FOSTERING SELF-CONCEPT AND INTEREST FOR STATISTICS THROUGH SPECIFIC LEARNING ENVIRONMENTS

UTE SPROESSER
Ludwigsburg, University of Education
sproesser@ph-ludwigsburg.de

JOACHIM ENGEL
Ludwigsburg, University of Education
engel@ph-ludwigsburg.de

SEBASTIAN KUNTZE
Ludwigsburg, University of Education
kuntze@ph-ludwigsburg.de

ABSTRACT

Supporting motivational variables such as self-concept or interest is an important goal of schooling as they relate to learning and achievement. In this study, we investigated whether specific interest and self-concept related to the domains of statistics and mathematics can be fostered through a four-lesson intervention focusing on statistics. Data about these motivational variables and achievement related to statistics were gathered from 503 eighth graders. Our results indicate that students perceived mathematics and statistics differently with 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 achievement in the domain of statistics.

Keywords: *Intervention study; Motivational variables; Confirmatory factor analysis; Statistics education research*

1. 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 these variables 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) emphasized 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 their impact on their educational choices. However, within statistics education the awareness related to affective and motivational variables still appears as rather low: Compared to mathematics education only a few studies have investigated these variables related to this domain (e.g., Bond, Perkins & Ramirez, 2012; Ramirez, Schau, & Emmioğlu, 2012). Although Hood, Creed,

Statistics Education Research Journal, 15(1), 28-54, http://iase-web.org/Publications.php?p=SERJ

and Neumann (2012) emphasized the necessity to foster such motivational variables in particular for learners with rather low achievement in 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.

Therefore, this study examines whether self-concept and interest both in mathematics and in statistics can be supported by a statistics-oriented and learner-centered intervention among eighth graders. The intervention implemented characteristics that are considered to have a positive impact on self-concept (such as provision of feedback, see, e.g., O'Mara, Marsh, Craven, & Debus, 2006) and interest (such as experience of autonomy, achievement and social relatedness, see, e.g., Krapp, 2005). As students often associate statistics with mathematics (Gal et al., 1997), we first investigated whether middle school students perceived these two domains as separate with respect to their interest and self-concept. Then we analyzed whether self-concept and interest increased during the course of the domain-specific intervention. In order to explore which characteristics of learners particularly favor a positive development of these motivational variables, we additionally assessed if previous statistics-related achievement was related to the development of self-concept and interest in the corresponding domain. The theoretical background leading to these research interests is presented next.

2. THEORETICAL BACKGROUND

Motivational variables such as self-concept and interest accompany the learning process and influence learners' achievement development at school and beyond (cf. Eccles & Wigfield, 2002; Helmke & Weinert, 1997; Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002). For instance, interest or confidence in one's own competencies determine if learning activities are performed even beyond external obligations. In this regard, it is not surprising that there are various reciprocal effects between achievement and motivational variables and that—besides fostering achievement—the development of students' motivational variables is considered to be an essential goal of schooling (e.g., Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005; 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[2] or TIMSS[3] (e.g., Artelt, Baumert, Julius-McElvany, & 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.

2.1. SELF-CONCEPT

According to Bong and Skaalvik (2003, p. 6) "academic self-concept refers to individuals' knowledge and perceptions about themselves in achievement situations" and it typically relates to particular academic domains such as school subjects. Hence, in the case of mathematics, academic self-concept expresses a person's confidence in his or her competencies in this particular domain. Moreover, self-concept is not only a domain-specific construct but also multidimensional and hierarchical, that is, self-concepts related to more specific domains are subordinated to and included into more general self-concepts (Shavelson, Hubner, & Stanton, 1976). In general, self-concept represents relatively stable

[2] Programme for International Student Assessment.
[3] Third International Mathematics and Science Study.

perceptions of the self, based on prior (achievement) experiences (Bong & Skaalvik). Crucial factors that influence self-concept in a particular domain are social comparisons with others, comparisons with former achievement, and comparisons between perceived achievement in different domains (Möller & Köller, 2004).

Another relevant construct within the field of self-competency beliefs is self-efficacy as the conviction to master given tasks (e.g., Bandura, 1982). As self-efficacy refers to particular tasks, it depends more on the actual context and hence is less stable than self-concept (Bong & Skaalvik, 2003). According to Bandura (1982), self-efficacy is mainly formed by achievement experience, comparisons with others, and verbal encouragement. Hence, achievement experiences as well as social comparisons are crucial for both self-efficacy and self-concept. Bong and Skaalvik even consider self-efficacy as a precursor for self-concept. Thus, in this study, we primarily focus on academic self-concept for mathematics and statistics.

Self-competency beliefs such as academic self-concepts are considered to be positively related to motivation and achievement (e.g., 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. Hansford and Hattie (1982), for instance, report findings of a meta-analysis indicating that domain-specific self-concept correlates at an average of r = 0.42 with related achievement compared to $r \leq 0.22$ for global self-concept. This is why Marsh and Craven (1997) advise researchers to focus on self-concept related to the domains of their particular research.

The relationship between (academic) self-concept and achievement is considered to be reciprocal, that is, achievement has an impact on self-concept and vice versa. High levels of achievement apparently lead to high levels of self-concept because of positive achievement experiences and favorable social comparisons. On the other hand, Valentine, DuBois, and Cooper (2004) mention several reasons for the influence of self-concept on 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; they choose more challenging tasks and develop better ways of coping with failure leading to more persistence and hence better school performance. In contrast, individuals with a lower positive academic self-concept are likely to develop avoidance strategies. Reciprocal effects models between self-concept and achievement could also be confirmed empirically (e.g., Marsh et al., 2005; Valentine et al., 2004).

Several studies have aimed at fostering academic self-concept. O'Mara and colleagues (2006) found in an extensive meta-analysis that self-concept interventions are particularly effective when they refer to the same domain the measured self-concept relates to. Moreover, interventions focusing particularly on praise and feedback but also on developing domain-specific skills have shown to be successful in fostering academic self-concepts. In line with Marsh et al. (2005), O'Mara et al. (2006) point out that the reciprocal effects between self-concept and achievement suggest fostering both variables simultaneously. In this case, the increase remained relatively stable over time.

Hattie and Timperley (2007) explain the impact of feedback on self-competency beliefs such as self-concept by claiming that "the main purpose of feedback is to reduce discrepancies between current understandings and performance and a goal" (p. 86). Hence, effective feedback gives information about the learning goals, the actual performance, and how to improve the learning process. Furthermore, they point out that feedback can considerably influence self-competency beliefs when it focuses on how to better perform on a given task. This means that, on the one hand, the students' achievement may be enhanced through their teachers' feedback, which may lead to positive temporal comparisons and hence to a more positive self-concept. On the other hand, feedback may provide a more realistic perception of their own achievement.

Beyond its relatedness to achievement, self-concept is also considered to be related to interest (e.g. Pekrun & Zirngibl, 2004). In addition to academic self-concept, this study investigated the development of interest over the course of the intervention. Hence, the next section outlines the theoretical background related to interest.

2.2. INTEREST

According to Krapp's (2007, p. 8) person-object theory of interest (POI), "interest represents or describes a more or less enduring specific relationship between a person and an object." Krapp further points out that interest always refers to particular contents or objects; hence, interest is a domain-specific variable. In general, interest is accompanied by a high readiness to acquire new knowledge and skills related to the object of interest as well as by the desire to apply such knowledge and skills.

In line with Hidi and Renninger (2006), Krapp claims that interest develops over time, for instance, through favorable learning environments. In this context, both authors distinguish situational from individual interest: The first step of interest development is characterized by focused attention, for example, due to surprising information or personal relevance of an object. Under certain conditions (e.g., subjective meaningfulness), this first state of triggered situational interest may develop to a more stabilized situational interest. Again, particular conditions may support the development from a maintained situational interest to an individual interest for a particular object such as a school subject. Contrary to the rather fluctuating situational interest that may depend on factors such as the topic, the context, and the kind of activity (Gardner, 1985), individual interest is relatively stable over time.

Two aspects are crucial for the development of interest: First, individuals attribute a personal importance to their objects of interest. Beyond these value-related valences, interest is typically accompanied by involvement and enjoyment (feeling-related valences). For this reason, interest can be considered as a trigger for intrinsic motivation. Interest will develop only if value- and feelings-related valences are experienced in a positive way (Krapp, 2005). Referring to self-determination theory (SDT; Deci & Ryan, 2002), such positive emotional experiences require the fulfillment of the basic needs of achievement experience, autonomy, and social relatedness (Krapp, 2005). In the context of school, this means that learners firstly desire to perceive themselves as being able to attain a desired outcome. Secondly, students strive to feel relatively independent from external obligations while working on a task; this second need is closely intertwined with the need for achievement experience because only when learners feel competent of mastering a task can they exert it autonomously; otherwise, when they work with little external support, they perceive themselves as competent. Finally, experiencing collaborative work in learning situations may support students' interest.

Several studies found that interest and achievement are correlated (see, e.g., the meta-analysis of Helmke & Weinert, 1997, with a mean correlation between interest and school achievement of $r = 0.41$). According to SDT or POI, (perceived) achievement is considered to have an impact on interest. On the other hand, Schiefele (1991) gives several reasons for the influence of interest on achievement: Highly interested students are more likely to engage in intense and meaning-oriented learning activities, whereby they use more elaborated learning strategies and invest more time and effort on learning. Empirical evidence for a reciprocal effects model between interest and achievement can be found, for instance, by Köller, Baumert and Schnabel (2001) or by Marsh et al. (2005).

The relationship between interest and achievement underlines the importance of paying attention to and fostering students' interest through specific learning activities. According to the model of interest development by Hidi and Renninger (2006), interest for a specific topic may grow within classrooms while dealing with the topic. In line with the fulfillment of the three basic needs (Deci & Ryan, 2002; Krapp, 2005), Hidi and Renninger suggest that teachers support their students' interest, for instance, by providing a task choice and accounting for own questions to promote a sense of autonomy, by giving helpful feedback and challenging, personally relevant tasks to build up their achievement as well as by enabling social learning with a partner or in a group.

In the next section, we will outline the theoretical background related to interest and self-concept in the specific domains of mathematics and statistics.

2.3. SELF-CONCEPT AND INTEREST: 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 et al. (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 expect that, in general, students of secondary school can hold different perceptions of mathematics and statistics from a motivational point of view (cf. Gundlach, Kuntze, Engel, & Martignon, 2010a, 2010b), even if this expectation has still to be verified empirically. 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 check whether this assumption is reflected by the empirical data. In a first analysis step, we will investigate by confirmatory factor analysis whether academic self-concept and interest 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 ($r = 0.55$). Klieme, Neubrand, and Lüdtke (2001) report a correlation of $r = 0.70$ within the PISA sample. Baumert and Köller (2000) found high correlations between achievement and self-concept ($r = 0.73$) as well as between achievement and interest ($r = 0.69$) in the TIMSS sample. A study by Marsh, Trautwein, Lüdtke, Köller and Baumert (2006) replicates this relationship between mathematics-related self-concept and achievement ($0.51 \leq r \leq 0.77$) whereas correlation coefficients near zero ($-0.03 \leq r \leq 0.05$) were found between general or non-academic self-concept and the achievement score. Marsh et al. (2005) empirically confirmed the general assumption of reciprocal effects between self-concept or interest on the one hand and achievement on the other hand for the domain of mathematics. In both cases, the impact of the motivational variable on achievement turned out to be stronger than the reciprocal impact. Moreover, effects appeared to be stronger for self-concept than for interest. Even if self-concept and (individual) interest are relatively stable over time (e.g., Bong & Skaalvik, 2003; Krapp, 2007), several studies (cf. Frenzel, Goetz, Pekrun, & Watt, 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; Gal et al., 1997; Hood et al., 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 relation to statistics. Emmioğlu and Çapa-Aydın (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 Emmioğlu (2012) indicates 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 observed a decline with increasing age for statistics-related interest. Furthermore, they found significant correlations between self-efficacy and interest referring to statistics (r = 0.63). Moreover, prior achievement in mathematics was positively related to self-efficacy in particular, but also to interest in the domain of statistics.

Even if the mentioned studies examined students' motivational variables related to statistics, they do not provide empirical evidence about ways of supporting domain-specific self-concept and interest. Furthermore, there is no insight into whether learners' prior statistics-related achievement is connected to the development of these motivational variables over time. As the study presented here focused on fostering students' domain-specific self-concept and interest through a statistics-related intervention, we will describe this intervention in the next section.

2.4. THE STATISTICS-RELATED INTERVENTION

According to the German standards referring to mathematics instruction (KMK, 2003) which includes the *leading idea* of *data and chance*, students ought to learn about data collection, data representation, data reduction, data interpretation, and probability (comparable to 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, tasks referring to statistics may encompass more. One important reason why statistics is widely needed and used is the lack of uniformity among all observable objects (e.g., Wild & Pfannkuch, 1999). Dealing adequately with such statistical variation thus constitutes an important goal of statistics education (Watson & Callingham, 2003). It implies judging within a concrete situation whether observed differences can be attributed to randomness or to an interpretable effect that is meaningful for the corresponding context (Wild & Pfannkuch). Complementary to considerations of variation is the reduction of data to a numerical or graphical summary. Such data reduction omits parts of the original data that often contain a large magnitude of relevant and irrelevant information so that reduced data mostly provide better overviews. Abilities in the domain of data reduction are not restricted to determining characteristic values but they include an awareness about the associated loss of information, for example, by smoothing out random variation through forming averages (Kröpfl, Peschek & Schneider, 2000).

Corresponding to the learning goals outlined above, the intervention of the study presented here focused on the one hand on (basic) elements of statistics that students already know from their mathematics classroom. On the other hand, we included tasks requiring dealing with variation and/or data reduction, as described above central aspects of statistics. Students were assigned to one of four treatments in a 2×2 design. Treatment 1 involved exclusively basic elements of statistics such as dealing with tables, bar charts and characteristic values. On the basis of these tasks, treatment 2 required the reflection of statistical variation, whereas treatment 3 requested to reflect about data reduction. Treatment 4 combined both requirements (see the Appendix for sample tasks). These treatments were designed to have an impact on students' achievement in the domain of statistics (see Sproesser, Engel, & Kuntze, 2015a) and also to support self-concept and interest (see Sproesser, Engel, & Kuntze, 2015b, 2015c). In particular, it is possible that the four special foci of the learning environments may cause differences in developing students' domain-specific self-concept and interest, for instance, in the sense that students may perceive tasks dealing with variation as more interesting or more difficult than basic tasks. Therefore, beyond the influence of the statistics-related intervention as a whole, the potential impact of the four different treatments on domain-specific self-concept and interest was explored in this study.

Apart from these differences, as far as reflective tasks are concerned, the treatments used the same contents and were implemented identically. As suggested by Scheaffer, Gnanadesikan, Watkins, and Witmer (1996), statistical contents provide opportunities for implementing student-centered activities in corresponding learning environments. In this sense, all four treatments provided learner-centered and hands-on activities in contexts relevant for eighth graders. For instance, students could discover characteristics of the distribution of chocolate lentils (see Appendix for sample tasks and an overview of the contexts and statistical concepts of the intervention). During the whole intervention, students worked with a partner and were relatively autonomous from the teacher on statistics-related problems. At any time, students had the possibility of getting help from flash cards and they could verify their answers with sample solutions. Upon finishing a topic, the students were given additional individual feedback as to whether the actual learning goal was reached and—if necessary—from which flash card to get help. The intervention was implemented so as to make students aware of their own achievement development through feedback (e.g., Hattie & Timperley, 2007) and hence to support their domain-specific self-concept. Furthermore, the intervention should foster interest by providing autonomy in cooperative learning situations (see e.g. Krapp, 2005). Therefore, we expected motivational variables to be supported through working on specific tasks and not through an explicit motivational training like, for instance, an explicit emphasis on the general contribution of statistics for life on a meta-level. Of course, during their work students could experience that statistical contents helped them to solve problems, but the students were not explicitly told to appreciate statistics for general reasons.

As the learning material was student-centered, the class teacher's role was mainly limited to observing students in their working process. The first author assured that students had access to sample solutions and flash cards, and she gave individual feedback.

2.5. RESEARCH QUESTIONS

The previous Sections 2.1 to 2.3 emphasized the 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. Hence, this is the main research interest of the present study.

As Gal et al. (1997) recommend researchers to discriminate between students' perceptions of the domains of statistics and mathematics before investigating motivational variables related to statistics, we first aim to explore whether students perceive statistics and mathematics differently from a motivational point of view. Next we examine the efficacy of our intervention with respect to supporting mathematics- and statistics-specific self-concept and interest. As the design of the study provided four treatment groups, also the potential impact of these different treatments is analyzed.

We focus on 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?

Moreover, it is possible that the intervention does not have the same effect on learners that differ in particular characteristics, such as prior domain-specific achievement. As argued above, mathematics-specific achievement has an influence on self-concept and interest in this domain (e.g., Marsh et al., 2005). Consequently, students' prior statistics-related achievement might influence the extent to which these motivational variables develop during the intervention. In this sense, the third research question is:

3) Does prior statistics-related achievement impact the development of students' statistics-related self-concept and interest?

3. METHODS

3.1. DESIGN AND SAMPLE

Data from 450 German eighth graders (212 female, 238 male), who completed pre- and post-tests and participated in the intervention were the basis of the analysis. These 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 public secondary schools (so-called "Realschule"). In addition to the students participating in the intervention, a baseline group consisting of 53 students (32 male, 21 female) between the ages of 13 and 15 years (M 13.73; SD 0.72) from four other classes was included in the study. The students of this baseline group continued with their regular mathematics classes instead of the intervention. They completed, in identical time intervals (approximately 2 weeks between pre- and posttest), the same tests as the students in the intervention group. In each class, the particular class teacher and the first author were present for the testing and during the intervention. We coordinated with the teachers of the baseline group that the students in this group did not work on statistical topics during this time. A control by an external observer of the classroom would have been a very unusual circumstance; we therefore trusted in the teachers' cooperation.

Analyzing the pretest scores of the variables included in this study revealed no significant differences between baseline and intervention group: (Latent) regressions in Mplus 7.1 specifying each pretest variable as outcome variable and the dummy-coded group assignment as independent variables resulted in *p*-values ranging from 0.064 to 0.852. Selection of participating schools was done to have a mix of inner city and rural schools in the federal state of Baden-Württemberg. Inclusion in the study required parental consent, which resulted in an overall 5.5% dropout rate.

We assert that prior to this study participating students had been instructed with some elements of descriptive statistics according to the German Standards for mathematics and the State Standards (KMK, 2003; Land Baden-Württemberg, 2004) such as numerical data summaries (e.g., average, median), reading and interpreting quantitative information (given) in tables, charts and diagrams. In other school subjects such as geography, history or science, students are typically also faced with representations of statistical data such as tables or diagrams; yet, the German Standards do not explicitly aim at teaching statistics in these subjects. Hence, students in our study ought to know the term statistics and associate it with the above-mentioned contents and activities. Against this background we expect that students will typically perceive statistics as a part of the mathematics classroom that is also relevant for other school subjects, in different social contexts or for their private lives (e.g., statistics in sports or data representations in the media). However, discussing the value of statistics for particular domains or everyday live is not explicitly prescribed in the German Standards for any school subject.

Over a period of four 45-minute classroom lessons within one week, the students participating in the intervention worked on learner-centered material dealing with different statistics-related topics. Curriculum validity of this intervention for grade 8 was checked on the basis of corresponding mathematics standards and textbooks (see, e.g., Aits, Berkemeier, Hecht, Heske, Koullen, & Ostrow, 2006; Griesel, Postel, & vom Hofe, 2005; Maroska, Olpp, Pongs, Stöckle, Wellstein, & Wontroba, 2006). In order to avoid bias through initial differences between the classes, students in each class were split up into four subgroups according to a 2×2 design. Assignment to the four different subgroups followed a randomized block design with pretest statistics-related achievement score as blocking factor: As mentioned above, students in group 1 (basic training, $n = 117$) 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 specific tasks that required them to reflect about statistical variation (treatment 2, $n = 120$) or data reduction (treatment 3, $n = 102$) or both variation and reduction (treatment 4, $n = 111$) (see Section 2.4 for further details).

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

3.2. MEASURES

Motivational dispositions referring to mathematics and statistics were measured in a multiple-choice format using a 5-point Likert scale (1 = strongly disagree, …, 5 = strongly agree). 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, vom Hofe, Blum, Frenzel, & Wartha (2007; see also Frenzel, Pekrun, Dicke, & Goetz, 2012). 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 et al. (2010a, 2010b) as parallel scales referring to statistics. In that prior study, it was shown by factor analysis that the statistics-related scales were empirically separable from their mathematics-related counterparts (see Table 2 in Section 4 for sample items or Appendix for all 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 constructs within our sample. These mathematics- and statistics-related scales have already been used in several studies (see, e.g., Frenzel et al. 2012; Gundlach et al., 2010a, 2010b; Pekrun et al., 2007; Schukajlow & Krug, 2014) in German secondary schools with similar age groups as included in this study. Hence, in the context of our study they appeared to be more appropriate than scales such as SATS (Survey of Attitudes Towards Statistics; see, e.g., Schau, Stevens, Dauphinee, & Del Vecchio, 1995) that has not been used with students of grade 8 yet. In order to provide the opportunity of comparing our results to studies that used other scales, the scales' items are displayed in the Appendix.

Additionally, as an alternative measure for students' self-concept 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 the Appendix for sample items) were presented to the learners. These tasks constitute typical activities that students may encounter in their statistics classroom (see, e.g., the textbooks from Aits et al., 2006; Griesel et al., 2005; Maroska et al., 2006). Comparisons between students' self-concept and interest concerning these tasks (summarized to a scale *task-specific self-concept* and a scale *task-specific interest*) and the non-task-specific statistics-related scales were made to assess whether students associated—from the point of view of self-concept and interest—the word statistics with dealing with statistical contents such as the ones presented in the tasks.

Furthermore, in pre- and posttest we gathered data about students' achievement in the domain of statistics by a test instrument that comprises 15 tasks. This achievement test was conceived and tested in samples from primary to tertiary education by Kuntze, Engel, Martignon, and Gundlach (2010) and Kuntze, Lindmeier, and Reiss (2008) (for sample items, see Lindmeier, Kuntze, & Reiss, 2007). In this preceding research, the instrument has shown to fit to a one-dimensional Rasch-model. In Sproesser, Kuntze, & Engel (submitted), more information about the achievement test instrument and details about the parameters can be found.

3.3. ANALYSIS

To investigate research question 1, confirmatory factor analysis of each motivational scale was conducted with the software AMOS 22.0 using maximum likelihood solutions. Checking the distributions of the corresponding items for normality revealed that skewness of six items in pre- and post-test exceeded the critical value 3. Logarithmic transformations were applied to these items to get their distribution closer to normal. After transformation, skewness and kurtosis for all items were in an acceptable range (skewness < 3; kurtosis < 8, see Kline, 2005). Furthermore, the items did not correlate above 0.85; hence, there was no severe collinearity that could cause estimation problems (Bühner, 2011). For the 503 cases, missing values for the items were below 5%. Before conducting confirmatory factor analysis, these missing data was imputed with the Expectation Maximization algorithm of SPSS 22.0. For each factor, the regression weight of one item was fixed to 1.0. As measures of global model fit, the comparative fit index (CFI) and the root-mean-square error of approximation (RMSEA) were assessed (Hu & Bentler, 1999, suggest CFI $\geq$ 0.95 and RMSEA $\leq$ 0.06 for a sufficient fit). As the χ^2 statistic depends on the sample size and hence for large sample sizes often significant differences between model and data are indicated, the χ^2/df ratio was additionally calculated (according to Bollen & Long, 1993, the model fits the data reasonably sufficiently when this ratio does not exceed 5). In order to compare the fit of nested models, chi-square difference tests were calculated (see Kline, 2005, p. 182, for further information).

Research questions 2 and 3 were examined using the software Mplus 7.1. We used the original dataset without imputations and transformations because the implemented Full Information Maximum Likelihood (FIML) procedure estimates missing values and the robust maximum likelihood estimator corrects for non-normality in the measures.

As the participants in our sample were clustered in school classes, cases could not be assumed to be independent. Disregarding this dependency structure might lead to inaccurate parameter estimates because of deflated standard errors (Snijders & Bosker, 1999). For the motivational scales, intraclass correlations were rather low (< 0.0050), indicating no need for multilevel analysis (Snijders & Bosker). 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 whether domain-specific self-concept and interest significantly increased due to the intervention (research question 2), we specified latent regression models predicting self-concept and interest, respectively (indicated by the actual items) in the posttest by the corresponding pretest values as well as by the group assignment (see Figure 1 for the structure of a regression model predicting self-concept). Such latent regressions also model the measurement error and hence provide more accurate estimates (Kline, 2005).

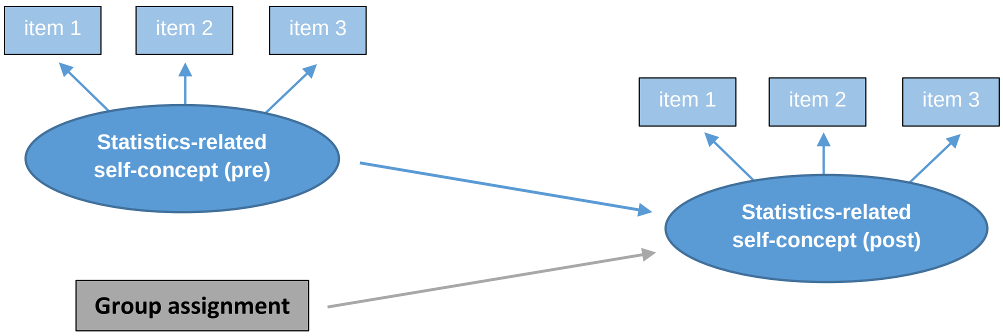


Figure 1. Example of a latent regression model (research question 2)

This procedure allowed us to examine differences in the development of self-concept and interest, respectively, between the different groups. Including the pretest score in the regression model controls for original differences in this variable. Here, the baseline group was used as a safeguard against effects that were not due to the intervention.

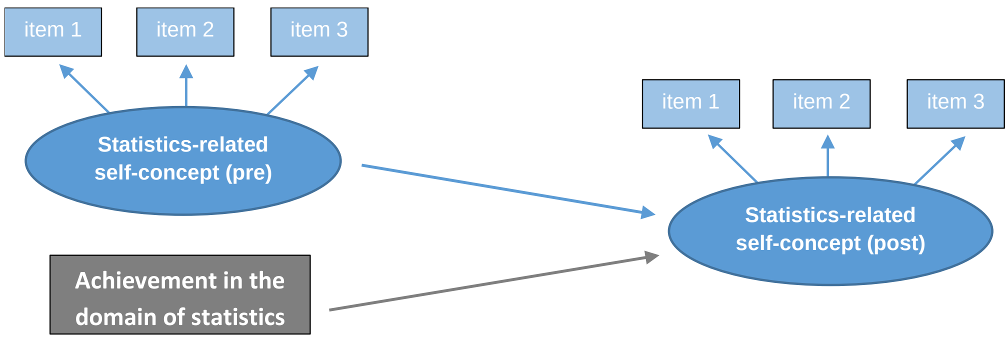


Figure 2. Example of a latent regression model (research question 3)

In order to evaluate whether students' pretest achievement in the domain of statistics influenced the development of their statistics-related self-concept and interest in the course of the intervention (research question 3), we specified similar latent regression models for self-concept and interest, respectively, within the intervention group. Figure 2 displays the structure of a regression model predicting statistics-related self-concept in the posttest by its pretest as well as pretest achievement in the domain of statistics.

Again, taking the pretest value of self-concept or interest into account as covariate controls for possible original differences in this variable, and allows the authors to determine how strong prior achievement in the domain of statistics affects the development of self-concept and interest, respectively.

4. RESULTS

4.1. 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 analysis of pre- and posttest. We compared global model fit (see Table 1) of the full model (Model D) with the fit of two reduced models: Model B includes one factor combining mathematics-related self-concept and interest and another factor combining statistics-related self-concept and interest; Model C includes one factor combining mathematics- and statistics-related self-concept and another factor combining 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 to one factor. Global fit indices displayed in Table 1 indicate that the full model (Model D) best fits the empirical data. Model comparisons by means of chi-square difference tests confirm that the full model is significantly better than the other ones. Model fit indices for the posttest were very similar to or even better than those reported for the pretest.

Table 1. Global fit indices for the tested models (pretest)

Model	χ^2 (df)	p	χ^2/df	CFI	RMSEA
Model A (1 factor): sc & int for math & stat	2279.37 (135)	< 0.001	16.88	0.69	0.18
Model B (2 factors): sc & int math vs. sc & int stat	1320,87 (134)	< 0.001	9.86	0.83	0.13
Model C (2 factors): sc math & stat vs. int math & stat	1537,54 (134)	< 0.001	11.47	0.80	0.14
Model D (4 factors): sc math vs. int math vs. sc stat vs. int stat	330.56 (129)	< 0.001	2.56	0.97	0.06

Note. sc … self-concept; int … interest; math … mathematics; stat … statistics.
See Section 3.3 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 (indicator reliabilities ≥ 0.4; statistically significant factor loadings; factor reliabilities ≥ 0.8, see Table 2; average variance extracted ≥ 0.7; Fornell-Larcker criterion met). Retest reliability of the baseline group was satisfactory (see also Table 2) for the four scales, further indicating that the same constructs were measured in pre- and posttests.

Table 2. Reliabilities and sample-items of mathematics- and statistics-related self-concept and interest

Construct	# of Items	Sample item	Reliability Pre	Reliability Post	Retest reliability (baseline)
Self-concept (Mathematics)	6	Understanding mathematics is easy for me.	0.90	0.93	0.93
Interest (Mathematics)	6	Doing mathematics is one of my favorite activities.	0.93	0.94	0.92
Self-concept (Statistics)	3	Understanding tasks with diagrams and statistical data is easy for me.	0.84	0.73	0.79
Interest (Statistics)	3	Doing statistics is one of my favorite activities.	0.85	0.87	0.78

Note. Scales adapted from Pekrun et al. (2007) as used by Gundlach et al. (2010a, 2010b). Factor reliabilities estimated by AMOS 22.0, retest reliability estimated by SPSS 22.0.

Precise factor loadings and intercorrelations as well as the structure of the confirmatory factor analysis for the full model (pretest) can be found in Figure 3. Similar factor parameters 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 between 0.34 and 0.78.

4.2. EFFECTS OF THE INTERVENTION

In a first step, we investigated potential differences in the development of mathematics- and statistics-related (general and task-specific) self-concept and interest between the four treatment groups. For this purpose, we specified regression models such as displayed in Figure 1 within the intervention group to test the treatments 2 (variation), 3 (reduction) and 4 (variation and reduction) against the basic training. All of the regression weights for the group assignment were non-significant with p-values ranging from 0.102 to 0.996. This means that there were no significant differences at the $\alpha = 0.05$ level concerning mathematics- and statistics-related self-concept and interest, respectively between the basic training and the other treatment groups.

In a second step, we examined whether there were differences between the four treatment groups and the baseline group with respect to domain-specific self-concept and interest. Hence, we combined the four treatment groups into one intervention group. We again used regression models such as illustrated in Figure 1 to investigate whether the development in this intervention group was significantly different from the baseline group (see p-value in Table 3).

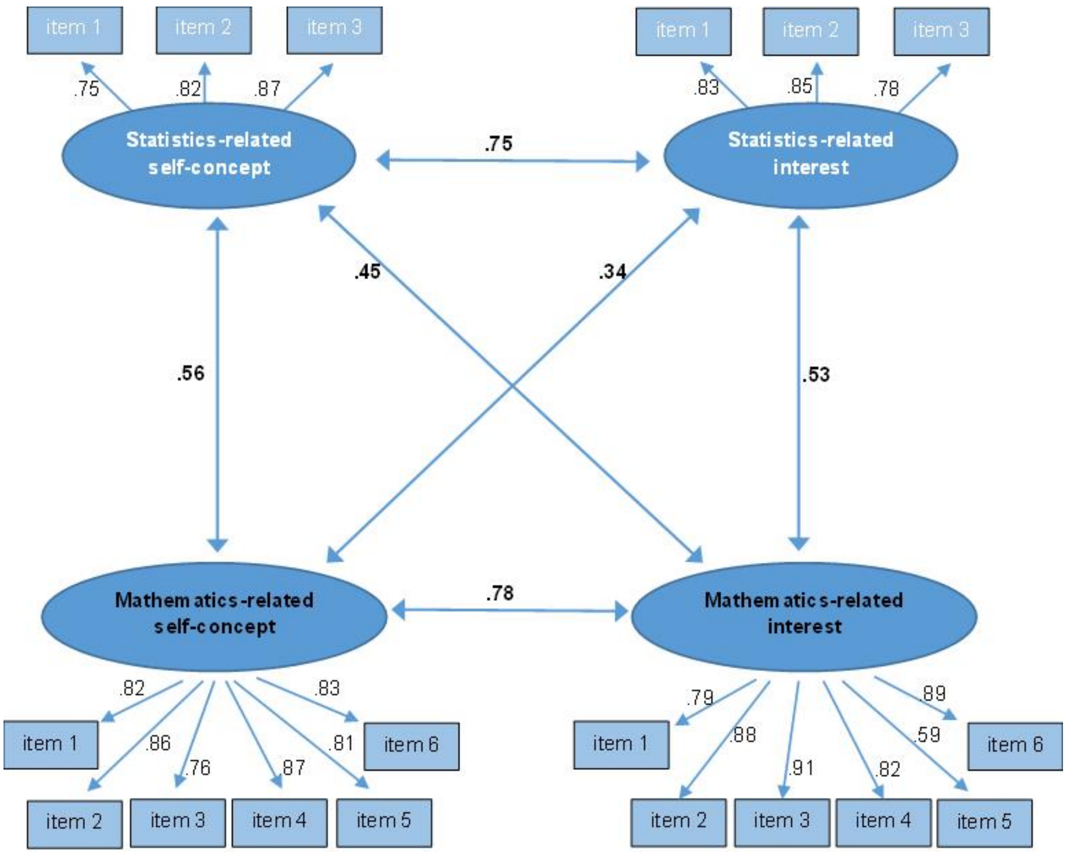


Figure 3. Factorial structure and parameters of the pretest items

Table 3 displays means and standard deviations of the motivational scales 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 (general, non-task-specific scales), both students' self-concept and interest significantly increased from pre- to post-test in the intervention group. The effect sizes as measured by Cohen's[4] *d* reached a small to medium magnitude (Cohen, 1988) indicating that the increase in the intervention group was relevant.

As explained above, statistics-related interest and self-concept were additionally measured by task-specific scales. Reliabilities (Cronbach's alpha) for these scales ranged from 0.67 to 0.89 in pre- and posttest. Taking into account that these scales were based on only three tasks, their reliabilities appear to range from acceptable to good. The absolute means of these scales were higher than for the more general statistics-related self-concept and interest scales (see Table 3). Similar to the statistics-related scales not referring to concrete tasks, the means of the task-specific self-concept and interest scales both increased significantly from pre- to posttest in the intervention group. The growth of these variables also ranged between small and medium effect sizes.

[4] Cut-off values for Cohen's *d*: small effect: $d = 0.2$; medium effect: $d = 0.5$; large effect: $d = 0.8$.

Latent correlations between the statistics-related task-specific and non-task-specific self-concept amounted to 0.62 ($p < 0.001$) for the pretest and 0.63 ($p < 0.001$) for the posttest. Latent correlations between the statistics-related task-specific and non-task-specific interest were even higher (pretest: 0.68, $p < 0.001$; posttest: 0.74, $p < 0.001$).

Table 3. Students' self-concept and interest related to mathematics and statistics in pre- and post-test

Construct	Pretest		Posttest		*p*-value	Cohen's ***d***
	M	(SD)	M	(SD)		
Self-concept (Mathematics)						
Intervention	3.28	(0.93)	3.29	(0.94)	0.064	0.01
Baseline	2.86	(0.96)	2.79	(1.00)		
Interest (Mathematics)						
Intervention	2.62	(1.05)	2.67	(1.06)	0.031	0.05
Baseline	2.24	(1.01)	2.19	(1.02)		
Self-concept (Statistics)						
Intervention	3.23	(0.96)	3.45	(0.90)	0.006	0.24
Baseline	3.13	(0.92)	3.13	(0.95)		
Interest (Statistics)						
Intervention	2.40	(1.03)	2.63	(1.07)	0.028	0.22
Baseline	2.42	(0.93)	2.44	(0.99)		
Self-concept (Task-specific)						
Intervention	3.73	(0.89)	4.12	(0.83)	0.001	0.45
Baseline	3.55	(0.85)	3.67	(0.96)		
Interest (Task-specific)						
Intervention	3.14	(1.03)	3.28	(1.11)	0.001	0.13
Baseline	2.97	(0.90)	2.78	(1.07)		

Note. M = mean; SD = standard deviation.

We did not observe a uniform increase in statistics-related self-concept and interest for all of the learners. Computing the differences from pre- to posttest within the intervention group revealed the changes of non-task-specific self-concept and interest as displayed in Figure 4.

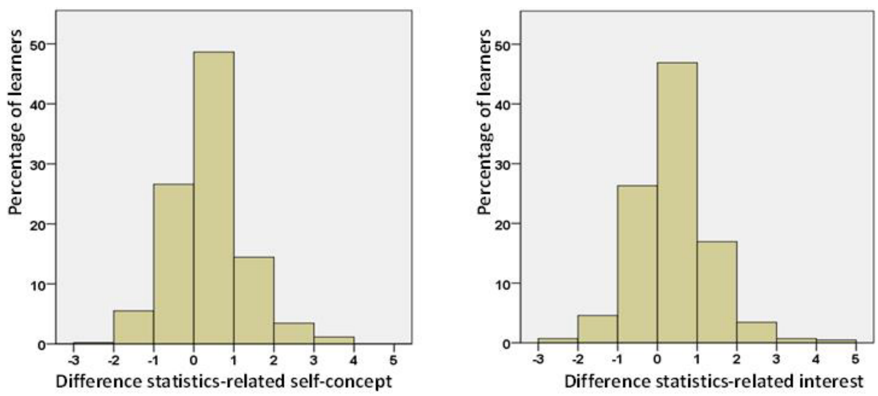


Figure 4. Differences in the development of statistics-related self-concept and interest

Note. Development measured in absolute point differences of the Likert scales.

As expected from the positive average increase, most of the learners developed their self-concept and interest in a positive way. However, some of them stayed at the same level and a number of students even showed lower self-concept and interest scores after the intervention. The design of this study does not afford full insight into the reasons for this variation. However, in the next section, we will examine the role of prior statistics-related achievement for the development of statistics-related self-concept and interest during the intervention.

4.3. PRIOR ACHIEVEMENT AND POST SELF-CONCEPT AND INTEREST

As achievement may influence self-concept and interest (e.g. Marsh et al., 2005), we also investigated whether the growth in statistics-related self-concept and interest in the course of the intervention depends on students' prior statistics-related achievement. Therefore, we first specified a multiple regression model for self-concept within the intervention group (see Figure 2).

Table 4 shows the precise parameter estimates of the latent regression analysis. Including the pretest score for self-concept controls for initial differences in this variable. Even after including pretest self-concept, the achievement score turned out as a significant predictor for the development of statistics-related self-concept. As achievement in the domain of statistics is positively correlated with pretest self-concept (latent correlation of 0.27 (SE 0.05), $p < 0.001$), the magnitude of the resulting regression weight for achievement is lowered through the inclusion of pretest self-concept into the model. The positive regression weight indicates that learners with higher initial achievement scores benefitted more from the intervention than did their lower-achieving peers, hence an aptitude-treatment interaction.

Table 4. Multiple regression model predicting statistics-related self-concept (posttest)

Predictor	Standardized regression weight	Standard error	*p*-value
Statistical self-concept (pretest)	0.64	0.04	< 0.001
Statistical achievement (pretest)	0.11	0.04	0.004

The same procedure was used to evaluate whether students' achievement score influenced the development of their statistics-related interest. The regression weight for achievement indicates a relationship in the same direction as seen for self-concept but its regression weight was non-significant (see Table 5).

Table 5. Multiple regression model predicting statistics-related interest (posttest)

Predictor	Standardized regression weight	Standard error	*p*-value
Statistical interest (pretest)	0.72	0.03	< 0.001
Statistical achievement (pretest)	0.06	0.04	0.176

5. DISCUSSION AND CONCLUSIONS

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.

5.1. IMPLICATIONS OF THE PRESENT STUDY

Following the recommendation of Gal et al. (1997), we first evaluated whether 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 measured the underlying variables reliably. Our finding confirms that students perceived mathematics and statistics differently from a motivational point of view. Hence, although in Germany statistics is taught as a part of mathematics education, students appear to recognize the specificity of statistics. This might be due to its interdisciplinary nature or the relevance of the context (Cobb & Moore, 1997), but it is also possible that students distinguished statistics from other content domains in the mathematics classroom according to specific contents such as, for instance, working with diagrams or calculating mean values. Therefore, in both cases 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 within the expected range (e.g., Carmichael et al., 2010; Klieme et al., 2001) and underline the proximity among these motivational variables. Latent correlations of self-concept and interest of approximately 0.50 between the variables related to the two domains confirmed, despite their separability, the relatedness between mathematics and statistics. Our findings extend prior research on the scales of Gundlach et al. (2010) by providing information about an adequate model fit obtained through confirmatory factor analysis. Thus, all in all, our analysis confirms that the measures can be used in future studies.

The second and main research interest was to investigate whether 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 learning (cf. Gal et al., 1997; Nagy et al., 2006), as well as a reported decline of these motivational variables in the course of schooling (e.g., Carmichael et al., 2010; Frenzel et al., 2010), emphasize the relevance of this focus. Concerning the domain of mathematics, self-concept and interest remained relatively stable from pre- to post-tests. As the intervention focused particularly on statistics, this finding meets the expectation and further underlines that mathematics-related motivational variables have shown to be separable from their statistics-specific counterparts. Moreover, in the sense of considering retest reliability, it can be seen as an indicator for the good quality of the measurement of the variables. However, there was a significant increase in statistics-related self-concept and interest with relevant effect sizes in the course of the intervention. This effect was observed consistently both for general and task-specific scales, respectively. Hence, in accordance with the literature (e.g., Hidi & Renninger, 2006; O'Mara et al., 2006), this finding implies that self-concept and interest for statistics could be fostered through the intervention 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 variants of the treatments were, above all, designed to have a different impact on students' statistics-related achievement, we did not have any specific expectations referring to potential differences related to motivational variables. It can be assumed that the implementation of the intervention as a whole—regardless of some difference in foci—was suitable to foster students' self-concept and interest. In accordance with the literature (Hattie & Timperley, 2007; O'Mara et al., 2006), the increase of statistics-related self-concept may be caused by experiencing an increase in achievement due to the intervention. Feedback might have played a key role here: As the students got feedback in regular intervals, they might have become aware of their improved achievement, which might have led to an increase in their self-concept. As autonomy, experience of achievement and social relatedness constitute crucial supporting factors for interest (Hidi & Renninger, 2006; Krapp, 2005), we conclude that working with a partner rather independently from the teacher on the statistics-related tasks was helpful for fostering interest in statistics.

As there was exclusively an increase in the statistics-related scales, it has to be noted that the implementation characteristics have to be seen in close connection with the topic domain in which the intervention was located. Hence, the combination of the properties of the learning environment and the focus on statistics may have led to this increase. Moreover, this finding confirms Marsh et al. (2005) and O'Mara et al. (2006) with respect to synergies in supporting achievement and motivation simultaneously. For educational practice, this suggests that students' self-concept and interest related to statistics can be fostered by working on statistical problems in a classroom setting that provides autonomy, the experience of achievement, and social learning.

There might also be alternative explanations for our findings. It might be argued, for instance, that prior to the intervention, students did not know much what statistics is all about, which would then imply that the difference between pre- and post-test may indicate that their idea about the nature of statistics had changed. However, according to the German standards, students in 8th grade are not unfamiliar with statistics. Another argument against this alternative explanation can be seen in our results referring to task-specific self-concept and interest. Although the tasks given represent a very limited part of the nature of statistics, they constitute typical activities in the domain of statistics as described by the German standards and are more easily seized and comprehended by students than purely word-based items such as the general statistics-related ones. 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 associated the notion of statistics with tasks such as the presented ones referring to tables, bar charts and characteristic values. Hence, their understanding of statistics may not have changed substantially in the course of the intervention. Altogether, it can be concluded that students had an adequate concept of the term statistics when responding to the statistics-related items.

Nevertheless, there were differences in the development of general statistics-related and task-specific self-concept and interest. For instance, the means of task-specific scales were higher than the statistics-related ones. Hence, students may imagine that statistics can also involve more complicated and less interesting problems than those presented in the corresponding questionnaire. Moreover, the increase of self-concept tended to be larger in the task-specific scales than in the general statistics-related scales; however, for interest the opposite was true, the increase in general statistics-related scales exceeded the increase in task-specific scales. This may indicate that students on the one hand had the perception that they had become more competent in these tasks compared to other statistics tasks but on the other hand that these tasks were perceived in the course of the intervention as less interesting than statistics in general. Such different perceptions of the two scale formats could have been expected as the task-specific scales were limited to particular contexts and activities and hence rather referred to self-efficacy or situational interest that can easily be influenced by specific task characteristics (Bong & Skaalvik, 2003; Gardner, 1985).

As in mathematics, achievement may also impact self-concept and interest in the domain of statistics (e.g. Marsh et al., 2005). The third research question addressed whether previous achievement in the domain of statistics influenced the development of statistics-related self-concept and interest. We found a significant aptitude-treatment interaction for self-concept, that is, students' self-concept developed more positively if they had already shown relatively high achievement referring to statistics in the pretest. Concerning statistics-related interest, prior achievement did not significantly impact its development. These findings are in line with Carmichael et al. (2010) or Marsh et al. (2005) who found domain-specific achievement to be more relevant for the development of self-concept than for interest. Referring to our intervention, this could be explained by the fact that in particular higher-achieving learners got positive feedback and hence improved their statistics-related self-concept in the course of the intervention. It further appears that although the experience of achievement during the intervention may have been important to develop interest (Krapp, 2007), the actual achievement before the intervention was not predictive for interest development. The finding that higher-achieving students increased their self-concept to a larger degree than lower-achieving students has implications for instructional practice related to our learning material. This implies, in accordance with Hood et al. (2012), that students with lower achievement in the domain of statistics do not only have to be particularly supported concerning their cognitive development but also concerning their self-concept.

Altogether, we conclude that it is possible to support statistics-related self-concept and interest through learning environments such as the presented one. For the above mentioned reasons, it appears to be crucial that such learning environments refer to the specific domain of statistics and that they provide opportunities for students to experience autonomy, achievement, and social relations. If these implementation characteristics are met, it appears to be possible to foster such motivational variables through interventions that also focus on students' cognitive development (see Sproesser et al., 2015a). As achievement turned out to influence the development of self-concept and in line with Marsh et al. (2005) or O'Mara et al., 2006), such combined interventions may be an appropriate means to support students' learning.

5.2. LIMITATIONS AND FURTHER RESEARCH

Although the statistical prerequisites necessary to apply the mentioned quantitative methods 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. Moreover, it should be noted that similar findings for other subject areas (e.g. Marsh et al., 2005) suggest that increasing self-concept and interest through experiencing autonomy, achievement and social relatedness is not limited to the particular domain of statistics. However, this study implies that it is possible to foster in students particular self-concept and interest related to statistics through learning environments with these characteristics.

One particular limitation of this study is that we do not exactly know which specific characteristic of the treatment and its implementation contributed most to the observed effects. This question beyond the research questions of this study should be explored in further research: Experimental designs should be used for gaining insight into the influence of the provision of feedback, the experience of autonomy, social relatedness, and further factors. Additionally, more detailed instruments could be used to investigate the reasons of the increase in self-concept and interest. Beyond questionnaires, structured interviews may reveal more detailed insight into the observed effects.

For the implementation of the regular mathematics classroom in the baseline group, we relied on the teachers' reports: For reasons of ecological validity, the researchers did not observe the classrooms of the baseline group. Hence, we trusted the participating teachers that they did not expose their students to statistical contents during this time.

Moreover, 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 achievement and also 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.

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

ACKNOWLEDGEMENTS

We thank Bob delMas for very helpful comments on a previous version of this paper. Ute Sproesser was 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-Württemberg. The research project ReVa-Stat is supported by research funds from Ludwigsburg University of Education.

REFERENCES

Aits, U., Berkemeier, H., Hecht, W., Heske, H., Koullen, R., & Ostrow, D. (2006). *Mathematik konkret, Realschule Baden-Württemberg, Band 4*. Berlin, Germany: Cornelsen.

Artelt, C., Baumert, J., Julius-McElvany, N., & Peschar, J. (2003). *Learners for life: Student approaches to learning. Results from PISA 2000*. Paris, France: OECD. [Online: www.oecd-ilibrary.org/education/learners-for-life-student-approaches-to-learning_9789264103917-en]

Baumert, J., Bos, W., & Lehmann, R. (Eds.) (2000). *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn Vol. 2: Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe.*. Opladen, Germany: 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, Germany: Leske + Budrich.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*(2), 122–147.

Bollen, K. A., & Long, J. S. (Eds.) (1993). *Testing structural equation models*. Newbury Park, CA: Sage Publications.

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.
[Online: iase-web.org/Publications.php?p=SERJ_issues]

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: how different are they really? *Educational Psychology Review*, *15*(1), 1–40.

Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3rd ed.). München, Germany: Pearson.

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, 10*, 83–93.

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, *104*(9), 801–823.
[Online: www.macalester.edu/~kaplan/statchat/2009-2010-year/2010-01/2010-01-agenda.html].

Cohen, J. (1988): *Statistical power analysis for the behavioral sciences (*2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Deci, E. L., & Ryan, R. M. (Eds.) (2002). *Handbook of self-determination research.* Rochester, NY: University of Rochester.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*, 109–132.

Emmioğlu, E., & Çapa-Aydın, Y. (2012). Attitudes and achievement in statistics: A meta-analysis study. *Statistics Education Research Journal*, *11*(2), 95–102.
[Online: iase-web.org/Publications.php?p=SERJ_issues].

Frenzel, A. C., Goetz, T., Pekrun, R., & Watt, H. M. G. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on Adolescence*, *20*(2), 507–537.

Frenzel, A. C., Pekrun, R., Dicke, A. L., & Goetz, T. (2012). Beyond quantitative decline: conceptual shifts in adolescents' development of interest in mathematics. *Developmental Psychology, 48*(4), 1069–1082.

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, The Netherlands: IOS Press.
[Online: iase-web.org/Books.php]

Gardner, P. L. (1985). Students' attitudes to science and technology: An international overview. In M. Lehrke, L. Hoffman, & P. L. Gardner (Eds.), *Interests in science and technology education*, *12th IPN-Symposium Kiel* (pp. 15–34). Kiel, Germany: Institute for Science Education.

Griesel, H., Postel, H., & vom Hofe, R. (2005). *Mathematik heute. Realschule Baden-Württemberg. Band 4.* Braunschweig, Germany: Schroedel.

Gundlach, M., Kuntze, S., Engel, J., & Martignon, L. (2010a). Einflussgrößen auf Statistical Literacy von Studierenden – Erste Ergebnisse aus dem Projekt RIKO-STAT unter besonderer Berücksichtigung motivationaler Dispositionen. In A. Lindmeier & S. Ufer (Hrsg.), *Beiträge zum Mathematikunterricht 2010* (pp. 373–376). Münster, Germany: WTM-Verlag.

Gundlach, M., Kuntze, S., Engel, J., & Martignon, L. (2010b). 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 International Conference on Teaching Statistics (ICOTS 8).* Voorburg, The Netherlands: ISI.
[Online: iase-web.org/Conference_Proceedings.php?p=ICOTS_8_2010]

Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and achievement / performance measures. *Review of Educational Research*, *52*(1), 123–142.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112.

Helmke, A., & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Ed.), *Enzyklopädie der Psychologie*, *Band 3: Psychologie der Schule und des Unterrichts* (pp. 71–176). Göttingen, Germany: 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. A., & 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.
[Online: iase-web.org/Publications.php?p=SERJ_issues]

Holmes, P. (1980). *Teaching statistics 11-16.* Berkshire, England: 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: A Multidisciplinary Journal*, *6*(1), 1–55.

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* (pp. 141–191). Opladen, Germany: Leske + Budrich.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford.

KMK (Kultusministerkonferenz) (2003). *Bildungsstandards im Fach Mathematik für den mittleren Schulabschluss.* München, Germany: Wolters Kluwer.

Köller, O., Baumert, J., & Schnabel, K. (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *Journal for Research in Mathematics Education*, *32*(5), 448–470.

Krapp, A. (2005). Basic needs and the development of interest and intrinsic motivational orientations. *Learning and Instruction 15*(5), 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*(2), 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 8th International Conference on Teaching Statistics.* Voorburg, The Netherlands: International Statistical Institute.
[Online: iase-web.org/Conference_Proceedings.php?p=ICOTS_8_2010]

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.* [Online: tsg.icme11.org/tsg/show/15]

Land Baden-Württemberg (2004). Bildungsstandards für Mathematik Realschule–Klassen 6, 8, 10. [Online: www.bildung-staerkt-menschen.de/service/downloads/]

Lindmeier, A., Kuntze, S., & Reiss, K. (2007). Representations of data and manipulations through reduction – competencies of German secondary students. In B. Phillips & L. Weldon (Eds.), *Proceedings of the IASE Satellite Conference on Assessing Student Learning in Statistic*s.
[Online: iase-web.org/Conference_Proceedings.php?p=Assess_Stud_Learn_2007]

Maroska, R., Olpp, A., Pongs, R., Stöckle, C., Wellstein, H., & Wontroba, H. (2006). *Schnittpunkt 4. Mathematik Baden Württemberg*. Stuttgart, Germany: Klett.

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. (2005). Academic self-concept, interest, grades, and standardized test scores: reciprocal effects models of causal ordering. *Child Development*, *76*(2), 397–416.

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*(2), 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, NY: Macmillan.

Möller, J., & Köller, O. (2004). Die Genese akademischer Selbstkonzepte: Effekte dimensionaler und sozialer Vergleiche. *Psychologische Rundschau, 55*(1), 19–27.

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*(4), 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., vom Hofe, R., Blum, W., Frenzel, A. C., Goetz, T., & Wartha, S. (2007). Development of mathematical competencies in adolescence: The PALMA longitudinal study. In M. Prenzel (Ed.), *Studies on the educational quality of schools. The final report on the DFG Priority Programme* (pp. 17-37). Münster, Germany: Waxmann.

Pekrun, R. & Zirngibl, A. C. (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 (Eds.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (pp. 191–210). Münster, Germany: Waxmann.

Ramirez, C., Schau, C., & Emmioğlu, E. (2012). The importance of attitudes in statistics education. *Statistics Education Research Journal*, *11*(2), 57–71.
[Online: iase-web.org/Publications.php?p=SERJ_issues]

Scheaffer, R. L., Gnanadesikan, M., Watkins, A., & Witmer, J. A. (1996). *Activity-based statistics: Instructor Resources.* New York, NY: Springer.

Schau, C., & Emmioğlu, E. (2012). Do introductory statistics courses in the United States improve students' attitudes? *Statistics Education Research Journal*, *11*(2), 86–94. [Online: iase-web.org/Publications.php?p=SERJ_issues]

Schau, C., Stevens, J., Dauphinee, T. L., & Del Vecchio, A. (1995). The development and validation of the Survey of Attitudes toward Statistics. *Educational and Psychological Measurement*, *55*(5), 868–875.

Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist, 26*(3&4), 299–323.

Schukajlow, S. & Krug, A. (2014). Do multiple solutions matter? Prompting multiple solutions, interest, competence, and autonomy. *Journal for Research in Mathematics Education, 45*(4), 497–533.

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: validation of construct interpretations. *Review of Educational Research, 46*(3), 407–441.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

Sproesser, U., Engel, J., & Kuntze, S. (2015a). Einflussfaktoren auf die Entwicklung von Statistical Literacy und von Sichtweisen auf zufallsbedingte Variabilität in der achten Realschulklasse. *Lernen und Lernstörungen 4*(2), 115–129.

Sproesser, U., Engel, J., & Kuntze, S. (2015b). Raising self-concept and interest for statistics through activity-based instruction. In M. A. Sorto (Ed.), Advances in statistics education: developments, experiences, and assessments. *Proceedings of the IASE Satellite Conference on Advances in Statistics Education: Developments, Experiences, and Assessments*. [Online: iase-web.org/Conference_Proceedings.php?p=Advances_in_Stats_Education_2015]

Sproesser, U., Engel, J., & Kuntze, S. (2015c). Supporting self-concept and interest in statistics. In K. Beswick, T. Muir, & J. Wells, (Eds.). *Proceedings of 39th Conference of the International Group for the Psychology of Mathematics Education*, (Vol. 1, p. 204). Hobart, Australia: PME.

Sproesser, U., Kuntze, S., & Engel, J. (submitted). Models and representations in statistical contexts: students' competencies and impact factors. *Journal für Mathema-tik-Didaktik.*

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. [Online: iase-web.org/Publications.php?p=SERJ_issues]

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 (with a discussion of T. M. F. Smith, D. Moore, N. E. Breslow, R. D. Snee, R. Biehler, and a response by the authors). *International Statistical Review 67*(3), 223–265.

UTE SPROESSER
Institut für Mathematik und Informatik, Abteilung Mathematik
PH Heidelberg
Im Neuenheimer Feld 561
69120 Heidelberg
GERMANY

APPENDIX

1. SAMPLE TASKS OF THE INTERVENTION

Tasks from the basic training:
Mary says: “Obviously, there are 24 chocolate lentils in each package. Assuming that the factory produces the same number of every chocolate lentil color, I expect 4 chocolate lentils of every color in my package.”

1. Predict the number of chocolate lentils of every color in your package (column *my prediction*).
2. Open your package and write down the actual numbers of each color.
3. Take 9[5] package-cards and compare the displayed numbers with the numbers of your own package.
4. Sum up the numbers of the red / orange / yellow / … chocolate lentils from the nine cards and from your own package and write them down in the table.
5. Calculate the average number of every color in the ten packages and write it in the table.

Color	Mary’s prediction	My prediction	Actual numbers in my package	Sum within the 10 packages	Average of the 10 packages
Brown	4				
Green	4				
Red	4				
Blue	4				
Yellow	4				
Orange	4				
Sum	24	24	24	240	—

Additional reflective tasks in the treatment “variation”

1. Are the prediction of Mary, your prediction and the actual numbers identical?
2. If they were not identical: What could be the reason for the differences? Have you made a mistake with your prediction?
3. How could the color distribution in another package be?
   Brown: ____, Green: ____, Red: ____, Blue: ____, Yellow: ____, Orange: ____
4. The averages that you have calculated are not all identical. Do you think that Mary’s assumption is wrong (“The factory produces the same number of chocolate lentils of every color”)?

Additional reflective tasks in the treatment “reduction”
Mary has based her prediction on the fact that the factory produces the same number of every chocolate lentil color. Explain in more detail how she deduced her prediction from this fact.
When you look at all the numbers of chocolate lentils from the different packages, you can quickly lose the overview. For this reason, calculating the total sum and the average of each color is often helpful because this procedure provides one value for every color and you can compare the numbers of the colors more easily. Particularly the average is an important

[5] Students of the basic training were requested to work with 19 additional package-cards in order to keep the learning times parallel despite the reflective tasks for the other treatments.

parameter to get an overview. However, much information can get lost when you calculate the average.

1. Which information gets lost when you calculate the average?
2. Which advantage do you see in calculating the average?

2. CONTEXTS AND CONTENTS ADDRESSED IN THE INTERVENTION

Within the four lessons of the intervention the students worked individually on a number of learning environments. The order of these learning environments was identical for all of the students. However, as each student could proceed at his or her own speed, most of them did not work on all environments.

No.	Context of the learning environment	Statistical content emphasized in learning environment beyond exploring data and distributions
1	Chocolate lentils	Average, range
2	Body height	Average, mode
3	Geyser	Ranked list, minimum, maximum, median
4	Paper planes	Ranked list, median
5	Salaries	2×2 table, comparing median, and average
6	Marbles	—

3. ITEMS OF THE MOTIVATIONAL SCALES USED IN THIS STUDY

<u>Mathematics-related items:</u>

Self-concept (Pekrun et al., 2007):

Understanding mathematics is easy for me.
Concerning mathematics, I am a talented student.
I usually know the answer when the mathematics teacher asks a question.
I am good at solving mathematical problems.
Mathematics tests are easy for me.
Learning for mathematics is easy for me.

Interest / intrinsic motivation (Pekrun et al., 2007; see also Frenzel et al., 2012):

I do my mathematics homework because I like this subject.
Doing mathematics is one of my favorite activities.
I contribute to the mathematics class because I am interested in mathematics.
In mathematics class I make an effort because I am interested in this subject.
I am interested in mathematics.
I like to read books and solve brain teasers related to mathematics.

<u>General statistics-related items:</u>

Self-concept (Gundlach et al., 2010a, 2010b):

Understanding tasks with diagrams and statistical data are easy for me.
I am good at solving statistical problems.
I think that I would perform well on the statistics tasks of a test.

Interest (Gundlach et al., 2010a, 2010b):

I am interested in statistics.
I like working with statistical data and diagrams.
Doing statistics is one of my favorite activities.

Task-specific statistics-related items:

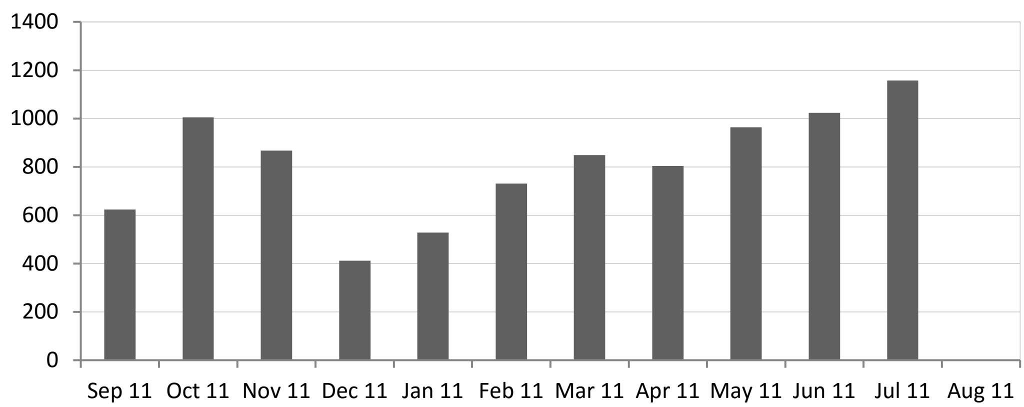
What do you think about this problem?					
A survey on the web usage per week revealed the following data (in hours): 10.5; 6.5; 9; 23; 7; 0; 6.5; 26; 3; 15; 12; 13.5; 1.5; 0.5; 12; 7; 3.5 Identify the minimum, the maximum and the range of this data!	Completely true.	Largely true.	Rather true.	Rather not true.	Not true at all.
I am certain that I can solve this problem correctly.	☐	☐	☐	☐	☐
I am interested in working on this problem.	☐	☐	☐	☐	☐

Other tasks used to measure self-concept and interest were, e.g.:

The table displays the grades in mathematics for class 8d from the last school year. Compute the mean of these grades!

grade	1	2	3	4	5	6
frequency	1	7	12	5	2	0

The Goethe school's caretaker records how many bottles of ice tea he has sold in his snack bar.



How many bottles of ice tea have approximately been sold in September and October 2011?