's  $d$ <sup>7</sup> reached a small to medium magnitude (Cohen, 1988) indicating that the increase in the intervention group was relevant.

Statistics-related interest and self-concept were additionally measured by students' ratings of three tasks referring to dealing with tables, bar charts and characteristic values. Reliabilities for these task-specific scales ranged from 0.67 to 0.94 in pre- and posttests. The absolute means of these scales were higher than for the more general statistics-related self-concept and interest scales (see Table 7-3). Similar to the statistics-related scales not referring to tasks, the task-based self-concept and interest scales both increased significantly from pretesting to posttesting. The growth of these variables ranged between a small and a medium effect size.

Pearson correlations between the statistics-related task-based and non-task-based self-concept scales amounted to  $r = .47$  ( $p < .001$ ) for the pretest and  $r = .48$  ( $p < .001$ ) for the posttest. Correlations between the statistics-related task-based and non-task-based interest scales were even higher (pretest:  $r = .57$ ,  $p < .001$ ; posttest:  $.66$ ,  $p < .001$ ). These correlations suggest that students associated the domain of statistics as referred to the non-task-based scales with tasks such as the presented ones, which referred to tables, bar charts and characteristic values.

---

<sup>7</sup> Cohen's  $d$  was calculated by  $d = \frac{M_1 - M_2}{\sqrt{(SD_1^2 + SD_2^2)/2}}$

Table 7-3

Students' self-concept and interest related to mathematics and statistics in pre- and posttesting

Construct	Pretest M <sub>1</sub> (SD <sub>1</sub> )	Posttest M <sub>2</sub> (SD <sub>2</sub> )	p-value	Cohen's d
<b>Self-concept (Mathematics)</b>				
Intervention group	3.28 (0.93)	3.29 (0.94)	.070	.01
Baseline group	2.86 (0.96)	2.79 (1.00)		
<b>Interest (Mathematics)</b>				
Intervention group	2.62 (1.05)	2.67 (1.06)	.029	.05
Baseline group	2.24 (1.01)	2.19 (1.02)		
<b>Self-concept (Statistics)</b>				
Intervention group	3.22 (0.96)	3.45 (0.90)	.007	.25
Baseline group	3.13 (0.92)	3.13 (0.95)		
<b>Interest (Statistics)</b>				
Intervention group	2.40 (1.03)	2.63 (1.07)	.045	.22
Baseline group	2.42 (0.93)	2.44 (0.99)		
<b>Self-concept (Task-specific)</b>				
Intervention group	3.73 (0.89)	4.12 (0.83)	.006	.45
Baseline group	3.55 (0.85)	3.67 (0.96)		
<b>Interest (Task-specific)</b>				
Intervention group	3.14 (1.03)	3.28 (1.11)	< .001	.13
Baseline group	2.97 (0.90)	2.78 (1.07)		

**Note:** M ... mean; SD ... standard deviation

### ***7.6.3 The impact of prior statistics-related competency on the development of self-concept and interest in this domain***

As an association between achievement and self-concept or interest is reported from various domains, we also investigated whether the growth in statistics-related self-concept and interest through the intervention depends on students' prior statistics-related competency.

In a first step, we specified among the students who participated in the intervention a multiple regression model (see Equation (7-2)) including self-concept in the pretest in order to determine the role of pretest competency for the development of self-concept. Table 7-4 shows the precise parameter estimates of the regression analysis. Competency in the pretest turned out to be a significant predictor for the development of posttest statistics-related self-concept after controlling for variance associated with pretest self-concept. The positive regression weight

indicates that learners with higher initial competency scores benefited more from the intervention than their less competent peers, hence an aptitude-treatment interaction.

Table 7-4

Multiple regression model predicting statistics-related self-concept (posttest)

Predictor	Standardized regression weight	Standard error	p-value
Statistical Self-concept (pretest)	.54	.04	< .001
Statistical Competency score (pre-test)	.10	.02	< .001

The same procedure was used to evaluate if students' pretest competency score influenced the development of their statistics-related interest. The regression weight of the competency score indicates a relationship in the same direction as seen in Table 7-4 for self-concept. However, this regression weight came close to a significance level of  $\alpha = 0.05$  (see Table 7-5).

Table 7-5

Multiple regression model predicting statistics-related interest (posttest)

Predictor	Standardized regression weight	Standard error	p-value
Statistical Interest (pretest)	.62	.02	< .001
Statistical Competency score (pre-test)	.05	.03	.056

## 7.7 Discussion

In this final section, we summarize and discuss the results in the order of the research questions. We will end up by stating limitations of this study and offering recommendations for further research.

Following the recommendation of Gal and colleagues (1997), we first evaluated if mathematics- and statistics-related self-concept and interest were empirically distinct variables within our sample. Model fit indices indicated that the four scales were empirically separable and reliably measured the underlying variables. This finding confirms that students perceived mathematics and statistics differently from a motivational point of view. Therefore, it is reasonable to evaluate such motivational variables for statistics separately from those related to mathematics. Latent correlations between self-concept and interest within both domains were

in the expected range (e.g. Carmichael et al., 2010; Klieme et al., 2001) and underline the proximity among these motivational variables. Correlations of self-concept and interest of approximately .50 between the variables related to the two domains confirmed the relatedness between mathematics and statistics. Our findings extend prior research on the consulted scales by Gundlach et al. (2010) by providing information about an adequate model fit obtained through confirmatory factor analysis. Thus, our analysis confirms the appropriateness of the measures for future studies.

The second and main research interest was to investigate if students' domain-specific self-concept and interest could be fostered through a statistics-oriented intervention. The importance of motivational variables referring to mathematics and statistics for learners' current and future lives (cf. Gal et al., 1997; Nagy et al., 2006) as well as a reported decline of these motivational variables in the course of schooling (see e.g. Carmichael et al., 2010; Frenzel et al., 2010; Jacobs et al., 2002; Watt, 2004) emphasize the relevance of this focus. Concerning the domain of mathematics, self-concept and interest remained relatively stable from pre- to posttests. As the intervention did not refer to mathematics in general, this finding was to be expected. In the sense of a retest-reliability, it can be considered as an indicator for the good quality of the measurement of the variables. However, there was a significant increase in general statistics-related and task-specific self-concept and interest with relevant effect sizes in the course of the intervention. Hence, in accordance with the literature (e.g. Frenzel et al., 2010; Marsh & Craven, 1997), this finding implies that self-concept and interest for statistics can be fostered through interventions in the corresponding domain.

Concerning the four variants of this intervention, we did not observe any significant differences in the increase of students' statistics-related self-concept and interest. As the treatment groups were principally designed to have an impact on students' competency of using representations and models in statistical contexts, we did not have any specific prior expectations related to such potential differences. It can be assumed that the learner-centered intervention providing hands-on activities, cooperative learning and different forms of feedback was – independently from variations in the focus of some of the tasks – suitable to foster students' self-concept and interest (see e.g. Hattie et al., 1996; Hidi & Renninger, 2006; Schukajlow et al., 2012). Moreover, as there was exclusively an increase in the statistics-related scales, the combination of the properties of the learning environment and the focus on the domain of statistics may have led to this increase. For educational practice, this means that students'

self-concept and interest related to statistics can be fostered by working on statistical problems in an appropriate classroom setting.

Furthermore, we investigated how the task-specific self-concept and interest scales developed over the course of the intervention. Although these tasks represent a very limited part of the nature of statistics, they are more sizeable and specific for students than broader and comprehensive items. The parallel increase in the general statistics-related scales and their task-specific equivalents as well as the magnitude of the correlations between them even before the intervention suggest that students saw a connection between the notion of statistics and the requirements of dealing with tables, bar charts and characteristic values. As the means of task-specific scales were higher than the general statistics-related ones, students may imagine that statistics can also involve more complicated and less interesting problems than those presented in the corresponding questionnaire. Altogether, it can be expected that students had an adequate concept of the term statistics when responding to the statistics-related items.

As in general, self-concept and interest are related to competencies in the corresponding domain; the third research question addressed if previous competency in the domain of statistics influenced the development of statistics-related self-concept and interest. It turned out that students' self-concept developed more positively if they had already shown relatively high competency referring to statistics in the pretest (see also Carmichael et al., 2010). Concerning statistics-related interest, there was also evidence for an aptitude treatment interaction, i.e. against the null hypothesis of no relationship, but this evidence was not strong ( $0.10 > p > 0.05$ ). For instructional practice related to our learning material, this finding implies in accordance with Hood et al. (2012) that students' with lower competency in the domain of statistics do not only have to be particularly supported concerning their cognitive but also concerning their motivational development.

#### *Limitations and further research*

Although the statistical prerequisites necessary to conducting the mentioned quantitative analyses are met, our findings should be interpreted carefully, given that the sample refers to a specific type of school and cannot be considered as representative for all German students.

This study was limited to investigate the effects of a short-term intervention. As motivational variables usually develop over a longer time period, it would be interesting to study long-term effects of such interventions. Such follow-up research should use a repeated measurement of motivational variables and competency and take into account potential motivational effects of

the regular classroom after the intervention, which could shape the further development of academic self-concept and interest.

Furthermore, the present study was not designed to provide reasons for the aptitude-treatment-interaction effect between self-concept and competency in the domain of statistics. Further - possibly qualitative - research should examine this interaction and explore why especially more competent students profited from the learning environment.

We want to thank Bob delMas for very helpful comments on a previous version of this paper.

Ute Sproesser is a member of the “Cooperative Research Training Group” of the University of Tübingen and the University of Education, Ludwigsburg, which is supported by the Ministry of Science, Research and the Arts in Baden-Wuerttemberg, Germany.

The research project “ReVa-Stat” is supported by research funds from Ludwigsburg, University of Education.

## 7.8 References

- Arbuckle, J. L. (1995-2010). *IBM SPSS AMOS 19.0 user's guide*. Chicago, IL: Small-Waters Corporation.
- Artelt, C., Baumert, J., Julius-McElvany, N. & Peschar, J. (2003). *Learners for life: Student approaches to learning. Results from PISA 2000*. Paris: OECD.
- Baumert, J., Bos, W. & Lehmann, R. (2000). *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn*. Opladen: Leske + Budrich.
- Baumert, J. & Köller, O. (2000). Unterrichtsgestaltung, verständnisvolles Lernen und multiple Zielerreichung im Mathematik- und Physikunterricht der gymnasialen Oberstufe. In J. Baumert, W. Bos & R. Lehmann (Eds.), *TIMSS III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn* (pp. 271-316) Opladen: Leske und Budrich.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review* 84(2), 191–215.
- Bollen, K., & Long, J. (1993). *Testing structural equation models*. Newbury Park, CA: Sage.
- Bond, M., Perkins, S.N. & Ramirez, C. (2012). Students' perceptions of statistics: An exploration of attitudes, conceptualizations, and content knowledge of statistics. *Statistics Education Research Journal*, 11(2), 6-25. <http://www.stat.auckland.ac.nz/serj> [17.07.2014].
- Brophy, J.E. & Good, T.L. (1986). Teacher behavior and student achievement. In M.C. Wittrock (Ed.), *Handbook of Research on Teaching* ( pp. 328-375). New York: Macmillan.
- Carmichael, C., Callingham, R., Hay, I & Watson, J. (2010). Statistical literacy in the middle school: The relationship between interest, self-efficacy and prior mathematics achievement. *Australian Journal of Educational & Developmental Psychology*. Vol 10, 2010, 83-93.
- Carmichael, C., Callingham, R., Watson, J., & Hay, I. (2009). Factors influencing the development of middle school students' interest in statistical literacy. *Statistics Education Research Journal*, 8(1), 62-81.



- Cobb, G.W. & Moore, D.S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, 104(9), 801-823. <http://www.macalester.edu/~kaplan/statchat/2009-2010-year/2010-01/January-2010.pdf> [15.08.2014].
- Cohen, J. (1988): *Statistical power analysis for the behavioral sciences*, 2. Edition, Hillsdale: Lawrence Erlbaum Associates.
- Eccles, J.S. & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology* 53, 109-132.
- Emmioglu, E. & Capa-Aydin, Y. (2012). Attitudes and achievement in statistics: A meta-analysis study. *Statistics Education Research Journal*, 11(2), 95-102. <http://www.stat.auckland.ac.nz/serj> [17.07.2014].
- Frenzel, A., Goetz, T., Pekrun, R., & Watt, H.M.G. (2010). Development of mathematics interest in adolescence: Influence of gender, family, and school context. *Journal of Research on Adolescence*, 20(2), 507-537.
- Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review* 70, 1-51.
- Gal, I., Ginsburg, L., & Schau, C. (1997). Monitoring attitudes and beliefs in statistics education. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 37–51). Amsterdam: IOS Press. <http://www.stat.auckland.ac.nz/~iase/publications/assessbk/chapter04.pdf> [06.08.2014].
- Gläser-Zikuda, M., Fuß, S., Laukenmann, M., Metz, K., & Randler, C. (2005). Promoting students' emotions and achievement—Instructional design and evaluation of the ECOLE approach. *Learning and Instruction*, 15, 481–495.
- Gundlach, M., Kuntze, S., Engel, J., & Martignon, L. (2010). Motivation and self-efficacy related to probability and statistics: Task-specific motivation and proficiency. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the 8th Int. Conf. on Teaching Statistics*. Voorburg, The Netherlands: ISI. [www.stat.auckland.ac.nz/~iase/publications.php](http://www.stat.auckland.ac.nz/~iase/publications.php) [04.06.2014].
- Hansford, B.C. & Hattie, J.A. (1982). The relationship between self and achievement / performance measures. *Review of Educational Research* 52, 123-142.

- Hattie, J., Biggs, J. B., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research*, 66, 99–136.
- Helmke, A. & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie, Band 3 (Psychologie der Schule und des Unterrichts)* (S. 71-176). Göttingen: Hogrefe-Verlag.
- Hidi, S. & Renninger, K.A. (2006). The four-phase model of interest development. *Educational Psychologist*, 41(2), 111-127.
- Hood, M., Creed, P. & Neumann, D.L. (2012). Using the expectancy value model of motivation to understand the relationship between student attitudes and achievement in statistics. *Statistics Education Research Journal*, 11(2), 72-85. <http://www.stat.auckland.ac.nz/serj> [17.07.2014].
- Holmes, P. (1980). *Teaching statistics 11-16*. Berkshire: Schools Council and Foulsham Education.
- Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis. Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. <http://dx.doi.org/10.1080/10705519909540118> [20.05.2014].
- Jacobs, E.J., Lanza, S., Osgood, D.W., Eccles, J.S. & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development* 73(2), 509-527.
- Klieme, E., Neubrand, M. & Lüdtke, O. (2001). Mathematische Grundbildung: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K. J. Tillmann & M. Weiß (Eds.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 141-191). Opladen: Leske + Budrich.
- KMK (Kultusministerkonferenz) (2003). *Bildungsstandards im Fach Mathematik für den mittleren Schulabschluss*. München: Wolters Kluwer.
- Kröpfl, B., Peschek, W., Schneider, E. (2000). Stochastik in der Schule: Globale Ideen, lokale Bedeutungen, zentrale Tätigkeiten. *mathematica didactica*, 23, 25–57.
- Kuntze, S., Engel, J., Martignon, L. & Gundlach, M. (2010). Aspects of statistical literacy between competency measures and indicators for conceptual knowledge – Empirical

- research in the project RIKO-STAT. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. [www.stat.auckland.ac.nz/~iase/publications.php](http://www.stat.auckland.ac.nz/~iase/publications.php) [14.6.2013].
- Kuntze, S., Lindmeier, A. & Reiss, K. (2008). "Using models and representations in statistical contexts" as a sub-competency of statistical literacy – Results from three empirical studies. *Proceedings of the 11th International Congress on Mathematical Education*. Monterrey (Mexico). <http://tsg.icme11.org/document/get/474>. [22.12.11].
- Marsh, H.W. & Craven, R. (1997). Academic self-concept: Beyond the dustbowl. In G. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement and adjustment* (pp. 131-198). Orlando, FL: Academic Press.
- Marsh, H.W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Integration of multidimensional self-concept and core personality constructs: Construct validation and relations to well-being and achievement. *Journal of Personality*, 74, 403-456.
- McLeod, D.B. (1992). Research on affect in mathematics education: A reconceptualization. In D.A. Grouws (Ed.), *Handbook of research on mathematics, teaching and learning* (pp. 575-596). New York: Macmillan.
- Möller, J. & Trautwein, U. (2009). Selbstkonzept. In E. Wild & J. Möller, *Pädagogische Psychologie* (pp. 180 – 203). Heidelberg: Springer.
- Muthén, L.K. and Muthén, B.O. (1998-2012). *Mplus user's guide. Seventh edition*. Los Angeles, CA: Muthén & Muthén.
- Nagy, G., Trautwein, U., Baumert, J., Köller, O., & Garrett, J. (2006). Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. *Educational Research and Evaluation*, 12, 323-345.
- O'Mara, A.J., Marsh, H.W., Craven, R.G., & Debus, R.L. (2006). Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis. *Educational Psychologist*, 41(3), 181-206.
- Pekrun, R., Götz, Jullien, S., Zirngibl, A., v. Hofe, R., & Blum, W. (2003). *Skalenhandbuch PALMA: 2. Messzeitpunkt (6. Klassenstufe)*. Universität München: Institut für Pädagogische Psychologie.

- Pekrun, R., Götz, J., Jullien, S., Zirngibl, A., v. Hofe, R., & Blum, W. (2002). *Skalenhandbuch PALMA: 1. Messzeitpunkt (5. Klassenstufe)*. Universität München: Institut Pädagogische Psychologie.
- Pekrun, R. & Zirngibl, A. (2004). Schülermerkmale im Fach Mathematik. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003* (S. 191-210). Münster: Waxmann.
- Ramirez, C., Schau, C. & Emmioglou, E. (2012). The importance of attitudes in statistics education. *Statistics Education Research Journal*, 11(2), 57-71. <http://www.stat.auckland.ac.nz/serj> [17.07.2014].
- Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical linear models* (2<sup>nd</sup> ed.). Thousand Oaks: Sage Publications.
- Rost, J. (2004). *Lehrbuch Testtheorie/ Testkonstruktion*. Bern: Hans Huber.
- Schau, C. & Emmioglou, E. (2012). Do introductory statistics courses in the United States improve students' attitudes? *Statistics Education Research Journal*, 11(2), 86-94. <http://www.stat.auckland.ac.nz/serj> [17.07.2014].
- Schiefele, U. (1992). Topic interest and levels of text comprehension. In K. A. Renninger, S. Hidi & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 151-182). Hillsdale, NJ: Erlbaum.
- Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist*, 26, 299-323.
- Schukajlow, S., Leiss, D., Pekrun, R., Blum, W., Müller, M., & Messner, R. (2012). Teaching methods for modelling problems and students' task-specific enjoyment, value, interest and self-efficacy expectations. *Educational Studies in Mathematics* 79(2), 215-237.
- Schunk, D.H. & Zimmerman, B.J. (2003). Self-regulation and learning. In W.M. Reynolds & G.E. Miller (Eds.), *Handbook of Psychology: Educational psychology* (Vol. 7, pp. 59-78). New York: Wiley.
- Shavelson, R.J., Hubner, J.J., & Stanton, G.C. (1976). Validation of construct interpretations. *Review of Educational Research*, 46, 407-441.
- Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

- Valentine, J.C., DuBois, D.L. & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, 39(2), 111-133.
- Watson, J. M. & Callingham, R. A (2003). Statistical literacy: a complex hierarchical construct. *Statistics Education Research Journal* 2, 3-46. [http://iase-web.org/documents/SERJ/SERJ2%282%29\\_Watson\\_Callingham.pdf](http://iase-web.org/documents/SERJ/SERJ2%282%29_Watson_Callingham.pdf) [12.10.2011].
- Watt, H.M.G. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th- through 11th-grade Australian students. *Child Development* 75(5), 1556-1574.
- Wild, C.J. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review* 67, 223-265.
- Zirngibl, A., Götz, T., Pekrun, R., vom Hofe, R., & Kleine, M. (2005). Aufgabenspezifische Erfassung von Mathematikemotionen. *Vortrag auf der Tagung der Sektion Empirische Bildungsforschung (AEPF-Tagung)*, Freie Universität Berlin, 17.03.2005.

## 8 Gesamtdiskussion

### 8.1 Zusammenfassung

In Anbetracht der Relevanz von Kompetenzen des Verstehens und Interpretierens von Daten für die informierte Teilhabe an der Gesellschaft und das Treffen von evidenzbasierten Entscheidungen im privaten und öffentlichen Leben zielte die vorliegende Dissertation auf die Untersuchung und Förderung verschiedener Lernervariablen im Bereich von Statistical Literacy ab: In einem ersten Schritt konnten Zusammenhänge zwischen allgemeinen Eingangsvoraussetzungen von Lernenden auf der Individual- und Klassenebene und der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ ermittelt werden. Des Weiteren beleuchtete die Dissertation, inwieweit sich diese Kompetenz, Sichtweisen auf zufallsbedingte Variabilität sowie das bereichsspezifische Selbstkonzept und Interesse im Verlauf einer statistikbezogenen Intervention entwickelten. Dabei wurden insbesondere individuelle Voraussetzungen sowie die Zugehörigkeit zu den verschiedenen Treatments als potentielle Einflussfaktoren analysiert. Tabelle 8-1 gibt einen Überblick über die wesentlichen Ergebnisse der drei Teilstudien dieser Dissertation. Die gewonnenen Erkenntnisse können einerseits für die fachdidaktische Theoriebildung und andererseits für die Konzeption von Materialien, die auf die Förderung von Kompetenz im Bereich von Statistical Literacy, von Sichtweisen auf Variabilität und von bereichsspezifischen motivationalen Variablen abzielen, genutzt werden. In den folgenden Abschnitten werden die Ergebnisse der drei Teilstudien in der Gesamtschau dieser Arbeit diskutiert.

Tabelle 8-1

Übersicht über Zielsetzungen und Ergebnisse der einzelnen Teilstudien

Zielsetzung der Teilstudien	Hauptergebnisse
Teilstudie 1: Analyse von Zusammenhängen zwischen der Kompetenz bezüglich Statistical Literacy und Voraussetzungen der Lernenden wie Leseverständnis und kognitiven Fähigkeiten. Abgrenzung des Kompetenzkonstrukts von allgemeinen Lernervariablen.	Während das „Nutzen von Darstellungen und Modellen“ kaum im Zusammenhang mit dem Leseverständnis der Lernenden zu stehen schien, konnten moderate Zusammenhänge mit kognitiven Fähigkeiten, der Mathematiknote und dem Geschlecht festgestellt werden. Auf der Klassenebene spielte der aggregierte sozioökonomische Status eine Rolle für die Ausprägung dieser Kompetenz. Diese Ergebnisse stellen die Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ als relativ eigenständig und bereichsspezifisch heraus. Der geringe Zusammenhang mit dem Leseverständnis legt nahe, dass die textgestützte Anlage der Lernmaterialien dieser Studie kein grundsätzliches Hindernis für Lernende mit niedrig ausgeprägtem Leseverständnis sein muss. Der Zusammenhang mit der Mathematiknote bestätigt frühere Arbeiten bezüglich der Bedeutung mathematischen Wissens für Statistical Literacy. Dass Jungen über

---

höhere durchschnittliche Kompetenzscores verfügen, ist konsistent mit Forschungsergebnissen bezüglich mathematischer Kompetenz. Die Erklärung von Varianz zwischen den Klassen durch den sozioökonomischen Status könnte an den unterschiedlichen Möglichkeiten der Unterrichtsgestaltung je nach Schülerschaft liegen.

---

Teilstudie 2:

Untersuchung des Einflusses von individuellen Voraussetzungen und der verschiedenen Treatments auf die Entwicklung von Kompetenz bezüglich Statistical Literacy im Verlauf einer Intervention.

Ermittlung des Einflusses von individuellen Voraussetzungen und der verschiedenen Treatments auf die Sichtweisen von Lernenden auf zufallsbedingte Variabilität während dieser Intervention.

Vergleichbar zu Teilstudie 1 stellte sich das Leseverständnis trotz textgestützter Anlage der Lernmaterialien als nicht bedeutsam für die Kompetenzentwicklung heraus. Kognitive Fähigkeiten und die Mathematiknote dagegen beeinflussten diese auch unter Kontrolle des Vortests signifikant. Jungen profitierten insgesamt mehr von der Intervention als Mädchen. Zwischen den Lernenden der verschiedenen Treatments wurden kaum Unterschiede in der Kompetenzentwicklung festgestellt.

Die Ergebnisse deuten darauf hin, dass ein schwach ausgeprägtes Leseverständnis den Kompetenzaufbau im Verlauf der Intervention nicht beeinträchtigt. Dagegen zeigte sich ein besonderer Förderbedarf für Mädchen, wobei auch die Gründe dieses Geschlechtereffekts weiter beleuchtet werden sollten. Diese Studie konnte keinen empirischen Beleg für die besonders positive Wirkung von Reflexionsanlässen bezüglich Reduktion und Variabilität auf den Kompetenzaufbau bezüglich Statistical Literacy finden.

Während die einbezogenen individuellen Lernervariablen keinen Einfluss auf Sichtweisen bezüglich Variabilität hatten, führte die Zugehörigkeit zu einem der reflexionsorientierten Treatments dazu, dass diese Lernenden im Vergleich zum Basis-Training zufallsbedingte Variabilität nach der Intervention stärker berücksichtigten als zuvor. Demnach wirkten Reflexionen über Reduktion bzw. Variabilität auf Sichtweisen bezüglich Variabilität förderlich.

---

Teilstudie 3:

Analyse der Wirkung der Intervention auf das bereichsspezifische Selbstkonzept und Interesse.

Die verwendeten Skalen zur Ermittlung des Selbstkonzepts und Interesses bezogen auf Mathematik und Statistik stellten sich als empirisch trennbare Faktoren heraus. Korrelationsanalytische Auswertungen legten es nahe, dass die Lernenden den Begriff Statistik mit dem Umgang mit Tabellen, Diagrammen und Kennwerten assoziierten. Die Intervention führte zu signifikant höheren Skalenmittelwerten des statistikspezifischen Selbstkonzepts und Interesses in allen vier Interventionsgruppen. Besonders Lernende mit hohen Kompetenzausprägungen im Vortest konnten durch die Intervention signifikant in ihrem Selbstkonzept und tendenziell auch in ihrem Interesse gefördert werden.

Die Auseinandersetzung mit statistischen Inhalten in einer schülerzentrierten Arbeitsform schien förderlich für das bereichsspezifische Selbstkonzept und Interesse zu sein, während die verschiedenen Treatments keine wesentlichen Unterschiede in der motivationalen Entwicklung nach sich zogen. Der Zusammenhang zwischen Leistungs- und motivationalen Variablen bekräftigt, dass beide Aspekte durch geeignete Materialien gefördert werden sollten.

---

## **8.2 Zusammenhänge zwischen der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ und Eingangsvoraussetzungen der Schüler**

In den Teilstudien 1 und 2 wurden Zusammenhänge zwischen verschiedenen Eingangsvoraussetzungen der Lernenden und der im Bereich von Statistical Literacy angesiedelten Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ bzw. deren Entwicklung im Verlauf einer Intervention untersucht. Die Verbindung der Ergebnisse beider Teilstudien ermöglicht Aussagen über die Ausprägung und Entwicklung der Kompetenz im Zusammenspiel mit den untersuchten Lernervariablen.

Sowohl Teilstudie 1 als auch Teilstudie 2 analysierten den Zusammenhang zwischen Leseverständnis und dem „Nutzen von Darstellungen und Modellen in statistischen Kontexten“, der aus verschiedenen Perspektiven naheliegender erschien. So betont Gal (2002) die Bedeutung von allgemeinen Fähigkeiten des Entnehmens und Verarbeitens von Information aus schriftlichen Darstellungen (sogenannte Literacy Skills) für Statistical Literacy. Dabei hebt er insbesondere hervor, dass neben speziellen statistischen Darstellungen oft auch komplexe Texte verstanden werden müssen, um eine Aufgabenstellung in statistischen Kontexten vollständig zu erfassen. Diese Überlegung erschien insbesondere für das vorliegende Forschungsprojekt relevant zu sein, in dem sowohl die Testitems als auch die Lernmaterialien in schriftlicher Form vorlagen und eine schriftliche Bearbeitung durch die Lernenden erforderten. Ein weiterer möglicher Verbindungspunkt zwischen textorientiertem Leseverständnis und der Kompetenz wurde darin gesehen, dass das „datenbezogene Lesen“ (vgl. Curcio, 1987) ein grundlegendes Element des Kompetenzkonstrukts darstellt. Auch wenn beim Umgang mit bzw. Lesen von statistischen Darstellungen zusätzlich fachspezifisches Wissen erforderlich ist (siehe z.B. Gal, 2002), ist es denkbar, dass sich Prozesse des Entnehmens von Information aus Texten und aus statistischen Darstellungen durchaus ähnlich sein können (Erkennen bedeutsamer Informationseinheiten, Einbetten einzelner Elemente in den Kontext, ...).

In der vorliegenden Dissertation zeigte sich das textorientierte Leseverständnis zwar als signifikant mit dem Vortestscore korreliert, stellte sich aber unter Kontrolle der kognitiven Fähigkeiten als unbedeutend für die Ausprägung und für die Entwicklung der Kompetenz im Verlauf der Intervention heraus. In Verbindung mit dem Befund, dass die Aufgaben des Kompetenztests zum ersten und zweiten Messzeitpunkt teilweise sehr erfolgreich von Lernern mit unterschiedlichen Ausprägungen von Leseverständnis gelöst werden konnten, deutet dies darauf hin, dass die Aufgabenformulierungen des Testinstruments und der Lernmaterialien für die Probanden angemessen waren. Relative Schwächen im Leseverständnis hinderten die Ler-



nenden demnach nicht daran, im Kompetenztest gut abzuschneiden. Auch die textgestützte Anlage der Intervention scheint Lernende mit eher schwachem Leseverständnis gleichermaßen in ihrer Kompetenz bezüglich Statistical Literacy gefördert zu haben wie Lernende mit starkem Leseverständnis. Das erforderliche Mindestmaß an Leseverständnis, um die Testaufgaben bearbeiten und die Lernmaterialien nutzen zu können, brachten die Schüler der achten Realschulklasse anscheinend als Eingangsvoraussetzung mit. Ein Grund für den eher geringen Zusammenhang könnte darin gesehen werden, dass das „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ in hohem Maße fachspezifische Aspekte des Lesens erfordert, also über textorientiertes Lesen hinaus auf fachspezifische Kenntnisse und Fähigkeiten beispielsweise über Diagramme und deren Eigenschaften rekurriert. Insofern geht dieser Befund mit Ergebnissen von Leiss et al. (2010) einher, die analog keinen Zusammenhang zwischen dem gleichen Maß für Leseverständnis und dem Lösen mathematischer Modellierungsaufgaben feststellten. Auch diese Autoren kamen zu der Schlussfolgerung, dass – obwohl das Verstehen und Bearbeiten von Modellierungsaufgaben sinnerfassendes Lesen voraussetzt – mathematikspezifischere Kompetenzen entscheidend für Modellierungsfähigkeiten sind.

Teilstudie 1 und 2 nahmen außerdem den Zusammenhang zwischen dem „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ und allgemeinen kognitiven Fähigkeiten in den Blick. Da kognitive Prozesse dem Lösen von (statistischen) Problemen zu Grunde liegen, war eine gewisse Verwobenheit der beiden Variablen zu erwarten. Signifikante Zusammenhänge zwischen kognitiven Fähigkeiten und Mathematical Literacy wurden in Schulleistungstests wie PISA (z.B. Klieme, Neubrand et al., 2001) bereits empirisch belegt. Klieme, Neubrand et al. (2001) stellten insbesondere fest, dass kognitive Fähigkeiten nicht nur direkt mit Mathematical Literacy korrelieren, sondern zusätzlich über die Lesekompetenz vermittelt werden. Dies weist auf einen Überschneidungsbereich zwischen kognitiven Fähigkeiten und Lesekompetenz hin. In Teilstudie 1 dieser Arbeit konnte ein signifikanter Zusammenhang zwischen der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ und kognitiven Fähigkeiten ermittelt werden. Durch die Modellaufnahme von kognitiven Fähigkeiten verschwand der zunächst als signifikant ermittelte Zusammenhang zwischen dem Leseverständnis und dieser Kompetenz. Es zeigten sich also auch in dieser Studie Überschneidungen zwischen dem Leseverständnis und kognitiven Fähigkeiten, wobei letztere anscheinend ein grundlegendes Leseverständnis, wie es hier notwendig war, miteinschlossen. Zudem zeigten sich kognitive Fähigkeiten in Teilstudie 2 als über den Vortest hinaus bedeutsam für die Entwicklung der Kompetenz im Bereich von Statistical Literacy. Lernende mit hohen kognitiven Fähigkeiten verfügten demnach nicht nur vor der Intervention bereits über

eine höhere durchschnittliche Kompetenzausprägung, sondern konnten die Lernmaterialien auch effektiver für ihren Kompetenzaufbau nutzen.

Obleich sich kognitive Fähigkeiten erwartungsgemäß als signifikante Prädiktoren für die Ausprägung und Entwicklung von Kompetenz im Bereich von Statistical Literacy herausstellten, so ist die Größe des Zusammenhangs doch relativ moderat (rund 16% an Varianzerklärung in Teilstudie 1; rund 3% an zusätzlicher Varianzaufklärung unter Kontrolle des Vortests in Teilstudie 2). Es kann demnach davon ausgegangen werden, dass das verwendete Testinstrument nicht nur allgemeine kognitive Fähigkeiten, sondern eine relativ eigenständige und bereichsspezifische Kompetenz misst. Die geringen bis moderaten Zusammenhänge zwischen dem „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ und den allgemeinen Lernvoraussetzungen Leseverständnis und kognitive Fähigkeiten bestätigen somit das Testinstrument sowie das zu Grunde liegende Kompetenzmodell als spezifisch für den entsprechenden Inhaltsbereich.

Über allgemeine kognitive Fähigkeiten hinaus wurden in Teilstudie 1 und 2 mit der Mathematiknote auch Bezüge zwischen fachspezifischen Lernvoraussetzungen und der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ analysiert. Auf die Bedeutung mathematischer Kenntnisse und Fähigkeiten für Statistical Literacy weist beispielsweise Gal (2002) in einer theoretischen Arbeit hin. Kuntze und Kollegen (2010) berichten von signifikanten Zusammenhängen zwischen der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ und Wissen in den Bereichen „Wahrscheinlichkeit“, „Funktionaler Zusammenhang“ und „Risiko“. Diese Zusammenhänge werden mit inhaltlichen Überschneidungen wie beispielsweise dem Umgang mit Darstellungen erklärt. Ein vergleichbarer Überschneidungsbereich ist auch zwischen dieser Kompetenz im Bereich von Statistical Literacy und allgemeinen mathematischen Fähigkeiten, wie sie durch die Mathematiknote erfasst werden, zu vermuten. Teilstudie 1 und 2 konnten einen signifikanten Zusammenhang zwischen der Mathematiknote und dem „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ jeweils über allgemeine kognitive Fähigkeiten hinaus finden. Demzufolge wiesen Schüler mit relativ guter Mathematiknote im Durchschnitt nicht nur vor der Intervention bereits höhere Kompetenzscores vor, sondern verzeichneten auch einen stärkeren Kompetenzzuwachs im Verlauf der Intervention. Dennoch war auch hier die Größe des Zusammenhangs eher moderat. In Teilstudie 1 erklärten allgemeine kognitive Fähigkeiten und die bereichsspezifischere Mathematiknote knapp 21% der Varianz des Kompetenzscores. Auch in Teilstudie 2 gab es nach Einbezug dieser beiden Prädiktoren noch einen großen Teil an uner-

klärter Varianz. Die Kompetenz wird also zusätzlich durch weitere, vermutlich noch bereichsspezifischere Variablen beeinflusst. In einer Studie von Sproesser et al. (2014) konnte gezeigt werden, dass Wissen in den Bereichen „Wahrscheinlichkeit“ und „Funktionaler Zusammenhang“ mehr Varianz dieser Kompetenz aufzuklären vermag als die allgemeineren Mathematiknoten. Eine Erklärung dafür sehen die Autoren darin, dass Wissen über „Funktionale Zusammenhänge“ Kenntnisse und Fähigkeiten über grafische Darstellungen miteinschließt beziehungsweise Wissen über „Wahrscheinlichkeit“ das Modellieren von Zufall und den Umgang mit Variabilität erleichtert. Diese Überlegungen sollten durch Anschlussforschung weiter untersucht werden.

Teilstudie 1 und 2 nahmen außerdem potentielle Geschlechterunterschiede in den Blick. Schulleistungstests wie PISA oder TIMSS hatten gezeigt, dass Jungen im Durchschnitt über höhere mathematikbezogene Kenntnisse und Fähigkeiten verfügen als Mädchen (z.B. Brunner, 2005; Hosenfeld et al., 1999; Klieme, Neubrand et al., 2001). Klieme (1997) zeigte in einer Meta-Analyse, dass diese mathematikbezogenen Geschlechterunterschiede besonders stark beim mathematischen Problemlösen ausgeprägt sind und sich mit zunehmendem Alter noch verstärken. Auf den Bereich Statistik übertragen berichtet Martignon (2010) von einer größeren Ablehnung gegenüber statistischer Information unter Mädchen verglichen mit ihren männlichen Mitschülern. In Anbetracht der notwendigen Auseinandersetzung mit statistischen Problemen im Kompetenztest sowie in den Lernmaterialien erschienen ähnliche Geschlechterunterschiede in der Ausprägung und Entwicklung der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ ebenso denkbar. Teilstudie 1 und 2 bestätigten diese Vermutung: Mädchen verfügten bereits im Vortest durchschnittlich über niedrigere Kompetenzscores als Jungen. Zudem profitierten Jungen stärker von der kurzen Intervention. Da sich die bestehenden Geschlechterunterschiede innerhalb von nur vier Unterrichtsstunden noch verstärkten, besteht Handlungsbedarf um die Diskrepanz zwischen den Geschlechtern nicht noch größer werden zu lassen. Ein Ansatzpunkt für die spezielle Förderung von Mädchen im Bereich von Statistical Literacy könnte die Studie von Carmichael und Hay (2009) sein, die unterschiedliche auf Statistik bezogene Interessenskontexte von Jungen und Mädchen identifizierte. Martignon (2010) stellt in diesem Zusammenhang fest, dass Mädchen besonders hohe Motivation und Leistung vorweisen, wenn mathematische Inhalte eine persönliche Bedeutung für sie haben. Ob durch solche spezifisch aufbereiteten Materialien Mädchen tatsächlich verstärkt gefördert werden können, sollte durch entsprechende Anschlussforschung näher beleuchtet werden. Zudem sollten mögliche Gründe für den gefundenen Geschlechtereffekt untersucht werden. Aus bestehender Forschung ist beispielsweise bekannt,

dass Mädchen teilweise auch deshalb schlechter in Mathematiktests abschneiden, weil sie häufig über ein schwächeres diesbezügliches Selbstkonzept verfügen (Hannover, 2010; Pekrun & Zirngibl, 2004). Ob das bereichsspezifische Selbstkonzept auch in der vorliegenden Studie für den Geschlechtereffekt verantwortlich ist, sollte in Anschlussforschung in den Blick genommen werden.

In Teilstudie 1 wurde außerdem der Zusammenhang zwischen der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ und dem sozioökonomischen Status untersucht. Auch hier legten verschiedene Studien bezüglich mathematischer Kompetenz (z.B. Ehmke, Hohensee, Siegle & Prenzel, 2006; Klieme, Neubrand et al., 2001; Opdenakker & Van Damme, 2001; Sirin, 2005; Thrupp et al., 2002) einen Zusammenhang zwischen dem sozioökonomischen Status und der Kompetenz bezüglich Statistical Literacy auf Individual- und Klassenebene nahe. Während in Teilstudie 1 der sozioökonomische Status auf der Individualebene nicht signifikant mit dem Kompetenzkonstrukt zusammenhing, erklärte der Klassendurchschnitt die eher geringe Klassenvarianz von 3,4% im Vortest fast vollständig auf. Der nicht signifikante Zusammenhang auf der Individualebene könnte damit erklärt werden, dass die im Testinstrument enthaltenen Aufgaben nur wenig mit typischen Schulbuch-Aufgaben gemein haben und deshalb Lernende mit relativ hohem sozioökonomischem Status diesbezüglich nicht stärker unterstützt werden als Lernende mit eher niedrigem sozioökonomischen Status. Eine mögliche Begründung für den signifikanten Zusammenhang auf der Klassenebene könnte darin bestehen, dass Klassen, deren Schülerschaft im Wesentlichen aus bildungsnahen Familien kommt, weniger Disziplinprobleme und ein lernförderlicheres Klima vorweisen (vgl. Willms, 1992). Ein aus diesen Gründen fruchtbarer Unterricht würde dann allen Lernenden einer Klasse zu Gute kommen und nicht nur Schüler aus bildungsnahen Familien fördern. Allerdings muss beachtet werden, dass im Rahmen dieser Dissertation nicht umfassend geklärt werden konnte, ob tatsächlich der sozioökonomische Status direkt die Kompetenzunterschiede zwischen den Klassen erklärt oder ob es sich dabei um einen Drittvariableneffekt handelt. Zumindest konnte hier aber festgestellt werden, dass im Vortest Klassenunterschiede bezüglich der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ bestehen, die durch andere Klassenvariablen wie gute durchschnittliche Mathematiknoten oder hohe Mittelwerte in den kognitiven Fähigkeiten nicht erklärt werden konnten. Eine Studie von Sproesser et al. (2014) wies Klassenunterschiede bezüglich dieser Kompetenz auch in der neunten Klassenstufe in stärkerem Maße nach, konnte die Ursache aber ebenso nicht klären. Nach erfolgter Intervention bestanden fast keine Klassenunterschiede mehr, so dass keine tiefer gehende Untersuchung stattfinden konnte. Da alle Klassen wäh-

rend der Intervention die gleiche Art von Unterricht erfahren, liegt der Schluss nahe, dass die im Vortest ermittelten Klassenunterschiede tatsächlich mit der unterschiedlichen Gestaltung des regulären Unterrichts in den Klassen zusammenhängen.

Insgesamt konnten im Rahmen dieser Arbeit kognitive Fähigkeiten, die Mathematiknote, das Geschlecht sowie der durchschnittliche sozioökonomische Status als bedeutsam für die Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ identifiziert werden. Entgegen der aus der Theorie ableitbaren Erwartung zeigten sich kaum signifikante Zusammenhänge zwischen einem textorientierten Leseverständnis und dieser Kompetenz. Durch die Abgrenzung von den genannten Lernervariablen wurden das Testinstrument und somit auch das dahinter stehende Kompetenzkonstrukt empirisch für den Inhaltsbereich Statistik bestätigt. Dies spricht für den sinnvollen Einsatz dieses Testinstruments in zukünftiger Forschung. Darüber hinaus sind die Befunde für die fachdidaktische Theoriebildung bezüglich möglicher Einflussfaktoren auf Kompetenz im Bereich von Statistical Literacy von Bedeutung.

### **8.3 Zur Wirkung der Treatments auf das Kompetenzkonstrukt, Sichtweisen bezüglich zufallsbedingter Variabilität und motivationale Variablen**

Neben der Verbreiterung der fachdidaktischen Basis durch die Untersuchung von Zusammenhängen zwischen Lernervoraussetzungen und der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ zielte die vorliegende Dissertation darauf ab, diese Kompetenz, Sichtweisen auf Variabilität sowie bereichsspezifische motivationale Variablen zu fördern und daraus Hinweise auf Gestaltungsmerkmale (lern)förderlicher Materialien abzuleiten. Wie in Teilstudie 1 und 2 dargelegt, führte die statistikbezogene Intervention je nach Zielvariable zu unterschiedlichen Ergebnissen, die im Folgenden noch einmal in der Gesamtschau diskutiert werden.

Aufgrund der Bedeutung des Umgangs mit Variabilität sowie von Kenntnissen und Fähigkeiten des Reduzierens von Daten für Statistical Literacy wurden Reflexionsanlässe bezüglich dieser beiden Aspekte als spezielle Förderansätze in den Lernmaterialien umgesetzt. Wie oben geschildert enthielten die Materialien aller vier Treatments im 2x2-Design zu einem gewissen Umfang die gleichen Grundaufgaben. Während das Basis-Training eine relativ große Anzahl solcher Grundaufgaben beinhaltet, wiesen die Treatments „Reduktion“, „Variabilität“ sowie „Reduktion und Variabilität“ ausgehend von einigen dieser Grundaufgaben zusätzlich auf Reduktion und / oder Variabilität abzielende Prompts auf. Die hier angeregten Reflexionen zielten darauf ab, diesbezügliche Lernervorstellungen aufzubauen und dadurch

die Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ sowie die Berücksichtigung von zufallsbedingter Variabilität in den Sichtweisen der Lernenden zu fördern.

Teilstudie 2 ermittelte zwischen den vier Treatmentgruppen kaum Unterschiede in der Kompetenzentwicklung durch die Intervention. Die individuellen Voraussetzungen der Lernenden stellten sich als bedeutsamer für diese heraus als die Treatmentzugehörigkeit. Insbesondere konnte durch die Anregung von Reflexionen über Reduktion oder / und Variabilität keine stärkere Kompetenzentwicklung erzielt werden als durch das Basis-Training, das eher auf technische und algorithmische Aktivitäten fokussierte. Die Gruppe „Reduktion“ wies sogar eine geringfügig schwächere Kompetenzentwicklung als diese Referenzgruppe vor. Ein möglicher Grund dafür wurde in Teilstudie 2 bereits diskutiert: Den Lernenden der reflexionsorientierten Treatments standen vermutlich weniger Zeit und Aufgaben zur Verfügung, die auf das grundlegende Nutzen von statistischen Darstellungen und Modellen abzielten. Dies war aber notwendig, um die Aufgaben der Kompetenzniveaus I, II und III zu lösen, auf denen die Kompetenzausprägungen der meisten Schüler zum Zeitpunkt des Vor- und Nachtests zu verorten waren. Die höchsten Kompetenzzuwächse konnten im Verlauf der Intervention auf Niveau II und III verzeichnet werden. Insofern ist es plausibel, dass sich die Lernenden der reflexionsorientierten Treatments im Vergleich zu den Lernenden des Basis-Trainings auf diesen Kompetenzniveaus eher weniger stark verbesserten. Bezüglich der Aufgaben von Kompetenzniveaus IV und V kam es auch unter den Schülern der reflexionsorientierten Treatments kaum zu höheren Lösungsraten von Vor- zu Nachtest, obwohl diese Treatments auf wesentliche Elemente abzielten, die auf diesen Kompetenzniveaus erforderlich waren. Im Anhang kann sich der Leser davon überzeugen, dass die Anregung von Reflexionen bezüglich Reduktion oder / und Variabilität in den Lernmaterialien der Intervention sorgfältig umgesetzt wurde. Dies legt die Vermutung nahe, dass die Intervention zu kurz war oder dass die Aufgaben der höchsten beiden Kompetenzniveaus insgesamt zu anspruchsvoll für die Lernenden der vorliegenden Stichprobe waren. Die Kompetenz der Lernenden dieser achten Klassen ist demnach im Wesentlichen auf den Kompetenzniveaus I bis III zu verorten.

Eine deutlichere Wirkung hatten die Reflexionsanlässe über Reduktion oder / und Variabilität auf Sichtweisen bezüglich Variabilität. Die vierstündige Intervention führte dazu, dass die Schüler der reflexionsorientierten Treatments statistische Variabilität in zufälligen Verteilungen signifikant stärker berücksichtigen als die Lernenden des Basis-Trainings. Auch wenn zufallsbedingte Variabilität selbst nach der Intervention noch eher mäßig beachtet wurde,

vermittelt der erzielte kleine bis mittlere Effekt in Anbetracht der Kürze der Intervention den Eindruck, dass Sichtweisen auf Variabilität durchaus zu beeinflussen sind. Anscheinend haben die Reflexionsanlässe bezüglich Reduktion oder / und Variabilität die Lernenden tendenziell dazu angeregt, allzu regelmäßige Strukturen in Verteilungen kritischer zu hinterfragen und den Zufall stärker in ihre diesbezüglichen Überlegungen zu integrieren. Dies kann auch als eine positive Entwicklung hinsichtlich des Aufbaus von Statistical Literacy gesehen werden, insbesondere da Shaughnessy (2007) darauf hinweist, dass Vorstellungen über Zufall in der Regel nur schwer und langfristig zu beeinflussen sind.

Insgesamt konnte also durch die Anregung von Reflexionen bezüglich Reduktion oder / und Variabilität – zumindest in dieser Studie – keine besondere Förderung von Kompetenz im Bereich von Statistical Literacy nachgewiesen werden. Andererseits wirkten diese Reflexionen förderlich auf die Sichtweisen der Lernenden bezüglich Variabilität. Über Gründe für diese – je nach Zielvariable unterschiedliche – Befundlage kann auf der Grundlage der vorliegenden Daten nur spekuliert werden: Es ist denkbar, dass der Kompetenzaufbau eine langfristige oder stärker diskursive Auseinandersetzung mit den komplexen Inhalten Reduktion beziehungsweise Variabilität erfordern würde. Denn gerade im Austausch mit den Meinungen, Einschätzungen und Erfahrungen von anderen kann die eigene Perspektive besonders von schwer zugänglichen Konzepten erweitert und bereichert werden. Auch wäre es möglich, dass die Lernenden der untersuchten achten Klassenstufe noch nicht empfänglich für diesbezügliche Reflexionen sind und noch ein gewisser Reifeprozess notwendig ist. Diese Überlegung wird durch Forschung von Kuntze et al. (2014) untermauert, die unter Neuntklässlern gute Lösungsraten auf Kompetenzniveau IV, das den Umgang mit Variabilität erforderte, feststellen konnten. Obwohl die Lernmaterialien der Intervention entsprechend theoretischer Überlegungen in mehreren Entwicklungsrunden sorgfältig konzipiert und auch durch Pilotierungen auf ihre grundsätzliche Eignung für Achtklässler erprobt wurden, deutet sich an, dass sie nur wenig den Bedürfnissen der vorliegenden Stichprobe in Bezug auf die Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“, insbesondere in Bezug auf die Förderung des Umgangs mit Variabilität, gerecht wurden. Um diesbezüglich passgenauere Fördermaterialien zu entwickeln, sollte untersucht werden, an welchen Verständnishaürden die Lernenden scheitern und wie diese überwunden werden können. In Anbetracht der alltagsbezogenen Relevanz von Kompetenz im Bereich von Statistical Literacy ist die weitere Erforschung von möglichen Förderansätzen dringend notwendig.

Durch die statistikbezogene Intervention sollten auch bereichsspezifische motivationale Variablen gefördert werden. Die Beachtung und Unterstützung von motivationalen Variablen im Unterricht ist auch deshalb wünschenswert, da diese das Lernen und somit Leistungsvariablen determinieren. In Teilstudie 3 konnte gezeigt werden, dass sich sowohl das statistikspezifische Selbstkonzept als auch das Interesse an Statistik durch die Intervention signifikant mit kleinen bis mittleren Effektstärken erhöhten. Signifikante Unterschiede zwischen den Treatments konnten nicht verzeichnet werden. Dies deutet darauf hin, dass die Auseinandersetzung mit statistischen Problemstellungen anhand von Materialien, die durch handlungsorientierte Aktivitäten, schülernahe Kontexte, Feedbackgabe und lernerzentrierte Arbeitsformen in allen vier Treatments die Erfahrung von Kompetenz, Autonomie und sozialer Eingebundenheit ermöglichen (vgl. Bong & Skaalvik, 2003; Krapp, 2005), für die Förderung des bereichsspezifischen Selbstkonzepts und Interesses geeignet war. Dagegen bewirkten die verschiedenen Treatments, die primär zur differenziellen Förderung der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ sowie von Sichtweisen bezüglich Variabilität konzipiert waren, wie erwartet keine Unterschiede in der motivationalen Entwicklung.

Insgesamt konnten in dieser Studie aufschlussreiche Erkenntnisse bezüglich der Förderung verschiedener Zielvariablen im Bereich von Statistical Literacy durch eine statistikbezogene Intervention erzielt werden. Die vorliegenden Treatments stellten sich in Teilen bzw. im Ganzen als geeignet heraus, Sichtweisen auf Variabilität sowie das statistikbezogene Selbstkonzept und Interesse zu fördern. Die hier zu Grunde liegenden Gestaltungsmerkmale könnten demnach auf weitere Lernmaterialien übertragen werden, die auf die Förderung dieser Variablen abzielen. Auch wenn durch die Intervention insgesamt eine Verbesserung der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ erreicht wurde, konnte das spezielle Förderpotential von Reflexionsanlässen über Reduktion beziehungsweise Variabilität auf diese Kompetenz in dieser Arbeit nicht bestätigt werden. Weiterer Forschungsbedarf bezüglich der Ausrichtung von entsprechenden Förderkonzepten erscheint deshalb als notwendig.

#### **8.4 Weitere Forschung**

Auch wenn diese Dissertation Kompetenz, Sichtweisen und motivationale Variablen im Bereich von Statistical Literacy in den Blick nahm, so ist die Verbindung dieser Bereiche nur in Ansätzen Gegenstand dieser Arbeit gewesen. Ein Zusammenhang zwischen Selbstkonzept beziehungsweise Interesse und der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ konnte in Teilstudie 3 nachgewiesen werden, da Lernende mit höhe-



rer Kompetenz im Vortest eine tendenziell stärkere Erhöhung dieser motivationalen Variablen verzeichnen als ihre schwächeren Peers. Im Allgemeinen wird davon ausgegangen, dass der Zusammenhang zwischen motivationalen Variablen und Leistungsvariablen nicht kausal in eine Richtung geht, sondern dass sich diese wechselseitig beeinflussen (z.B. Krapp, 1998; Marsh et al., 2005; Pekrun et al., 2006; Zirngibl & Pekrun, 2004). Insofern ist zu vermuten, dass motivationale Variablen wie das bereichsspezifische Selbstkonzept und Interesse wiederum eine Wirkung auf die Ausprägung und Entwicklung von Kompetenz haben. Inwieweit das Selbstkonzept und Interesse auch über die in dieser Studie identifizierten Lernervoraussetzungen hinaus geeignet sind, um Kompetenzunterschiede beziehungsweise –entwicklungen vorherzusagen, könnte in weiteren Auswertungsschritten beleuchtet werden. Ergebnisse in diese Richtung würden die Notwendigkeit der Förderung bereichsspezifischer motivationaler Variablen noch weiter unterstreichen. Wie oben bereits erwähnt, könnten diese motivationalen Variablen auch für den beobachteten Geschlechterunterschied verantwortlich sein: Aus der mathematikdidaktischen Forschung ist bekannt, dass Jungen in der Regel nicht nur stärkere Leistungen, sondern auch ein höheres diesbezügliches Selbstkonzept und Interesse vorweisen (Frenzel et al., 2010; Jacobs et al., 2002; Watt, 2004; Weinert, 1997). Vor diesem Hintergrund wäre es denkbar, dass die stärkere Kompetenzausprägung und -entwicklung der Jungen im Verlauf der Intervention über ein höheres bereichsspezifisches Selbstkonzept oder Interesse vermittelt wurde. Diese Anschlussfragen können in weiteren Analyseschritten des vorliegenden Datensatzes untersucht werden und somit das Zusammenspiel der beiden Bereiche weiter beleuchten.

Ebenso ist Anschlussforschung notwendig, um effektive Förderansätze für Kompetenz im Bereich von Statistical Literacy zu identifizieren bzw. die im Rahmen dieser Studie entwickelten Lernmaterialien zu optimieren. Da Wissen und Fähigkeiten bezüglich datenbezogener Reduktion und statistischer Variabilität aus theoretischer Sicht bedeutsam für Statistical Literacy sind, erscheint es sinnvoll an diesen bei der Konzeption von Fördermaterialien festzuhalten. Abgesehen von der Möglichkeit, neue diesbezügliche Fördermaterialien zu konzipieren, legen es die Befunde von Kuntze et al. (2014; vgl. auch Kuntze, 2013) nahe zu untersuchen, ob Reflexionsanlässe bezüglich Reduktion und Variabilität bei älteren Lernenden wirksam sind. Ebenso könnten die bestehenden Materialien in anderen Unterrichtssettings, die den gemeinsamen Austausch über diese Konzepte stärker fokussieren, erprobt werden. Wenn auch auf diesem Weg keine besseren Ergebnisse insbesondere in Bezug auf Aufgaben, die den Umgang mit Variabilität erfordern, erzielt werden, erscheint es lohnenswert, beispielsweise über die Methode des lauten Denkens (z.B. Konrad, 2010) wesentliche Verständnishürden der

Lernenden zu identifizieren, Methoden, um diese zu überwinden, zu entwickeln und diese in die bestehenden Lernmaterialien zu integrieren.

Bei der Konzeption von Fördermaterialien sollte insbesondere darauf geachtet werden, auch den Kompetenzaufbau von Mädchen zu fördern, da diese in der vorliegenden Studie hinter ihren männlichen Klassenkameraden zurückblieben. Studienergebnisse aus mathematikdidaktischer Forschung legen eine Verstärkung von Geschlechterdifferenzen mit zunehmendem Alter auch im Bereich von Statistical Literacy nahe. In Anbetracht des Einflusses von motivationalen Variablen auf Lernen und Leistung könnten entsprechend den Vorschlägen von Carmichael und Hay (2009) bzw. Martignon (2010) Lernkontexte und Methoden gewählt werden, die speziell für Mädchen von Interesse sind.

Für die Evaluation von (neu konzipierten) Förderansätzen in homogenen Stichproben wie der vorliegenden könnte es sinnvoll sein, ein stärker differenzierendes Testinstrument zu entwickeln, um Unterschiede zwischen Individuen und Messzeitpunkten trennschärfer ermitteln zu können. Bei einem solchen jahrgangsstufenspezifischen Testinstrument sollte eine größere Anzahl und Bandbreite von Aufgaben auf den für die jeweilige Stichprobe relevanten Niveaustufen vorhanden sein, da dadurch feinere Veränderungen im Verlauf einer Intervention feststellbar sein könnten. Ein solches Testinstrument könnte auch zur Optimierung der im Rahmen dieser Dissertation entwickelten Lernmaterialien dienen.

Die vorliegende Dissertation konnte verschiedene Bedingungsvariablen für die Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ identifizieren. Dabei konnte nicht abschließend geklärt werden, welche Variablen Kompetenzunterschiede im Bereich von Statistical Literacy zwischen Klassen zu erklären vermögen. Dies wäre aber insbesondere von Interesse, da sich diesbezügliche Klassenunterschiede in verschiedenen Stichproben gezeigt haben (Sproesser et al., 2014). Vor dem Hintergrund der Ergebnisse aus Teilstudie 1 ist zu beleuchten, ob der durchschnittliche sozioökonomische Status tatsächlich zur Erklärung von Klassenunterschieden geeignet ist oder ob es sich dabei um einen Drittvariableneffekt handelt. Dazu sollten geeignete Variablen auf Klassenebene, beispielsweise den Lehrer oder den Unterricht betreffend, erhoben werden. Letztendlich könnte in diesem Zusammenhang auch geklärt werden, ob es sich bei den von Sproesser et al. (2014) beobachteten stärkeren Klassenunterschieden in Klassenstufe neun um eine tatsächliche Erhöhung mit fortschreitender Unterrichtszeit oder um einen Zufallseffekt handelte.

Zuletzt ist hervorzuheben, dass einige der im Rahmen dieser Studie erhobenen Variablen (siehe Kapitel 4) bisher nicht ausgewertet werden konnten und demnach Gegenstand zukünftiger Auswertungen sein können. Diese Variablen betrafen - wie einige in diesem Abschnitt genannten, weiteren Fragestellungen - nicht den Hauptfokus dieser Dissertation.

## **8.5 Grenzen der Studie**

Neben den gegebenenfalls im Rahmen der jeweiligen Teilstudie angegebenen Grenzen müssen einige Einschränkungen bei der Interpretation der Ergebnisse dieser Dissertation berücksichtigt werden. Diese werden im Folgenden erläutert.

### ***8.5.1 Einschränkungen der Testinstrumente***

Insbesondere im Rahmen von Teilstudie 1 und 2 wurde das Testinstrument zur Ermittlung der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ durch IRT-basierte Methoden untersucht. Es stellte sich heraus, dass der verwendete Test in der vorliegenden Stichprobe zu wenig zu differenzieren scheint: Die Aufgaben von Kompetenzniveau I und II konnten von einem Großteil der Probanden gelöst werden, während die Aufgaben der höchsten beiden Niveaus geringe Lösungsraten vorwiesen. Es wurden also manche Niveaus sehr gut, andere kaum erreicht. Diese Boden- und Deckeneffekte könnten möglicherweise zu reduzierten Trennschärfen und der recht niedrigen EAP/PV-Reliabilität geführt zu haben. Dies bedeutet, dass mit dem verwendeten Testinstrument die Kompetenz zwischen verschiedenen Lernenden dieser Stichprobe bzw. zwischen verschiedenen Messzeitpunkten nicht sehr trennscharf ermittelt werden konnte und somit auch Befunde bezüglich der Entwicklung der Kompetenz im Verlauf der Intervention vorsichtig interpretiert werden sollten.

Da sich die Lernenden der beteiligten achten Realschulklassen anscheinend insgesamt recht ähnlich waren, machten sich Kompetenzunterschiede im Wesentlichen auf Kompetenzniveau III bemerkbar. Um in Folgestudien die Wirkung einer Intervention in eher homogenen Stichproben differenzierter analysieren zu können, erscheint die Erweiterung des verwendeten Testinstruments beispielsweise um zusätzliche Aufgaben auf Kompetenzniveau III sinnvoll (siehe auch 8.4). Im Hinblick auf die Vergleichbarkeit bietet das verwendete Testinstrument jedoch den Vorteil, dass die Kompetenzausprägungen in der vorliegenden Stichprobe mit der Kompetenz von Lernenden aus Vorgängeruntersuchungen (z.B. Kuntze et al., 2014) verglichen werden kann.

Auch bezüglich anderer Erhebungsinstrumente müssen einschränkende Bemerkungen gemacht werden. Beispielsweise wurde der sozioökonomische Status nur über ein Item abge-

fragt. Obwohl Paulus (2009) diesem eine hohe Aussagekraft zumisst, sollten diesbezügliche Ergebnisse mit Vorsicht interpretiert werden. Da verbale und nonverbale kognitive Fähigkeiten ebenso nur über je eine Subskala des Kognitiven Fähigkeitstests KFT 4-12+R (Heller & Perleth, 2000) erhoben wurden, sollten auch diese Werte lediglich als Indikatoren für kognitive Leistungsfähigkeit angesehen werden. Analoge Überlegungen gelten für den relativ kurzen Leseverständnistest LGVT 6-12 (Schneider et al., 2007). Insofern sind die ermittelten Zusammenhänge zwischen Leseverständnis bzw. kognitiven Fähigkeiten und der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ stets nur als Tendenz zu betrachten. Einschränkungen müssen ebenso bezüglich des Testinstruments zur Erhebung von Sichtweisen auf Variabilität gemacht werden. Durch dieses wird lediglich ein Aspekt der Berücksichtigung zufallsbedingter Variabilität ermittelt. Zudem wären auch andere Möglichkeiten der Bildung eines Gesamtscores denkbar gewesen, beispielsweise durch den Einbezug von mehr oder weniger Abstufungen. Auch die im Zusammenhang mit Sichtweisen auf Variabilität ermittelten Befunde sind dementsprechend lediglich im Sinne eines Trends zu verstehen.

### ***8.5.2 Repräsentativität der Stichprobe***

Die 25 Interventionsklassen dieser Studie stammen aus acht Schulen in Baden-Württemberg. Von vier dieser Schulen nahmen alle parallelen 8. Klassen an der Intervention teil, aus den vier weiteren Schulen konnte jeweils eine Klasse nicht mitwirken. Die Nicht-Teilnahme lag laut dem jeweiligen Mathematiklehrer an einem Mangel an verbleibender Unterrichtszeit. Dennoch kann nicht ausgeschlossen werden, dass es dadurch zu bewussten oder unbewussten Selektionseffekten kam. Auch bei der Rekrutierung der Schulen sind Selektionseffekte nicht ausgeschlossen, da wissenschaftliche Studien an Realschulen in Baden-Württemberg zunächst von der Schulleitung genehmigt werden müssen und erst dann an die betroffenen Lehrer weitergegeben werden. Die letztendlich an dieser Studie teilnehmenden Schulen und Lehrer wurden insofern nicht randomisiert ausgewählt, sondern zeichneten sich durch eine grundsätzliche Offenheit gegenüber empirischer Unterrichtsforschung aus. Bei der Auswahl der Schulen wurde auf ein ausgeglichenes Verhältnis von ländlichem und städtischem Einzugsgebiet geachtet. Dennoch kann nicht grundsätzlich davon ausgegangen werden, dass die hier erzielten Ergebnisse uneingeschränkt verallgemeinerbar sind, da es sich um keine repräsentative Stichprobe deutscher (Real-)Schüler handelt.

### ***8.5.3 Beständigkeit der geschilderten Entwicklungen***

Die vorliegende Dissertation beschreibt in Teilstudie 2 und 3, wie sich Sichtweisen auf zufallsbedingte Variabilität, Kompetenz im Bereich von Statistical Literacy sowie das statistik-

bezogene Selbstkonzept und Interesse im Verlauf der Intervention entwickelten. Dabei muss berücksichtigt werden, dass die Durchführung des Nachtests immer in der ersten Mathematikstunde nach der vierstündigen Intervention, in der Regel zu Beginn der Folgewoche, stattfand. Die gefundenen Effekte sind deshalb zunächst als kurzfristige Interventionseffekte zu interpretieren und erlauben keine Aussagen über langfristige Entwicklungen. Aufgrund nicht zu kontrollierender Unterrichtsbedingungen zwischen Nach- und Follow-Up-Test können auch durch die Analyse des Follow-Up-Tests keine Kausalaussagen über die langfristige Entwicklung gemacht werden. Insofern ist die Auswertung des Follow-Up-Tests nicht Gegenstand dieser Dissertation, die sich im Wesentlichen mit der Untersuchung von direkten Interventionseffekten beschäftigt.

### **8.6 Abschließende Bemerkungen**

Trotz einiger Einschränkungen hat diese Arbeit wichtige Erkenntnisse für Theorie und Praxis hervorgebracht. Die Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ konnte im Zusammenspiel mit einigen Eingangsvoraussetzungen der Lernenden analysiert werden. Dadurch wurden bestehende theoretische und empirische Arbeiten teilweise bestätigt, aber auch relativiert. Diese Ergebnisse können neben der Verbreiterung der fachdidaktischen Theorie auch bei der Konzeption von Lernmaterialien genutzt werden: Beispielsweise könnten Materialien konzipiert werden, die speziell für Mädchen ansprechend sind. Es hat sich aber auch gezeigt, dass lernförderliche Materialien durchaus textorientiert angelegt sein dürfen. Zielen Fördermaterialien auf Sichtweisen bezüglich Variabilität oder auf motivationale Variablen ab, so können die in diesen Materialien umgesetzten Gestaltungsmerkmale relativ direkt auf andere Materialien übertragen werden.

### **9 Darlegung des eigenen Anteils an dieser Studie**

Die vorliegende Dissertation entstand in einem Teilprojekt des Kooperativen Promotionskollegs „Effektive Lehr- Lernarrangements“ der Pädagogischen Hochschule Ludwigsburg und der Universität Tübingen, das vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg gefördert wurde. In diesem Teilprojekt waren hauptsächlich Ute Sproesser, die mit dem Ziel der Promotion von August 2011 bis Juli 2014 an die Pädagogische Hochschule Ludwigsburg abgeordnet war, sowie die beiden Betreuer der Promotion Prof. Dr. Joachim Engel und Prof. Dr. Sebastian Kuntze tätig. Das in dieser Studie eingesetzte Testinstrument zur Messung der Kompetenz „Nutzen von Darstellungen und Modellen in statistischen Kontexten“ wurde im Vorfeld in Arbeitsgruppen um Prof. Dr. Sebastian Kuntze an der Ludwig-Maximilians-Universität München (Projekt KOMMA) und der Pädagogischen Hoch-

schule Ludwigsburg (Projekt RIKO-STAT) entwickelt und in verschiedenen Studien bereits eingesetzt. Die Auswahl und Zusammenstellung bzw. Konzeption der weiteren Fragebogenteile erfolgte in Zusammenarbeit zwischen Ute Sproesser, Prof. Dr. Joachim Engel, Prof. Dr. Sebastian Kuntze und Prof. Dr. Hans-Christoph Nürk: Items und Tests zur Erhebung des sozioökonomischen Status<sup>8</sup>, des Arbeitsgedächtnisses, von kognitiven Fähigkeiten und Lesegeschwindigkeit bzw. -verständnis wurden von Prof. Dr. Hans-Christoph Nürk empfohlen. Skalen zur Erfassung mathematik- und statistikbezogener motivationaler Variablen wurden auf Anraten von Prof. Dr. Sebastian Kuntze einbezogen. Ein Item zur Erfassung von Sichtweisen auf Variabilität wurde aus früherer Forschung von Prof. Dr. Joachim Engel und Prof. Dr. Sebastian Kuntze übernommen, drei weitere diesbezügliche Items sowie Items zur Erfassung von aufgabenspezifischer Motivation wurden von Ute Sproesser neu konzipiert. Ebenso wurden die hier verwendeten Lernmaterialien in Kooperation zwischen Ute Sproesser, Prof. Dr. Joachim Engel und Prof. Dr. Sebastian Kuntze entwickelt.

Für die Rekrutierung der an der Studie teilnehmenden Schülerinnen und Schüler aus 29 verschiedenen Schulklassen, und für die Organisation und Durchführung der Pilotierung sowie der Hauptstudie war Ute Sproesser zuständig. Bei der Auswertung der Daten wurde die Doktorandin durch die Expertise ihrer Betreuer sowie von Methodik-Experten des Promotionskollegs und der PH Ludwigsburg durch Diskussion der Ergebnisse und weiterer Analysemöglichkeiten unterstützt. Die in dieser Dissertation abgedruckten wissenschaftlichen Artikel wurden von Ute Sproesser in enger Zusammenarbeit mit ihren beiden Betreuern Prof. Dr. Joachim Engel und Prof. Dr. Sebastian Kuntze verfasst. Bei der Überarbeitung des Artikels „Einflussfaktoren auf die Entwicklung von Statistical Literacy und von Sichtweisen auf zufallsbedingte Variabilität in der achten Realschulklasse“ wirkte außerdem Prof. Dr. Laura Martignon als „action editor“ mit. In die hier abgedruckten Versionen aller drei Artikel flossen zudem die Anmerkungen und Vorschläge von Reviewern und Herausgebern der jeweiligen Zeitschriften sowie von Prof. Dr. Robert delMas ein. Darüber hinaus gingen in die Abfassung der Aufsätze wertvolle Anregungen ein, die Ute Sproesser bei der Vorstellung von Zwischenergebnissen auf mehreren nationalen und internationalen Tagungen erhalten hat.

---

<sup>8</sup> Weitere Informationen und Verweise bezüglich der Herkunft etablierter Skalen bzw. bezüglich des Einsatzes eigener Skalen in früheren Studien sind dem Methodenteil zu entnehmen.

## 10 Literatur

- Agresti, A. & Franklin, C. (2013). *Statistics: The art and science of learning from data*. Pearson: Boston.
- Arbuckle, J. L. (1995-2010). *IBM SPSS AMOS 19.0 user's guide*. Chicago, IL: Small-Waters Corporation.
- Artelt, C., Baumert, J., Julius-McElvany, N. & Peschar, J. (2003). *Learners for life: Student approaches to learning. Results from PISA 2000*. Paris: OECD.
- Australian Education Council (1991). *A national statement on mathematics for Australian schools*. Carlton: Curriculum Corporation.
- Batanero, C., Burril, G. & Reading, C. (2011). *Teaching statistics in school mathematics – Challenges for teaching and teacher education*. Dordrecht: Springer Science + Business Media.
- Baum, D. & Klein, H. (Hrsg.). (2004). *Mathematik 1 XQuadrat*. München: Oldenbourg Schulbuchverlag.
- Baumert, J., Bos, W. & Lehmann, R. (2000). *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn*. Opladen: Leske + Budrich.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.J. & Weiß, M. (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Baumert, J. & Köller, O. (2000). Unterrichtsgestaltung, verständnisvolles Lernen und multiple Zielerreichung im Mathematik- und Physikunterricht der gymnasialen Oberstufe. In J. Baumert, W. Bos & R. Lehmann (Eds.), *TIMSS III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn* (S. 271-316) Opladen: Leske und Budrich.
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review* 84(2), 191–215.
- Ben-Zvi, D., & Garfield, J. (2005). Statistical literacy, reasoning, and thinking: goals, definitions, and challenges. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing*

- statistical literacy, reasoning, and thinking* (pp. 3-15). Dordrecht, The Netherlands: Kluwer Academic Publishers (Springer).
- Bidgood, P. (2014). Towards statistical literacy – relating assessment to the real world. *Proceedings of the 9th International Conference in Teaching Statistics ICOTS-9*, Flagstaff, USA. Voorburg, The Netherlands: International Statistical Institute. [http://icots.info/9/proceedings/pdfs/ICOTS9\\_7C1\\_BIDGOOD.pdf](http://icots.info/9/proceedings/pdfs/ICOTS9_7C1_BIDGOOD.pdf) [16.12.2015].
- Biehler, R. & Hartung, R. (2006). Die Leitidee Daten und Zufall. In W. Blum, C. Drüke-Noe, R. Hartung & O. Köller (Hrsg.), *Bildungsstandards Mathematik: konkret. Sekundarstufe I* (S. 51–80). Berlin: Cornelsen Scriptor.
- Biggs, J. & Collis, K. (1982). *Evaluating the quality of learning: the SOLO taxonomy*. New York: Academic.
- Blum, W. & Leiss, D. (2005). Modellieren im Unterricht mit der „Tanken“-Aufgabe. *mathematik lehren*, 128, 18-21.
- Blum, W., Neubrand, M., Ehmke, T., Senkbeil, M., Jordan, A., Ulfig, F. & Carstensen, C.H. (2004). Mathematische Kompetenz. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff & U. Schiefele, *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 47-92). Münster: Waxmann.
- Böttner, J., Maroska, R., Olpp, A., Pongs, R., Stöckle, C., Wellstein, H. & Wontroba, H. (2005). *Schnittpunkt 1*. Stuttgart: Ernst Klett.
- Bollen, K. & Long, J. (1993). *Testing structural equation models*. Newbury Park, CA: Sage.
- Bond, M., Perkins, S.N. & Ramirez, C. (2012). Students' perceptions of statistics: An exploration of attitudes, conceptualizations, and content knowledge of statistics. *Statistics Education Research Journal*, 11(2), 6-25. <http://www.stat.auckland.ac.nz/serj> [17.07.2014].
- Bong, M. & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: how different are they really? *Educational Psychology Review*, 15(1), 1-40.
- Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241-258). New York: Greenwood.



- Bradley, R.H. & Corwyn, R.F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, 53, 371–399.
- Breiman, L., Friedman, J., Olshon, R. Stone, C. (1984). *Classification and regression trees*. Wadsworth: Belmont, CA.
- Broers, N. (2006). Learning Goals: The primacy of statistical knowledge. *Proceedings of the 7th International Conference in Teaching Statistics ICOTS-7*, Salvador, Brazil. [http://iase-web.org/documents/papers/icots7/6G2\\_BROE.pdf](http://iase-web.org/documents/papers/icots7/6G2_BROE.pdf) [16.12.2015].
- Brophy, J.E. & Good, T.L. (1986). Teacher behavior and student achievement. In M.C. Wittrock (Ed.), *Handbook of Research on Teaching* (pp. 328-375). New York: Macmillan.
- Brunner, M. (2008). No g in education? *Learning and individual differences* 18 (2008), 152-165.
- Brunner, M. (2005). *Mathematische Schülerleistung: Struktur, Schulformunterschiede und Validität*. Humboldt Universität zu Berlin. <http://edoc.hu-berlin.de/dissertationen/brunner-martin-2006-02-08/PDF/brunner.pdf> [20.05.2014].
- Carmichael, C., Callingham, R., Hay, I & Watson, J. (2010). Statistical literacy in the middle school: The relationship between interest, self-efficacy and prior mathematics achievement. *Australian Journal of Educational & Developmental Psychology*. Vol 10, 2010, 83-93.
- Carmichael, C., Callingham, R., Watson, J., & Hay, I. (2009). Factors influencing the development of middle school students' interest in statistical literacy. *Statistics Education Research Journal*, 8(1), 62-81.
- Carmichael, C. & Hay, I. (2009). Gender differences in middle school students' interest in a statistical literacy context. In R. Hunter, B. Bicknell, & T. Burgess (Eds.), *Crossing divides: Proceedings of the 32nd annual conference of the Mathematics Education Research Group of Australasia* (Vol. 1). Palmerston North, NZ: MERGA.
- Chance, B.L. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education*, 10(3). <http://www.amstat.org/publications/jse/v10n3/chance.html> [10.09.2014].

- Cobb, G.W. & Moore, D.S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, 104(9), 801-823. <http://www.macalester.edu/~kaplan/statchat/2009-2010-year/2010-01/January-2010.pdf> [15.08.2014].
- Cohen, J. (1988): *Statistical Power Analysis for the Behavioral Sciences*, 2. Edition. Hillsdale: Lawrence Erlbaum Associates.
- Curcio, F.R. (2010). *Developing data-graph comprehension in grades K-8*. Reston, VA: National Council of Teachers of Mathematics.
- Curcio, F.R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for research in mathematics education*, 18 (5), 382 – 393.
- Ditton, H. (1998). *Mehrebenenanalysen. Grundlagen und Anwendungen des Hierarchischen Linearen Modells*. Weinheim: Juventa.
- Deci, E.L. & Ryan, R.M. (2002). *Handbook of self-determination research*. Rochester: University of Rochester.
- delMas, R. (2002). Statistical literacy, reasoning and learning: A commentary. *Journal of Statistics Education* 10(3). [http://www.amstat.org/publications/jse/v10n3/delmas\\_discussion.html](http://www.amstat.org/publications/jse/v10n3/delmas_discussion.html) [10.09.2014].
- delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, 6(2), 28-58. <http://www.stat.auckland.ac.nz/serj> [28.8.2014].
- Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. *Educational studies in mathematics*, 61, 103-131.
- Eccles, J.S. & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology* 53, 109-132.
- Ehmke, T., Hohensee, F., Siegle, T. & Prenzel, M. (2006). Soziale Herkunft, elterliche Unterstützungsprozesse und Kompetenzentwicklung. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost & U. Schiefele (Hrsg.), *PI-SA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 225-248). Münster: Waxmann.
- Eichler, A. & Vogel, M (2012). Basic modelling of uncertainty: young students' mental models. *ZDM Mathematics Education*, 44, 841-854.

- Eid, M., Gollwitzer, M. & Schmitt, M. (2011). *Statistik und Forschungsmethoden*. Weinheim: Beltz.
- Emmioglu, E. & Capa-Aydin, Y. (2012). Attitudes and achievement in statistics: A meta-analysis study. *Statistics Education Research Journal*, 11(2), 95-102. <http://www.stat.auckland.ac.nz/serj> [17.07.2014].
- Enders, C.K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12(2), 121-138.
- Engel, J. (2014). Change point detection tasks to explore students' informal inferential reasoning. In T. Wassong, D. Frischmeier, P.R. Fischer, R. Hochmuth & P. Bender (Hrsg.), *Mit Werkzeugen Mathematik und Stochastik lernen* (S. 113-125). Wiesbaden: Springer Spektrum.
- Engel, J. (2011). Über ikonische Repräsentationen von zufallsbedingter Variabilität. In R. Haug & L. Holzäpfel, *Beiträge zum Mathematikunterricht 2011* (S. 235-238). Münster: WTM.
- Engel, J. & Kuntze, S. (2011). From data to functions: Connecting modelling competencies and statistical literacy. In: G. Kaiser et al. (Eds.), *Trends in teaching and learning of mathematical modelling* (pp. 397-406) Heidelberg: Springer.
- Engel, J. & Sedlmeier, P. (2005). On middle-school students' comprehension of randomness and chance variability in data. *Zentralblatt Didaktik der Mathematik*, 37(3), 168-177.
- Engel, J., Sedlmeier, P. & Wörn, C. (2008). Modeling scatterplot data and the signal-noise metaphor. Towards statistical literacy for pre-service teachers. *Proceedings of the ICMI Study 18 Conference and IASE 2008 Round Table Conference*. Monterrey: ICMI/IASE.
- Engel, J. & Vogel, M. (2005). Von M&Ms und bevorzugten Farben: ein handlungsorientierter Unterrichtsvorschlag zur Leitidee Daten und Zufall in der Sekundarstufe I. *Stochastik in der Schule* 25(2), 11-18.
- Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Reidel: Dordrecht.
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) pro-*

- ject: A pre-k-12 curriculum framework. <http://www.amstat.org/education/gaise/> [04.09.2014].
- Frensch, P.A. & Funke, J. (1995). Definitions, traditions, and a general framework for understanding complex problem solving. In P.A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 3-25). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Frenzel, A., Goetz, T., Pekrun, R., & Watt, H.M.G. (2010). Development of mathematics interest in adolescence: Influence of gender, family, and school context. *Journal of Research on Adolescence*, 20(2), 507-537.
- Fröhlich, A., Kuntze, S. & Lindmeier, A. (2007). Testentwicklung und -evaluation im Bereich von „Statistical Literacy“. *Beiträge zum Mathematikunterricht 2007* (S. 783-786). Hildesheim: Franzbecker.
- Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review* 70, 1-51.
- Gal, I., Ginsburg, L., & Schau, C. (1997). Monitoring attitudes and beliefs in statistics education. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 37–51). Amsterdam: IOS Press. <http://www.stat.auckland.ac.nz/~iase/publications/assessbk/chapter04.pdf> [06.08.2014].
- Gal, I., & Murray, S. (2011). Users' statistical literacy and information needs: Institutional and educational implications. *Statistical Journal of the IAOS*, 27(3-4), 185-195.
- Garfield, J., Aliaga, M., Cobb, G., Cuff, C., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, R., Utts, J., Velleman, P., & Witmer, J. (2005). *Guidelines for assessment and instruction in statistics education (GAISE) project: College report*. <http://www.amstat.org/education/gaise> [04.09.2014].
- Garfield, J., & delMas, R. (2010). A web site that provides resources for assessing students' statistical literacy, reasoning and thinking. *Teaching Statistics*, 32(1), 1-7.
- Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM – The International Journal on Mathematics Education*, 44(7), 883-898.

- Garfield, J. & Gal, I. (1999). Teaching and assessing statistical reasoning. In L. Stiff (Ed.), *Mathematical reasoning* (pp. 207-219). Reston, Virginia: National Council Teachers of Mathematics.
- Gläser-Zikuda, M., Fuß, S., Laukenmann, M., Metz, K., & Randler, C. (2005). Promoting students' emotions and achievement—Instructional design and evaluation of the ECOLE approach. *Learning and Instruction, 15*, 481–495.
- Green, D.R. (1990). *A longitudinal study of pupils' probability concepts*. Loughborough: Loughborough University.
- Green, D.R. (1986). Children's understanding of randomness. In R. Davidson & J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics* (pp. 287-291). Victoria, British Columbia: University of Victoria.
- Green, D.R. (1982). A survey of probabilistic concepts in 3000 students aged 11-16 years. In D.R. Green (Ed.), *Proceedings of the First International Conference on Teaching Statistics* (pp. 766-783). University of Sheffield: Teaching Statistics Trust.
- Griesel, H., Postel, H. & vom Hofe, R. (2004). *Mathematik heute 1*. Braunschweig: Schroedel.
- Gundlach, M., Kuntze, S., Engel, J., & Martignon, L. (2010). Motivation and self-efficacy related to probability and statistics: Task-specific motivation and proficiency. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the 8th Int. Conf. on Teaching Statistics*. Voorburg, The Netherlands: ISI. [www.stat.auckland.ac.nz/~iase/publications.php](http://www.stat.auckland.ac.nz/~iase/publications.php) [04.06.2014].
- Hannover, B. (2010). Lernen Mädchen anders? In M. Matzner & I. Wyrobnik (Hrsg.), *Handbuch Mädchen-Pädagogik* (S. 95-107). Weinheim: Beltz.
- Hansford, B.C. & Hattie, J.A. (1982). The relationship between self and achievement / performance measures. *Review of Educational Research 52*, 123-142.
- Hattie, J., Biggs, J. B., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research, 66*, 99–136.
- Heller, A. K. & Perleth, C. (2000). *KFT 4-12+R. Manual*. Göttingen: Hogrefe.

- Helmke, A. & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie, Band 3 (Psychologie der Schule und des Unterrichts)* (S. 71-176). Göttingen: Hogrefe-Verlag.
- Helmstaedter, C., Lendt, M. & Lux, S. (2001). *Verbaler Lern- und Merkfähigkeitstest: VLMT, Manual*. Weinheim: Beltz-Verlag.
- Hidi, S. & Renninger, K.A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111-127.
- Holling, H., Preckel, F. & Vock, M. (2004) *Intelligenzdiagnostik*. Kompendien Psychologische Diagnostik, Bd. 6. Göttingen: Hogrefe.
- Holmes, P. (1980). *Teaching statistics 11-16*. Berkshire: Schools Council and Foulsham Education.
- Hood, M., Creed, P. & Neumann, D.L. (2012). Using the expectancy value model of motivation to understand the relationship between student attitudes and achievement in statistics. *Statistics Education Research Journal*, *11*(2), 72-85. <http://www.stat.auckland.ac.nz/serj> [17.07.2014].
- Hosenfeld, I., Köller, O. & Baumert, J. (1999). Why sex differences in mathematics achievement disappear in German secondary schools: A reanalysis of the German TIMSS-data. *Studies in Educational Evaluation*, *25*, 143-162.
- Hox, J. (2010). *Multilevel analysis. Techniques and applications*. New York: Routledge.
- Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis. Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. <http://dx.doi.org/10.1080/10705519909540118> [20.05.2014].
- Jacobs, E.J., Lanza, S., Osgood, D.W., Eccles, J.S. & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development* *73*(2), 509-527.
- Jones, G.A., Langrall, C.W., Mooney, E.S. & Thornton, C.A. (2005). Models of development in statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 97-117). Dordrecht, The Netherlands: Kluwer Academic Publishers (Springer).

- Kamil, M. Mosenthal, P., Pearson, P. & Barr, R. (2000). *Handbook of reading research, III*. New York: Routledge.
- Kaun, Andreas (2006). Stochastik in deutschen Lehrplänen allgemeinbildender Schulen. *Stochastik in der Schule* 26 (3), 11–17.
- Kirsch, I.S., Jungeblut, S.S. & Mosenthal, P.M. (1998). The measurement of adult literacy. In S.T. Murray, I.S. Kirsch & L.B. Jenkins (Eds.), *Adult literacy in OECD countries: Technical report on the first International Adult Literacy Survey* (pp. 105-134). Washington, DC: National Center of Education Statistics, U.S. Department of Education.
- Klieme, E. (1997). Gender-related differences in mathematical abilities: Effect-size, spatial mediation, and item content. *Paper presented at the 7<sup>th</sup> Conference of the European Association for Research in Learning and Instruction (EARLI)*, Athens, Greece.
- Klieme E., Funke, J., Leutner, D, Reimann, P. & Wirth, J (2001). Problemlösen als fächerübergreifende Kompetenz – Konzeption und erste Resultate aus einer Schulleistungsstudie. *Zeitschrift für Pädagogik* 47 (2001), 179–200.
- Klieme, E., Neubrand, M. & Lüdtke, O. (2001). Mathematische Grundbildung: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K. J. Tillmann & M. Weiß (Eds.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 141-191). Opladen: Leske + Budrich.
- KMK (Kultusministerkonferenz) (2003). *Bildungsstandards im Fach Mathematik für den mittleren Schulabschluss*. München: Wolters Kluwer.
- Konold, C. & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education* 33(4), 259-289.
- Konrad, K. (2010). Lautes Denken. In G. Mey & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 476-490). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Krapp, A. (1998). Entwicklung und Förderung von Interesse im Unterricht. *Psychologie in Erziehung und Unterricht*, 45, 186-203.
- Krapp, A. (2005). Basic needs and the development of interest and intrinsic motivational orientations. *Learning and Instruction* 15, 381-395.

- Krapp, A. (2007). An educational-psychological conceptualisation of interest. *International Journal of Educational and Vocational Guidance*, 7(1), 5-21.
- Kröpfl, B., Peschek, W., Schneider, E. (2000). Stochastik in der Schule: Globale Ideen, lokale Bedeutungen, zentrale Tätigkeiten. *mathematica didactica*, 23, 25–57.
- Kuntze, S. (2013). Modellieren beim Nutzen von Darstellungen in statistischen Kontexten. Hierarchische Beschreibung und Bedingungsvariablen eines Aspekts mathematischer Kompetenz. In R. Borromeo Ferri, G. Greefrath & G. Kaiser (Eds.), *Mathematisches Modellieren in Schule und Hochschule* (S. 71-94). Wiesbaden: Springer Spektrum.
- Kuntze, S. (2012). Sichtweisen von Lernenden zu statistischer Variabilität – Vorstellungen von Grundschüler(inne)n, Realschüler(inne)n und Studierenden. In M. Ludwig & M. Kleine (Hrsg.), *Beiträge zum Mathematikunterricht 2012* (S. 505-508). Münster: WTM.
- Kuntze, S., Engel, J., Martignon, L. & Gundlach, M. (2010). Aspects of statistical literacy between competency measures and indicators for conceptual knowledge – Empirical research in the project RIKO-STAT. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. [http://iase-web.org/documents/papers/icots8/ICOTS8\\_C159\\_KUNTZE.pdf](http://iase-web.org/documents/papers/icots8/ICOTS8_C159_KUNTZE.pdf) [14.6.2013].
- Kuntze, S., Lindmeier, A. & Reiss, K. (2008a). „Daten und Zufall“ als Leitidee für ein Kompetenzstufenmodell zum „Nutzen von Darstellungen und Modellen“ als Teilkomponente von Statistical Literacy. *Anregungen zum Stochastikunterricht*, 4 (S. 111-122). Hildesheim: Franzbecker.
- Kuntze, S., Lindmeier, A. & Reiss, K. (2008b). “Using models and representations in statistical contexts” as a sub-competency of statistical literacy – Results from three empirical studies. *Proceedings of the 11th International Congress on Mathematical Education*. Monterrey (Mexico). <http://tsg.icme11.org/document/get/474> [22.12.11].
- Kuntze, S., Vargas, F., Martignon, L. & Engel, J. (2014). Competencies in understanding statistical information in primary and secondary school levels: An inter-cultural empirical study with German and Colombian students. *Avances de Investigación en Educación Matemática (AIEM)*. 7, 5-25.



- Land Baden-Württemberg (2004). *Bildungsstandards für Mathematik Realschule - Klassen 6, 8, 10*. [http://www.bildung-staerkt-menschen.de/service/downloads/Bildungsstandards/Rs/Rs\\_M\\_bs.pdf](http://www.bildung-staerkt-menschen.de/service/downloads/Bildungsstandards/Rs/Rs_M_bs.pdf). [15.11.11].
- Land Baden-Württemberg (2003). *Verordnung des Kultusministeriums über die Erste Staatsprüfung für das Lehramt an Realschulen*. <http://www.landesrecht-bw.de/jportal/?quelle=purl&psml=bsbawueprod.psml&max=true&docId=jlr-RSchulLehr1StPrOBWrahmen&doc.part=X> [28.5.2015].
- Land Baden-Württemberg (1999). *Verordnung des Kultusministeriums über die Erste Staatsprüfung für das Lehramt an Realschulen*. [http://www.llpa-bw.de/site/pbs-bw/get/documents/KULTUS.Dachmandant/KULTUS/import/kv/pdf/RPO-I-1999\\_AUSSERKRAFT.pdf](http://www.llpa-bw.de/site/pbs-bw/get/documents/KULTUS.Dachmandant/KULTUS/import/kv/pdf/RPO-I-1999_AUSSERKRAFT.pdf) [28.5.2015].
- Land Baden-Württemberg (1994). Bildungsplan für die Realschule. *Kultus und Unterricht. Amtsblatt des Ministeriums für Kultus und Sport Baden-Württemberg*. <http://www.lsbw.de/bildungsplaene/allgbilschulen/lp1994/bprs.pdf> [27.5.2015].
- Lehrer, R. (2014). *Data modeling supports development of statistical reasoning*. [http://peabody.vanderbilt.edu/research/pro/about\\_peabody\\_research/funded\\_projects/data\\_modeling\\_\\_home.php](http://peabody.vanderbilt.edu/research/pro/about_peabody_research/funded_projects/data_modeling__home.php) [05.09.2014].
- Leiss, D., Schukajlow, S., Blum, W., Messner, R. & Pekrun, R. (2010). The role of the situation model in mathematical modelling – Task analyses, student competencies, and teacher interventions. *Journal für Mathematik-Didaktik*, 31(1), 119-141.
- Lindmeier, A., Kuntze, S. & Reiss, K. (2007). Representations of data and manipulations through reduction – competencies of German secondary students. In B. Philips & L. Weldon (Eds.), *Proceedings of the IASE/ISI Satellite Conference on Statistical Education, Guimarães, Portugal, 19-21 August 2007*. Voorburg, NL: International Statistical Institute. [http://www.stat.auckland.ac.nz/~iase/publications/sat07/Lindmeier\\_et\\_al.pdf](http://www.stat.auckland.ac.nz/~iase/publications/sat07/Lindmeier_et_al.pdf) [16.11.11].
- Marsh, H.W. & Craven, R. (1997). Academic self-concept: Beyond the dustbowl. In G. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement and adjustment* (pp. 131-198). Orlando, FL: Academic Press.

- Marsh, H.W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Integration of multidimensional self-concept and core personality constructs: Construct validation and relations to well-being and achievement. *Journal of Personality, 74*, 403-456.
- Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: reciprocal effects models of causal ordering. *Child Development, 76*(2), 397-416.
- Martignon, L. (2010). Mädchen und Mathematik. In M. Matzner & I. Wyrobnik (Hrsg.), *Handbuch Mädchen-Pädagogik* (S. 220-232). Weinheim: Beltz.
- McLeod, D.B. (1992). Research on affect in mathematics education: A reconceptualization. In D.A. Grouws (Ed.), *Handbook of research on mathematics, teaching and learning* (pp. 575-596). New York: Macmillan.
- Ministry of Education (1992). *Mathematics in the New Zealand Curriculum*. Wellington: Learning Media.
- Möller, J. & Trautwein, U. (2009). Selbstkonzept. In E. Wild & J. Möller, *Pädagogische Psychologie* (S. 180–203). Heidelberg: Springer.
- Moore, D. (1997): New pedagogy and new content: The case of statistics, *International Statistical Review, 65*(2),123-165.
- Moore, P. (1990). Uncertainty. In L. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95-137). Washington, DC: National Academy Press.
- Muthén, L.K. and Muthén, B.O. (1998-2012). *Mplus user's guide. Seventh edition*. Los Angeles, CA: Muthén & Muthén.
- Nagy, G., Trautwein, U., Baumert, J., Köller, O., & Garrett, J. (2006). Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. *Educational Research and Evaluation, 12*, 323-345.
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for K–12 mathematics*. Reston VA: The Council.
- OECD (2003). *The PISA 2003 Assessment Framework. Mathematics, Reading, Science and Problem Solving Knowledge and Skills*. Paris: OECD.

- O'Mara, A.J., Marsh, H.W., Craven, R.G., & Debus, R.L. (2006). Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis. *Educational Psychologist*, 41(3), 181-206.
- Opdenakker, M.C. & Van Damme, J. (2001). Relationship between school composition and characteristics of school process and their effect on mathematics achievement. *British Educational Research Journal*, 27, 407-432.
- Paulus, C. (2009). Die „Bücheraufgabe“ zur Bestimmung des kulturellen Kapitals bei Grundschulern. [http://bildungswissenschaften.uni-saarland.de/personal/paulus/Artikel/BA\\_Artikel.pdf](http://bildungswissenschaften.uni-saarland.de/personal/paulus/Artikel/BA_Artikel.pdf) [15.06.2013].
- Pekrun, R., Götz, Jullien, S., Zirngibl, A., v. Hofe, R., & Blum, W. (2003). *Skalenhandbuch PALMA: 2. Messzeitpunkt (6. Klassenstufe)*. Universität München: Institut für Pädagogische Psychologie.
- Pekrun, R., Götz, Jullien, S., Zirngibl, A., v. Hofe, R., & Blum, W. (2002). *Skalenhandbuch PALMA: 1. Messzeitpunkt (5. Klassenstufe)*. Universität München: Institut Pädagogische Psychologie.
- Pekrun, R., Hofe, R. vom, Blum, W., Götz, T., Wartha, S., Frenzel, A. & Jullien, S. (2006). Projekt zur Analyse der Leistungsentwicklung in Mathematik (PALMA). Entwicklungsverläufe, Schülervoraussetzungen und Kontextbedingungen von Mathematikleistungen in der Sekundarstufe I. In M. Prenzel & L. Allolio-Näckle (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule: Abschlussbericht des DFG-Schwerpunktprogramms* (S. 21-53). Münster: Waxmann.
- Pekrun, R. & Zirngibl, A. (2004). Schülermerkmale im Fach Mathematik. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003* (S. 191-210). Münster: Waxmann.
- Piaget, J. (1962). *The origins of intelligence in the child*. New York: Norton.
- Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.
- Piaget, J. & Inhelder, B. (1975). *The origin of the idea of chance in children*. London: Routledge & Kegan Paul.

- Ramirez, C., Schau, C. & Emmioglu, E. (2012). The importance of attitudes in statistics education. *Statistics Education Research Journal*, 11(2), 57-71. <http://www.stat.auckland.ac.nz/serj> [17.07.2014].
- Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical linear models (Second edition)*. Thousand Oaks: Sage Publications.
- Raudenbush, S.W., Bryk, A.S., Cheong, Y.F., Congdon, R. & Du Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling*. Lincolnwood IL: Scientific Software International.
- Reading, C. (2002). Profile for statistical understanding. In B. Philips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics: Developing a statistically literate society, Cape Town, South Africa*. Voorburg, the Netherlands: International Statistics Institute.
- Rost, J. (2004). *Lehrbuch Testtheorie/ Testkonstruktion*. Bern: Hans Huber.
- Schau, C. & Emmioglu, E. (2012). Do introductory statistics courses in the United States improve students' attitudes? *Statistics Education Research Journal*, 11(2), 86-94. <http://www.stat.auckland.ac.nz/serj> [17.07.2014].
- Schiefele, U. (1992). Topic interest and levels of text comprehension. In K. A. Renninger, S. Hidi & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 151-182). Hillsdale, NJ: Erlbaum.
- Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist*, 26, 299-323.
- Schild, M. (2011). Statistical literacy: A new mission for data producers. *Statistical Journal of the IAOS* 27 (2011), 173-183.
- Schild, M. (2010). Statistical literacy: A short introduction. <http://www.statlit.org/pdf/2010Schild-StatLit-Intro4p.pdf> [16.06.2013].
- Schild, M. (1998). Statistical literacy and evidential statistics. *Proceedings of the ASA Section on Statistical Education*, 187-192. [www.StatLit.org/pdf/1998SchildASA.pdf](http://www.StatLit.org/pdf/1998SchildASA.pdf) [11.07.2014].
- Schneider, W., Schlagmüller, M. & Ennemoser, M. (2007). *LGVT 6-12. Manual*. Göttingen: Hogrefe.

- Schukajlow, S., Leiss, D., Pekrun, R., Blum, W., Müller, M., & Messner, R. (2012). Teaching methods for modelling problems and students' task-specific enjoyment, value, interest and self-efficacy expectations. *Educational Studies in Mathematics* 79(2), 215-237.
- Schunk, D.H. & Zimmerman, B.J. (2003). Self-regulation and learning. In W.M. Reynolds & G.E. Miller (Eds.), *Handbook of Psychology: Educational psychology* (Vol. 7, pp. 59–78). New York: Wiley.
- Shaughnessy, J.M. (2007). Research on statistics learning and reasoning. *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 957-1009). Greenwich, CT: Information Age Publishing, Inc. and NCTM.
- Shavelson, R.J., Hubner, J.J., & Stanton, G.C. (1976). Validation of construct interpretations. *Review of Educational Research*, 46, 407-441.
- Sirin, S.R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75, 417-453.
- Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.
- Sproesser, U., Kuntze, S. & Engel, J. (2014). A multilevel perspective on factors influencing students' statistical literacy. *Proceedings of the 9th International Conference in Teaching Statistics ICOTS-9*, Flagstaff, USA. Voorburg, The Netherlands: International Statistical Institute. [http://icots.info/9/proceedings/pdfs/ICOTS9\\_7E2\\_SPROESSER.pdf](http://icots.info/9/proceedings/pdfs/ICOTS9_7E2_SPROESSER.pdf) [16.12.2015].
- Sproesser, U., Engel, J. & Kuntze, S. (angenommen). Fostering self-concept and interest for statistics through statistics-specific learning environments. *Statistics Education Research Journal*.
- Sproesser, U., Kuntze, S. & Engel, J. (eingereicht). Is statistical literacy interrelated to reading comprehension and cognitive abilities? A multilevel analysis.
- Stanat, P., Schwippert, K. & Gröhlich, C. (2010). Der Einfluss des Migrantenanteils in Schulklassen auf den Kompetenzerwerb. Längsschnittliche Überprüfung eines umstrittenen Effekts. In C. Allemann-Ghionda, C., P. Stanat, K. Göbel & C. Röhner (Hrsg.), *Migration, Identität, Sprache und Bildungserfolg* (S. 147-164). Weinheim: Beltz.

- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Allyn and Bacon.
- Tent, L. (2001). Zensuren. In D.H. Rost (Eds.), *Handwörterbuch Pädagogische Psychologie* (S. 805-811). Weinheim: Psychologie Verlags Union.
- Therneau, T., Atkinson, E. (1997). *An introduction into recursive partitioning using the RPART routines*. <http://www.mayo.edu/hsr/techrpt/61.pdf> [20.02.2012].
- Thorndike, R.L. & Hagen, E. (1971). *Cognitive Abilities Test (CAT). Examiner's manual*. Boston: Multi-Level Edition.
- Thrupp, M., Lauder, H. & Robinson, T. (2002). School composition and peer effects. *International Journal of Educational Research* 37(2002), 483-504.
- Valentine, J.C., DuBois, D.L. & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, 39(2), 111-133.
- Vogel, M. (2010) The significance of residuals for modelling data. *Proceedings of the Eighth International Conference on Teaching Statistics ICOTS8*, Ljubljana, Slovenia. Voorburg, The Netherlands: International Statistical Institute. [http://iase-web.org/documents/papers/icots8/ICOTS8\\_C259\\_VOGEL.pdf](http://iase-web.org/documents/papers/icots8/ICOTS8_C259_VOGEL.pdf) [16.12.2015].
- Wallman, K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, 88(421), 1-8.
- Watermann, R. & Baumert, J. (2006). Entwicklung eines Strukturmodells zum Zusammenhang zwischen sozialer Herkunft und fachlichen und überfachlichen Kompetenzen: Befunde national und international vergleichender Analysen. In J. Baumert, P. Stanat & R. Watermann (Eds.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit* (S. 61-94). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Watson, J.M. (2011). Foundations for improving statistical literacy. *Statistical Journal of the IAOS* 27, 27, 197-204.
- Watson, J. M. (1997). Assessing statistical literacy using the media. In I. Gal and J.B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 107-121). Amsterdam: IOS Press and The International Statistical Institute.

- Watson, J.M. & Callingham, R.A. (2003). Statistical literacy: a complex hierarchical construct. *Statistics Education Research Journal* 2, 3-46.
- Watson, J.M., & Kelly, B.A. (2004). Expectation versus variation: Students' decision making in a chance environment. *Canadian Journal of Science, Mathematics and Technology Education*, 4, 371-396.
- Watson, J.M., Kelly, B.A., Callingham, R.A. & Shaughnessy, J.M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology*, Vol. 34, 1-29.
- Watt, H.M.G. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th- through 11th-grade Australian students. *Child Development* 75(5), 1556-1574.
- Weinert, F. E. (2001): Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessung in Schulen* (S. 17-31). Weinheim: Beltz.
- Wild, C.J. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review* 67, 223-265.
- Willms, J. D. (1992). *Monitoring school performance*. London: Falmer Press.
- Wu, M. L., Adams, R. J., Wilson, M. R. & Haldane, S.A. (2007). *ACERConQuest Version 2: Generalised item response modelling software*. Camberwell: Australian Council for Educational Research.
- Ziegler, L.A. (2014). Reconceptualizing statistical literacy: developing an assessment for the modern introductory statistics course. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9<sup>th</sup> International Conference on Teaching Statistics*, Flagstaff, Arizona, USA. Voorburg, The Netherlands: International Statistical Institute. [http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9\\_7F1\\_ZIEGLER.pdf](http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_7F1_ZIEGLER.pdf) [16.12.2015].
- Zirngibl, A., Götz, T., Pekrun, R., vom Hofe, R., & Kleine, M. (2005). Aufgabenspezifische Erfassung von Mathematikemotionen. *Vortrag auf der Tagung der Sektion Empirische Bildungsforschung (AEPF-Tagung)*, Freie Universität Berlin, 17.03.2005.

## 11 Anhang

### 11.1 Lernumgebung „M&Ms“ des Treatments „Basis-Training“

Laufzettel BT
---------------

<b>Mein Code:</b>			
<b>Code meines Partners:</b>			

### Lernumgebung 1: M&Ms

	Erledigt? 	Bestätigt (Lehrer)
Ich habe diese Lernumgebung vollständig bearbeitet.		
Ich habe meine Lösungen regelmäßig mit der Musterlösung verglichen.		
Ich habe alle offenen Fragen geklärt. (Wenn nicht: schau noch einmal in die Hilfekarten und die Musterlösung oder frage deinen Lehrer.)		
<b>Abschlussfragen:</b> Der Begriff Spannweite bedeutet _____ _____. Um die durchschnittliche Anzahl der roten M&Ms in 20 Packungen zu berechnen, muss ich _____ _____. (Wenn du diese Fragen nicht beantworten kannst, lies noch einmal auf deinem Arbeitsblatt bzw. in der Musterlösung nach.)		
Insgesamt habe ich ca. _____min für diese Lernumgebung gebraucht.		
Ich fand die Aufgaben dieser Lernumgebung	Schwer <input type="checkbox"/>	Mittel <input type="checkbox"/>
	Einfach <input type="checkbox"/>	

Du bist fertig mit dieser Lernumgebung? Super!

Dann tausche beim Lehrer dein Arbeitsblatt gegen die Musterlösung und beginne mit der nächsten Lernumgebung.

	Erledigt? 	Bestätigt (Lehrer)
Ich habe mein Arbeitsblatt gegen die Musterlösung getauscht.		



Marie meint:

„Anscheinend sind in einer Packung immer 24 M&Ms. Wenn ich davon ausgehe, dass von jeder der 6 Farben gleich viele M&Ms hergestellt werden, kann ich mit 4 M&Ms jeder Farbe rechnen.“

1. Mache eine eigene Vorhersage, wie viele M&Ms von jeder Farbe in deiner Packung enthalten sein werden (**Tabellenspalte „Meine Vorhersage“**).
2. Öffne nun deine Packung und notiere in der Tabelle die Anzahlen der jeweiligen Farbe.
3. Nimm dir 19 Packungs-Kärtchen und vergleiche die angegebenen Häufigkeiten mit deiner Packung.
4. Addiere nun die Anzahlen der roten / orangen / gelben / ... M&Ms der **19 Kärtchen sowie deiner eigenen Packung** und notiere sie in der Tabelle.

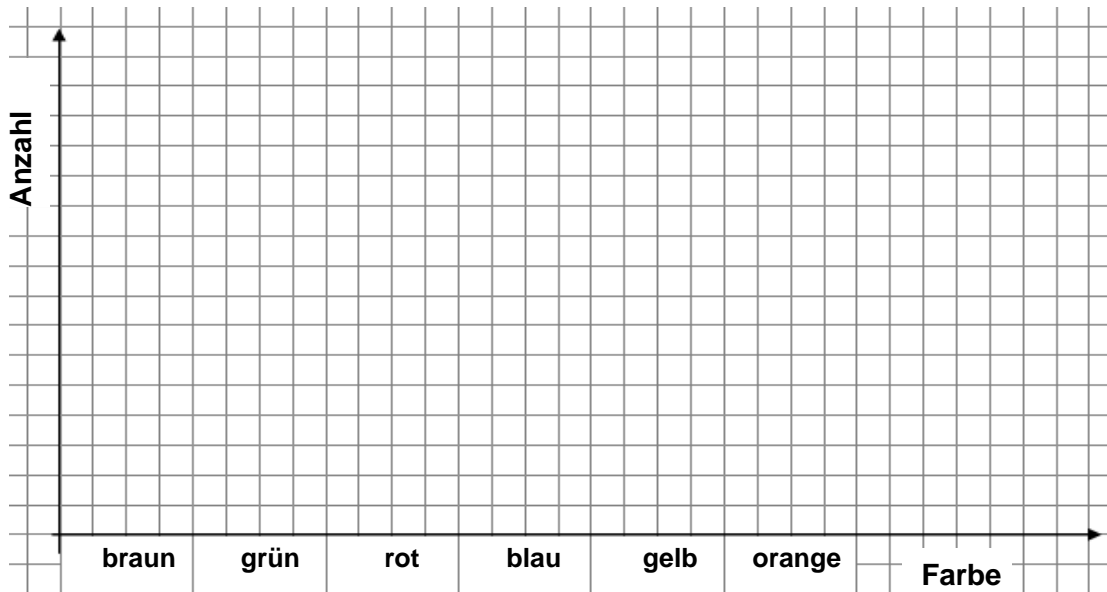
Platz für Rechnungen

5. Berechne für jede Farbe die durchschnittliche Anzahl in allen 20 Packungen und schreibe sie in die letzte Spalte der Tabelle.

Platz für Rechnungen

Farbe	Maries Vorhersage	Meine Vorhersage	Anzahl in 1 Packung	Gesamtanzahl in allen 20 Packungen	Durchschnitt in allen 20 Packungen
Braun	4				
Grün	4				
Rot	4				
Blau	4				
Gelb	4				
Orange	4				
Summe	24	24	24	480	-----

6. Stelle die Anzahlen deiner Packung in einem Säulendiagramm dar.  
Teile zunächst die y-Achse sinnvoll ein.



Spätestens jetzt solltest du **KONTROLLIEREN!**

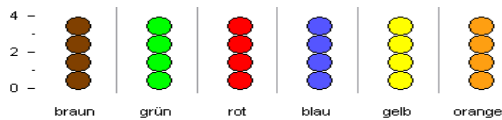
7. Einige Schüler aus Maries Klasse sollten weitere M&M-Packungen öffnen und die Anzahlen in einem Säulendiagramm darstellen.

Marie ist sich aber sicher:

„Ein paar von meinen Mitschülern haben geschummelt! Sie haben gar nicht wirklich gezählt, sondern sich die Anzahlen ausgedacht!“

Bei welchen Schaubildern wurde deiner Meinung nach geschummelt? Begründe!

a)



Echt  Geschummelt  
Weil \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

b)



Echt  Geschummelt  
Weil \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

c)



Echt  Geschummelt  
Weil \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_



Spätestens jetzt solltest du **KONTROLLIEREN!**

8. Fritz und Arno berechnen die Spannweiten der M&M-Anzahlen. Dazu ziehen sie von der häufigsten Anzahl (z. B. in Aufgabenteil 7b): 8 grüne M&Ms) die niedrigste Anzahl (z. B. in Aufgabenteil 7b): 1 braunes M&M) ab: → Spannweite in Aufgabenteil 7b):  $8 - 1 = 7$ . Die Spannweite gibt also an, wie stark sich die häufigste Anzahl von der niedrigsten Anzahl unterscheidet.

Bestimme die Spannweiten der M&M-Anzahlen der Aufgabenteile 7a) und 7c)!

→ Spannweite in Aufgabenteil 7a): \_\_\_\_\_

→ Spannweite in Aufgabenteil 7c): \_\_\_\_\_

- a) Nimm dir vom Materialtisch noch einmal 20 Packungs-Kärtchen und bestimme von ihnen jeweils die Spannweite.

→ Spannweite Packung 1: \_\_\_\_\_

→ Spannweite Packung 2: \_\_\_\_\_

→ Spannweite Packung 3: \_\_\_\_\_

→ Spannweite Packung 4: \_\_\_\_\_

→ Spannweite Packung 5: \_\_\_\_\_

→ Spannweite Packung 6: \_\_\_\_\_

→ Spannweite Packung 7: \_\_\_\_\_

→ Spannweite Packung 8: \_\_\_\_\_

→ Spannweite Packung 9: \_\_\_\_\_

→ Spannweite Packung 10: \_\_\_\_\_

→ Spannweite Packung 11: \_\_\_\_\_

→ Spannweite Packung 12: \_\_\_\_\_

→ Spannweite Packung 13: \_\_\_\_\_

→ Spannweite Packung 14: \_\_\_\_\_

→ Spannweite Packung 15: \_\_\_\_\_

→ Spannweite Packung 16: \_\_\_\_\_

→ Spannweite Packung 17: \_\_\_\_\_

→ Spannweite Packung 18: \_\_\_\_\_

→ Spannweite Packung 19: \_\_\_\_\_

→ Spannweite Packung 20: \_\_\_\_\_

- b) Wie hoch ist die größte, wie hoch ist die kleinste von dir berechnete Spannweite?

→ größte Spannweite: \_\_\_\_\_

→ kleinste Spannweite: \_\_\_\_\_



Spätestens jetzt solltest du **KONTROLLIEREN!**

## 11.2 Lernumgebung „M&Ms“ des Treatments „Reduktion“

Laufzettel Re
---------------

<b>Mein Code:</b>	<table border="1" style="display: inline-table; border-collapse: collapse;"> <tr> <td style="width: 30px; height: 20px;"></td> <td style="width: 30px; height: 20px;"></td> <td style="width: 30px; height: 20px;"></td> </tr> </table>			
<b>Code meines Partners:</b>	<table border="1" style="display: inline-table; border-collapse: collapse;"> <tr> <td style="width: 30px; height: 20px;"></td> <td style="width: 30px; height: 20px;"></td> <td style="width: 30px; height: 20px;"></td> </tr> </table>			

### Lernumgebung 1: M&Ms

	Erledigt? 	Bestätigt (Lehrer)
Ich habe diese Lernumgebung vollständig bearbeitet.		
Ich habe meine Lösungen regelmäßig mit der Musterlösung verglichen.		
Ich habe alle offenen Fragen geklärt. (Wenn nicht: schau noch einmal in die Hilfekarten und die Musterlösung oder frage deinen Lehrer.)		
<p><b>Abschlussfragen:</b></p> <p>Du hast in dieser Lernumgebung einen Nachteil des Bildens des Durchschnitts (Reduktion) erkannt, nämlich _____</p> <p>_____.</p> <p>Der Begriff Spannweite bedeutet _____</p> <p>_____.</p> <p>(Wenn du diese Fragen nicht beantworten kannst, lies noch einmal auf deinem Arbeitsblatt bzw. in der Musterlösung nach.)</p> <p>Insgesamt habe ich ca. _____min für diese Lernumgebung gebraucht.</p>		
Ich fand die Aufgaben dieser Lernumgebung	Schwer <input type="checkbox"/>	Mittel <input type="checkbox"/>
	Einfach <input type="checkbox"/>	

Du bist fertig mit dieser Lernumgebung? Super!

Dann tausche beim Lehrer dein Arbeitsblatt gegen die Musterlösung und beginne mit der nächsten Lernumgebung.

	Erledigt? 	Bestätigt (Lehrer)
Ich habe mein Arbeitsblatt gegen die Musterlösung getauscht.		

Marie meint:

„Anscheinend sind in einer Packung immer 24 M&Ms. Wenn ich davon ausgehe, dass von jeder der 6 Farben gleich viele M&Ms hergestellt werden, kann ich mit 4 M&Ms jeder Farbe rechnen.“

Marie hat ihre Vorhersage für die Farbverteilung nur aufgrund einer Information gemacht. Sie meint nämlich, dass von allen Farben gleich viele M&Ms hergestellt werden.

Erkläre, wie Marie von dieser Information zu ihrer Vorhersage gekommen ist.

---



---

1. Mache nun eine eigene Vorhersage, wie viele M&Ms von jeder Farbe in deiner Packung enthalten sein werden (**Tabellenspalte „Meine Vorhersage“**).
2. Öffne deine Packung und notiere in der Tabelle die Anzahlen der jeweiligen Farbe.
3. Nimm dir 9 Packungs-Kärtchen und vergleiche die angegebenen Häufigkeiten mit deiner Packung.
4. Addiere nun die Anzahlen der roten / orangen / gelben / ... M&Ms der **9 Kärtchen sowie deiner eigenen Packung** und notiere sie in der Tabelle.
5. Berechne für jede Farbe die durchschnittliche Anzahl in allen 10 Packungen und schreibe sie in die letzte Spalte der Tabelle.

Farbe	Maries Vorhersage	Meine Vorhersage	Anzahl in 1 Packung	Gesamtanzahl in allen 10 Packungen	Durchschnitt in allen 10 Packungen
Braun	4				
Grün	4				
Rot	4				
Blau	4				
Gelb	4				
Orange	4				
Summe	24	24	24	240	-----

Wenn man alle Anzahlen von M&Ms nebeneinander vor sich sieht, kann man leicht den Überblick verlieren. Darum macht es Sinn, die Gesamtanzahl und den Durchschnitt zu berechnen, denn damit erfasst man für jede M&M-Farbe einen Wert und kann besser vergleichen.

Vor allem der Durchschnitt ist ein wichtiger Wert, um sich einen Überblick zu verschaffen. Es gehen aber andererseits auch viele Informationen verloren, wenn man ihn berechnet.

- I. Welche Informationen gehen hier beim Bilden des Durchschnitts verloren?

---



---

- II. Welche Vorteile siehst du darin den Durchschnitt zu bilden?

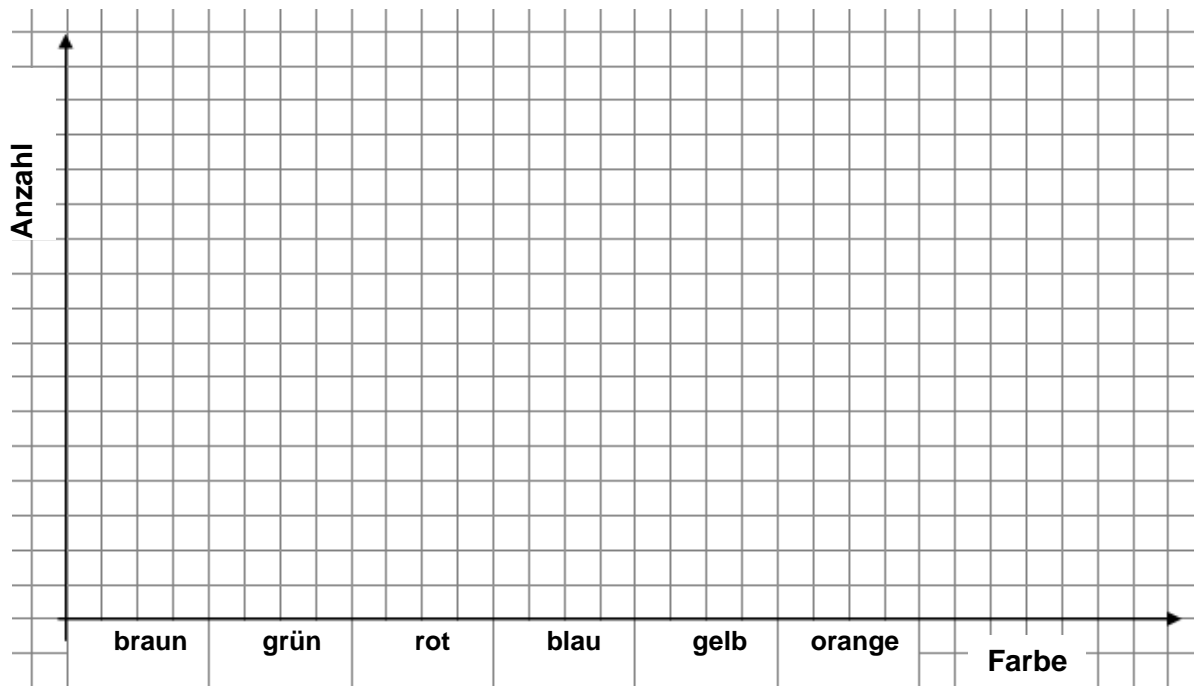
---

Auf Blatt 3 hast du Platz für Rechnungen



Spätestens jetzt solltest du **KONTROLLIEREN!**

6. Stelle die Anzahlen deiner Packung in einem Säulendiagramm dar.  
Teile zunächst die y-Achse sinnvoll ein.



7. Einige Schüler aus Maries Klasse sollten weitere M&M-Packungen öffnen und die Anzahlen in einem Säulendiagramm darstellen.  
Marie ist sich aber sicher:  
„Ein paar von meinen Mitschülern haben geschummelt! Sie haben gar nicht wirklich gezählt, sondern sich die Anzahlen ausgedacht!“

Bei welchen Schaubildern wurde deiner Meinung nach geschummelt? Begründe!

a)  Echt  Geschummelt

Weil \_\_\_\_\_

b)  Echt  Geschummelt

Weil \_\_\_\_\_

c)  Echt  Geschummelt

Weil \_\_\_\_\_



Spätestens jetzt solltest du **KONTROLLIEREN!**

8. Fritz und Arno berechnen die Spannweiten der M&M-Anzahlen. Dazu ziehen sie von der häufigsten Anzahl (z. B. in Aufgabenteil 7b): 8 grüne M&Ms) die niedrigste Anzahl (z. B. in Aufgabenteil 7b): 1 braunes M&M) ab:  $\rightarrow$  Spannweite in Aufgabenteil 7b):  $8 - 1 = 7$ . Die Spannweite gibt also an, wie stark sich die häufigste Anzahl von der niedrigsten Anzahl unterscheidet.

Bestimme die Spannweiten der M&M-Anzahlen der Aufgabenteile 7a) und 7c)!

$\rightarrow$  Spannweite in Aufgabenteil 7a): \_\_\_\_\_

$\rightarrow$  Spannweite in Aufgabenteil 7c): \_\_\_\_\_

- I. Du hast eine Tüte mit sehr großer Spannweite (z. B. Spannweite 10).

Was sagt dir diese Information über die jeweilige Packung?

---

---

- II. Du hast eine Tüte mit sehr kleiner Spannweite (z. B. Spannweite 0).

Was sagt dir diese Information über die jeweilige Packung?

---

---

- III. Welche Informationen gehen verloren, wenn du weißt, dass eine Packung ...

die Spannweite 10 hat? \_\_\_\_\_

---

---

die Spannweite 0 hat? \_\_\_\_\_

---

---

Platz für Rechnungen



Spätestens jetzt solltest du **KONTROLLIEREN!**

### 11.3 Lernumgebung „M&Ms“ des Treatments „Variabilität“

Laufzettel Va
---------------

<b>Mein Code:</b>	<table border="1" style="width: 100%; height: 20px;"> <tr> <td style="width: 33%;"></td> <td style="width: 33%;"></td> <td style="width: 33%;"></td> </tr> </table>			
<b>Code meines Partners:</b>	<table border="1" style="width: 100%; height: 20px;"> <tr> <td style="width: 33%;"></td> <td style="width: 33%;"></td> <td style="width: 33%;"></td> </tr> </table>			

## Lernumgebung 1: M&Ms

	Erledigt? 	Bestätigt (Lehrer)
Ich habe diese Lernumgebung vollständig bearbeitet.		
Ich habe meine Lösungen regelmäßig mit der Musterlösung verglichen.		
Ich habe alle offenen Fragen geklärt. (Wenn nicht: schau noch einmal in die Hilfekarten und die Musterlösung oder frage deinen Lehrer.)		
<b>Abschlussfragen:</b> In dieser Lernumgebung zeigt sich Variabilität darin, dass _____ _____. Der Begriff Spannweite bedeutet _____ _____.  (Wenn du diese Fragen nicht beantworten kannst, lies noch einmal auf deinem Arbeitsblatt bzw. in der Musterlösung nach.)		
Insgesamt habe ich ca. _____min für diese Lernumgebung gebraucht.		
Ich fand die Aufgaben dieser Lernumgebung	Schwer <input type="checkbox"/>	Mittel <input type="checkbox"/>
	Einfach <input type="checkbox"/>	

Du bist fertig mit dieser Lernumgebung? Super!

Dann tausche beim Lehrer dein Arbeitsblatt gegen die Musterlösung und beginne mit der nächsten Lernumgebung.

	Erledigt? 	Bestätigt (Lehrer)
Ich habe mein Arbeitsblatt gegen die Musterlösung getauscht.		



Marie meint:

„Anscheinend sind in einer Packung immer 24 M&Ms. Wenn ich davon ausgehe, dass von jeder der 6 Farben gleich viele M&Ms hergestellt werden, kann ich mit 4 M&Ms jeder Farbe rechnen.“

Stimmst du Marie zu? Begründe!

---



---

1. Mache nun eine eigene Vorhersage, wie viele M&Ms von jeder Farbe in deiner Packung enthalten sein werden (**Tabellenspalte „Meine Vorhersage“**).
2. Öffne deine Packung und notiere in der Tabelle die Anzahlen der jeweiligen Farbe.

I. Sind Maries Vorhersage, deine Vorhersage und die tatsächlichen Anzahlen identisch?

II. Wenn nein: Woran liegt das? Hast du bei der Vorhersage einen Fehler gemacht?

---



---

III. Wie könnte die Farbverteilung einer weiteren Packung aussehen?

Braun: \_\_\_\_\_, Grün: \_\_\_\_\_, Rot: \_\_\_\_\_, Blau: \_\_\_\_\_, Gelb: \_\_\_\_\_, Orange: \_\_\_\_\_

3. Nimm dir 9 Packungs-Kärtchen und vergleiche die angegebenen Häufigkeiten mit deiner Packung.
4. Addiere nun die Anzahlen der roten / orangen / gelben / ... M&Ms der **9 Kärtchen sowie deiner eigenen Packung** und notiere sie in der Tabelle.
5. Berechne für jede Farbe die durchschnittliche Anzahl in allen 10 Packungen und schreibe sie in die letzte Spalte der Tabelle.

Farbe	Maries Vorhersage	Meine Vorhersage	Anzahl in 1 Packung	Gesamtanzahl in allen 10 Packungen	Durchschnitt in allen 10 Packungen
Braun	4				
Grün	4				
Rot	4				
Blau	4				
Gelb	4				
Orange	4				
Summe	24	24	24	240	-----

Die von dir berechneten durchschnittlichen Anzahlen sind nicht alle gleich. Glaubst du, dass Maries Annahme („von allen Farben werden es gleich viele M&Ms hergestellt“) falsch ist?

---



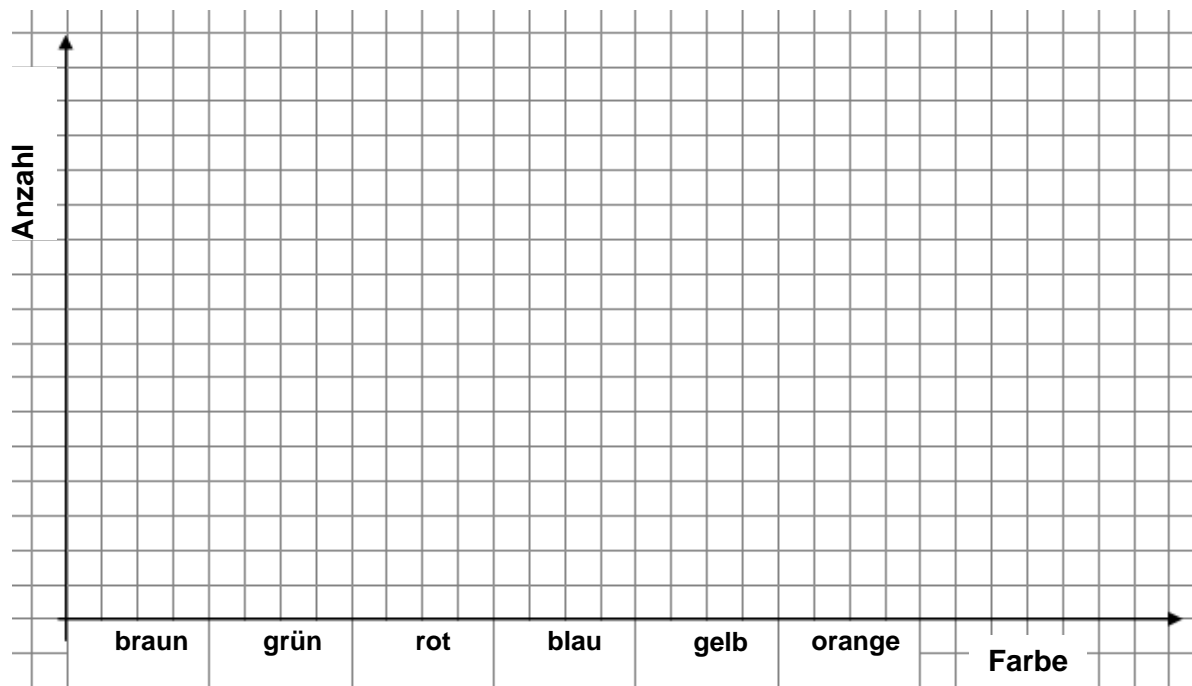
---

Auf Blatt 3 hast du Platz für Rechnungen



Spätestens jetzt solltest du **KONTROLLIEREN!**

6. Stelle die Anzahlen deiner Packung in einem Säulendiagramm dar.  
Teile zunächst die y-Achse sinnvoll ein.



7. Einige Schüler aus Maries Klasse sollten weitere M&M-Packungen öffnen und die Anzahlen in einem Säulendiagramm darstellen.  
Marie ist sich aber sicher:  
„Ein paar von meinen Mitschülern haben geschummelt! Sie haben gar nicht wirklich gezählt, sondern sich die Anzahlen ausgedacht!“

Bei welchen Schaubildern wurde deiner Meinung nach geschummelt? Begründe!

a)  Echt  Geschummelt

Weil \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

b)  Echt  Geschummelt

Weil \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

c)  Echt  Geschummelt

Weil \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

Woran machst du deine Vermutung fest, bei welchen Schaubildern geschummelt wurde?

---

---



Spätestens jetzt solltest du **KONTROLLIEREN!**

8. Fritz und Arno berechnen die Spannweiten der M&M-Anzahlen. Dazu ziehen sie von der häufigsten Anzahl (z. B. in Aufgabenteil 7b): 8 grüne M&Ms) die niedrigste Anzahl (z. B. in Aufgabenteil 7b): 1 braunes M&M) ab:  $\rightarrow$  Spannweite in Aufgabenteil 7b):  $8 - 1 = 7$ . Die Spannweite gibt also an, wie stark sich die häufigste Anzahl von der niedrigsten Anzahl unterscheidet. Bestimme die Spannweiten der M&M-Anzahlen der Aufgabenteile 7a) und 7c)!

$\rightarrow$  Spannweite in Aufgabenteil 7a): \_\_\_\_\_

$\rightarrow$  Spannweite in Aufgabenteil 7c): \_\_\_\_\_

Fritz und Arno haben zwei Packungen mit gleicher Spannweite. Sind die Inhalte ihrer Tüten völlig gleich? Begründe mit einem Beispiel!

---

---

---

---

Platz für Rechnungen



Spätestens jetzt solltest du **KONTROLLIEREN!**

## 11.4 Lernumgebung „M&Ms“ des Treatments „Reduktion und Variabilität“

Laufzettel ReVa
-----------------

<b>Mein Code:</b>			
<b>Code meines Partners:</b>			

### Lernumgebung 1: M&Ms

	Erledigt? 	Bestätigt (Lehrer)	
Ich habe diese Lernumgebung vollständig bearbeitet.			
Ich habe meine Lösungen regelmäßig mit der Musterlösung verglichen.			
Ich habe alle offenen Fragen geklärt. (Wenn nicht: schau noch einmal in die Hilfekarten und die Musterlösung oder frage deinen Lehrer.)			
<p><b>Abschlussfragen:</b></p> <p>In dieser Lernumgebung zeigt sich Variabilität darin, dass _____ _____.</p> <p>Du hast in dieser Lernumgebung einen Nachteil des Bildens des Durchschnitts (Reduktion) erkannt, nämlich _____ _____.</p> <p>(Wenn du diese Fragen nicht beantworten kannst, lies noch einmal auf deinem Arbeitsblatt bzw. in der Musterlösung nach.)</p> <p>Insgesamt habe ich ca. _____min für diese Lernumgebung gebraucht.</p>			
Ich fand die Aufgaben dieser Lernumgebung	Schwer <input type="checkbox"/>	Mittel <input type="checkbox"/>	Einfach <input type="checkbox"/>

Du bist fertig mit dieser Lernumgebung? Super!

Dann tausche beim Lehrer dein Arbeitsblatt gegen die Musterlösung und beginne mit der nächsten Lernumgebung.

	Erledigt? 	Bestätigt (Lehrer)
Ich habe mein Arbeitsblatt gegen die Musterlösung getauscht.		

Marie meint:

„Anscheinend sind in einer Packung immer 24 M&Ms. Wenn ich davon ausgehe, dass von jeder der 6 Farben gleich viele M&Ms hergestellt werden, kann ich mit 4 M&Ms jeder Farbe rechnen.“

Marie hat ihre Vorhersage für die Farbverteilung nur aufgrund der Annahme gemacht, dass von allen Farben gleich viele M&Ms hergestellt werden.

Erkläre, wie Marie von dieser Annahme zu ihrer Vorhersage gekommen ist.

---



---

Stimmst du Marie zu? Begründe! \_\_\_\_\_

---

1. Mache nun eine eigene Vorhersage, wie viele M&Ms von jeder Farbe in deiner Packung enthalten sein werden (**Tabellenspalte „Meine Vorhersage“**).
2. Öffne deine Packung und notiere in der Tabelle die Anzahlen der jeweiligen Farbe.

I. Sind Maries Vorhersage, deine Vorhersage und die tatsächlichen Anzahlen identisch?

II. Wenn nein: Woran liegt das? Hast du bei der Vorhersage einen Fehler gemacht?

---



---

III. Wie könnte die Farbverteilung einer weiteren Packung aussehen?

Braun: \_\_\_\_\_, Grün: \_\_\_\_\_, Rot: \_\_\_\_\_, Blau: \_\_\_\_\_, Gelb: \_\_\_\_\_, Orange: \_\_\_\_\_

3. Nimm dir 9 Packungs-Kärtchen und schau dir die angegebenen Häufigkeiten auch im Vergleich zu deiner Packung an.
4. Addiere nun die Anzahlen der roten / orangen / gelben / ... M&Ms **der 9 Kärtchen sowie deiner eigenen Packung** und notiere sie in der Tabelle.
5. Berechne für jede Farbe die durchschnittliche Anzahl in allen 10 Packungen und schreibe sie in die letzte Spalte der Tabelle.

Farbe	Maries Vorhersage	Meine Vorhersage	Anzahl in 1 Packung	Gesamtanzahl in allen 10 Packungen	Durchschnitt in allen 10 Packungen
Braun	4				
Grün	4				
Rot	4				
Blau	4				
Gelb	4				
Orange	4				
Summe	24	24	24	240	-----



Spätestens jetzt solltest du **KONTROLLIEREN!**

Die von dir berechneten durchschnittlichen Anzahlen sind nicht alle gleich. Glaubst du, dass Maries Annahme („von allen Farben werden gleich viele M&Ms hergestellt“) falsch ist?

---



---



---

Wenn man alle Anzahlen von M&Ms nebeneinander vor sich sieht, kann man leicht den Überblick verlieren. Darum macht es Sinn, die Gesamtanzahl und den Durchschnitt zu berechnen, denn damit erfasst man für jede M&M-Farbe einen Wert und kann besser vergleichen.

Vor allem der Durchschnitt ist ein wichtiger Wert, um sich einen Überblick zu verschaffen. Es gehen aber andererseits auch viele Informationen verloren, wenn man ihn berechnet.

I. Welche Informationen gehen hier beim Bilden des Durchschnitts verloren?

---



---



---

II. Welche Vorteile siehst du darin den Durchschnitt zu bilden?

---



---



---

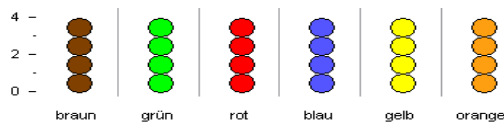
6. Einige Schüler aus Maries Klasse sollten weitere M&M-Packungen öffnen und die Anzahlen in einem Säulendiagramm darstellen.

Marie ist sich aber sicher:

„Ein paar von meinen Mitschülern haben geschummelt! Sie haben gar nicht wirklich gezählt, sondern sich die Anzahlen ausgedacht!“

Bei welchen Schaubildern wurde deiner Meinung nach geschummelt? Begründe!

a)



Echt  Geschummelt

Weil \_\_\_\_\_

---



---

b)



Echt  Geschummelt

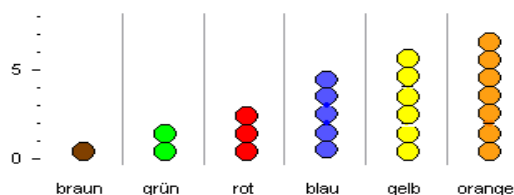
Weil \_\_\_\_\_

---



---

c)



Echt  Geschummelt

Weil \_\_\_\_\_

---



---

Woran machst du deine Vermutung fest, bei welchen Schaubildern geschummelt wurde?

---

---



Spätestens jetzt solltest du **KONTROLLIEREN!**

7. Fritz und Arno berechnen die Spannweiten der M&M-Anzahlen. Dazu ziehen sie von der häufigsten Anzahl (z. B. in Aufgabenteil 6b): 8 grüne M&Ms) die niedrigste Anzahl (z. B. in Aufgabenteil 6b): 1 braunes M&M) ab:  $\rightarrow$  Spannweite in Aufgabenteil 6b):  $8 - 1 = 7$ . Die Spannweite gibt also an, wie stark sich die häufigste Anzahl von der niedrigsten Anzahl unterscheidet.

Bestimme die Spannweiten der M&M-Anzahlen der Aufgabenteile 6a) und 6c)!

$\rightarrow$  Spannweite in Aufgabenteil 6a): \_\_\_\_\_

$\rightarrow$  Spannweite in Aufgabenteil 6c): \_\_\_\_\_

- I. Du hast eine Tüte mit sehr großer Spannweite (z. B. Spannweite 10).

Was sagt dir diese Information über die jeweilige Packung?

---

---

- II. Du hast eine Tüte mit sehr kleiner Spannweite (z. B. Spannweite 0).

Was sagt dir diese Information über die jeweilige Packung?

---

---

- III. Welche Informationen gehen verloren, wenn du weißt, dass eine Packung ...

die Spannweite 10 hat? \_\_\_\_\_

---

---

die Spannweite 0 hat? \_\_\_\_\_

---

---

- IV. Fritz und Arno haben zwei Packungen mit gleicher Spannweite. Sind die Inhalte ihrer Tüten völlig gleich? Begründe mit einem Beispiel! \_\_\_\_\_

---

---



Spätestens jetzt solltest du **KONTROLLIEREN!**