Evaluation of a 'Painting and Puzzles Exercise Book for Statistics' for psychology first year students

Kim L. Austerschmidt, Denise Kerkhoff, Sarah Bebermeier & Anne Hagemann

Statistics courses are challenging for many psychology students. We designed a 'Painting and Puzzles Exercise Book for Statistics' for first year psychology undergraduates, to repeat and deepen the course content and get prepared for the exam. We describe the development and characteristics of the book and report a longitudinal evaluation study. Users (N=72) rated the book positively and judged it as helpful, easy to use, and enjoyable. Students with initially higher skills, younger students and, those who graduated from school recently were more likely to use it. We matched users to non-users from the preceding and the evaluation cohort, respectively, on characteristics at study entry and found positive effects on achievement (exam grade and perceived management of course content) for users. We reason that the book is a valuable support that can be embedded in an ongoing course easily.

Keywords: *statistics; drawing; psychology first year students; learning support; active learning.*

PSYCHOLOGY is an empirical discipline with a strong focus on research methods and statistics. Hence, in undergraduate psychology major programs, a comparatively great number of courses in research techniques and statistical analyses has to be completed and is obligatorily anchored in university curricula (American Psychological Association, 2013; British Psychological Society, 2019; Stoloff et al., 2009). However, it is well known that many students struggle with mathematical and statistical content, especially at the very beginning of their studies (Doyle, 2017; Heublein et al., 2017; Onwuegbuzie & Wilson, 2003).

Beginners in the psychology major program are often overwhelmed by the unexpectedly large proportion of mathematical content (Ruggeri et al., 2008) as many of them aim to work as clinical psychologists or in related fields and do not expect to require extensive statistical knowledge in their future career (Leino et al., 2019). Regularly, students prefer application-oriented courses (Holmes, 2014; Rajecki et al., 2005; Reiß et al., 2011) over the methodological content which they perceive as complex and abstract (Garfield & Ben-Zvi, 2008) and commonly do not recognize the scientific character of the subject of psychology (Fonteye et al., 2015; Holmes, 2014; Veilleux & Chapman, 2017). On the other hand, students are often not adequately prepared for the requirements in statistics courses and lack mathematical and especially statistical skills (Field, 2014; Fonteye et al., 2015) which are crucial for statistics achievement (Chiesi & Primi, 2010; Hood et al., 2012). Overall, subject related mathematical skills of first year students in psychology have declined over the last decades (Carpenter & Kirk, 2017).

A lack of confidence and negative attitudes towards statistics (Dempster & McCorry, 2009; Vanhoof et al., 2006) but also statistics anxiety (Chiesi & Primi, 2010; Macher et al., 2012) are therefore common among psychology students, especially for those with initially low mathematical skills (Dempster & McCorry, 2009; Finney & Schraw, 2003). However, low self-efficacy, negative attitudes, and anxiety impede success and decrease performance (Chew & Dillon, 2014; Dempster, & McCorry, 2009; Kennett et al., 2009; Schau, 2003; Tremblay et al., 2000; Vanhoof et al., 2006) because of the lower motivation and the reduced time students spend on learning statistics (Budé et al., 2007; Hood et al., 2012; Macher et al., 2012).

The rationale for a 'Painting and Puzzles Exercise Book for Statistics'

Because of the high relevance and challenging nature of methods and statistics courses, many tools have been developed to improve teaching, change negative attitudes, support students' learning and thus enhance their academic success (e.g. GAISE College Report ASA Revision Committee 2016; Garfield & Ben-Zvi, 2008, 2009; Zieffler et al., 2008). Whereas technology-based support has been applied and investigated thoroughly in the last years (e.g. Aberson et al., 2000; Austerschmidt & Bebermeier, 2019; Bebermeier et al., 2015; Chance et al., 2007; Goode et al., 2018), analog methods, especially self-learning materials, were not in the focus of research.

In general, during the process of learning, hands-on and autonomous activities support learning outcomes (Herrmann, 2009) and it is known that providing or letting learners generate multiple representations of a concept on their own improves a deeper understanding (Ainsworth, 2014). Especially beginners, who are faced with a lot of new content, could be addressed and supported by graphics and illustrations, which can be preferred over dynamic representations that need much more working memory capacity (Mayer et al., 2005). One way to obtain alternative representations and graphical interaction with the content can be drawing.

While drawing, the learner cannot only consume information but must act intentionally (Fiorella & Zhang, 2018). By integrating verbal and pictorial models, monitoring and regulation activities – and thus learners' cognitive and metacognitive processes – are activated, which in turn enhance deep learning (Leutner & Schmeck, 2014). Besides that, drawing has been shown to improve students' interest and involvement with the content (Meter & Garner, 2005). Furthermore, autonomous visualisation enhances learning success (Schwamborn et al., 2010).

Certain requirements must be met to make drawing an effective learning tool. In their meta-analysis, Cromley et al. (2020) found drawing to especially affect the more basic outcomes of factual and inferential learning, whereas transfer learning is affected to a minor degree. Thus, drawing seems to be most applicable to support the understanding of fundamental concepts at the beginning of education. Nückles and Wittwer (2014) stated that a precondition for the use of pictures in learning is that learners must be familiar with symbols and elements (and able to interpret them correctly) or, alternatively, symbols and elements must be illustrated in legends. Furthermore, learners need time and support to integrate new representations of content into their prior knowledge and earlier representations (Ainsworth, 2014; Meter & Garner, 2005) and accurate instructions seem to make drawing even more effective (Cromley et al., 2020). Finally, self-generated visualizations can function as a kind of self-assessment which can uncover misconceptions and gaps in knowledge (Bobek & Tversky, 2016). Such feedback about knowledge (gaps) gives an indication for the next steps to take in the learning process and affects feelings of competence and self-determination (Eccles & Wigfield, 2002; Herrmann, 2009).

The present research

To offer psychology students an innovative analog learning tool in their statistics education and assess its usability and value, we developed and evaluated a 'Painting and Puzzles Exercise Book for Statistics' (PPB)[1]. The PPB provides alternative support for dealing with statistical content and combines

[1] A German version of the PPB as well as sample solutions can be requested for teaching purposes. If required, please contact the authors.

explanations of content and visualisation of statistical concepts by drawing and calculating statistical parameters by hand. This should train application, enhance students' motivation to work with the course content relevant for the exam, and promote a deeper learning experience and greater understanding by active engagement with the content. Since 2013, we regularly survey our psychology students longitudinally during their first year regarding their academic achievement. In the student cohort 2018, we added items to the second questionnaire at the beginning of the third semester to evaluate the PPB and were thus able to compare users' and non-users' achievement as well as users' achievement with the achievement of non-users from the preceding cohort.

Our research has three main goals. First, we want to explore how users get along with the PPB and depict how they evaluate it. Therefore, students who had used the PPB were asked to rate it regarding different aspects. We expect users to evaluate overall properties positively and rate the PPB as helpful, easy to use, and enjoyable. Also, structure and design and task difficulty as well as the amount of content covered should be rated as suitable. To answer our first research question (RQ1), we examine descriptive values regarding the rating of:

- RQ1a: overall properties
- RQ1b: structure and design and
- RQ1c: task difficulty and amount of content covered.

As a second goal, we investigate which students do (not) use the PPB. We compare users to non-users regarding their characteristics at study entry. Although the offer addresses the enhancement of basic skills and learning motivation, a certain affinity to statistics and level of motivation might make the usage of support services more likely (Bebermeier & Nussbeck, 2014). For psychology students, it has been shown that low self-efficacy beliefs and uncertainty as well as high achievement motivation can lead to the use of basic services despite sufficient objective competencies (Austerschmidt & Bebermeier, 2020). Because the PPB is a newly developed offer, and because we want to explore if users of the PPB differ in aspects of competency, motivation, and demographics, we test for differences regarding skills (i.e. scores on the math assessment), and motivational characteristics (i.e. interest in and perceived relevance of statistics). Furthermore, older participants or those who graduated from school a longer time ago probably have further obligations, like family care or side jobs (Heublein et al., 2010), and are less familiar with mathematical content (Austerschmidt & Bebermeier, 2020). This may decrease time and commitment for the additional offer. On the other hand, it has already been shown that those students prefer flexible support services that do not require attendance (Bebermeier & Nußbeck, 2014). To determine the importance of characteristics (at study entry) for using the PPB, we examine the following exploratory hypotheses:

- H1a: Initial mathematical skills of users and non-users differ.
- H1b: Initial interest in and perceived relevance of statistics of users and non-users differ.
- H1c: Age and high school graduation year of users and non-users differ.

Our third and last goal is to assess the benefits of the PPB since we expect positive effects on learning outcomes. We therefore compare the achievement of users with two samples of non-users: (1) students from the preceding cohort without availability of the PPB, and (2) students from the same cohort who did not use the PPB. To draw reliable conclusions, compared groups should be as similar as possible. We thus match users to non-users of each comparison group (with respect to math skills, attitudes towards statistics and demographic characteristics at study entry) and test for differences in their achievement.

Generally, academic success can be measured by objective as well as subjective criteria

(Sutherland, 2017) and self-perceived competencies are closely related to (later) objective measures of success, such as grades (Austerschmidt & Bebermeier, 2019; Bebermeier et al., 2019). We thus use the statistics grade as an objective and perceived management of the study requirements in statistics as a subjective measure of academic achievement. We hypothesise that, given comparable math skills, attitudes towards statistics, and demographic characteristics at study entry:

- H2a: Users achieve better statistics grades than non-users from the preceding cohort.
- H2b: Users report better management of requirements in statistics than non-users from the preceding cohort.
- H3a: Users achieve better statistics grades than non-users from the same cohort.
- H3b: Users report better management of requirements in statistics than non-users from the same cohort.

Course structure

The course 'Statistics for Psychologists I & II' is obligatory for all psychology students at the beginning of the bachelor's program and is taught four hours weekly during the first (descriptive statistics) and second semester (inferential statistics). Thematic units are presented in Table 1. The course reported here took place in winter term 2018/19 and summer term 2019 and addressed 132 students enrolled in the psychology major programme who had to pass the final exam covering the content of both semesters.

Table 1: Thematic units of the course 'Statistics for Psychologists I & II'.

Unit	Content	Term
1	measurement theory	winter
2	univariate descriptive statistics	winter
3	bivariate descriptive statistics	winter
4	simple and multiple linear regression	winter
5	probability and probability distribution	winter
6	basics of inferential statistics (sampling distributions, confidence intervals)	summer
7	hypothesis tests for metric variables (t-tests, analysis of variance, regression)	summer
8	hypothesis tests for nominal and ordinal variables (chi-square-test, Mann-Whitney U-test, Wilcoxon-test)	summer
9	power analyses	summer

Development and provision of the 'Painting and Puzzles Exercise Book for Statistics'

Five research assistants of the working group 'methods and evaluation' created exercises on the course content (see Table 1), thereby focusing on topics (a) of particular importance for the general understanding of statistics and (b) suitable for calculation tasks, visualisation, and drawing activities. The exercises were peer-reviewed and revised several times. The final PPB consists of 36 pages including a cover page, introduction, and table of contents, introductory texts and 28 exercises (measurement theory: 3, univariate descriptive statistics: 1, bivariate descriptive statistics: 3, simple and multiple linear regression: 6, probability and probability distribution: 7, basics of inferential statistics: 2, hypotheses tests for metric variables: 5, power analyses: 1). We show sample exercises in Figure 1A to 1C.

Figure 1A: Linear regression: Drawing regression lines for different sets of coefficients.

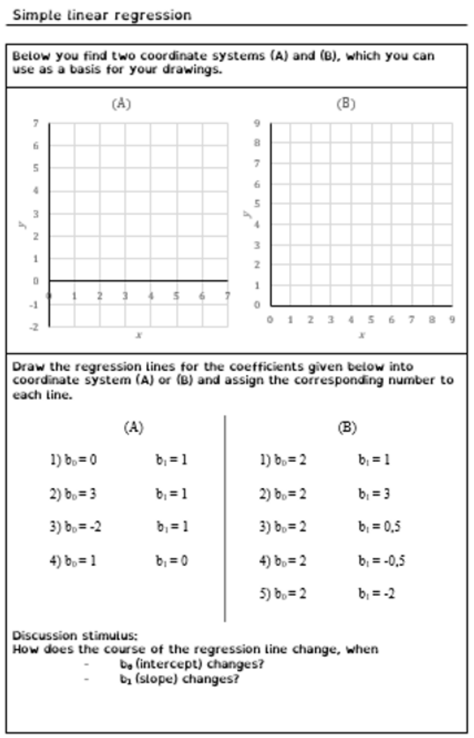


Figure 1B: Odds and probabilities: Drawing diagrams to visualize the difference between odds and probabilities.

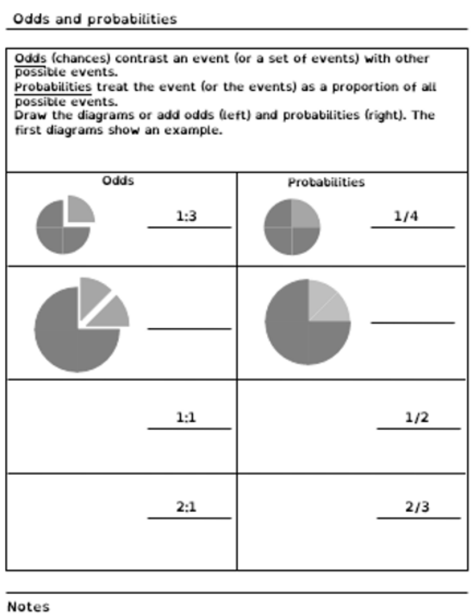
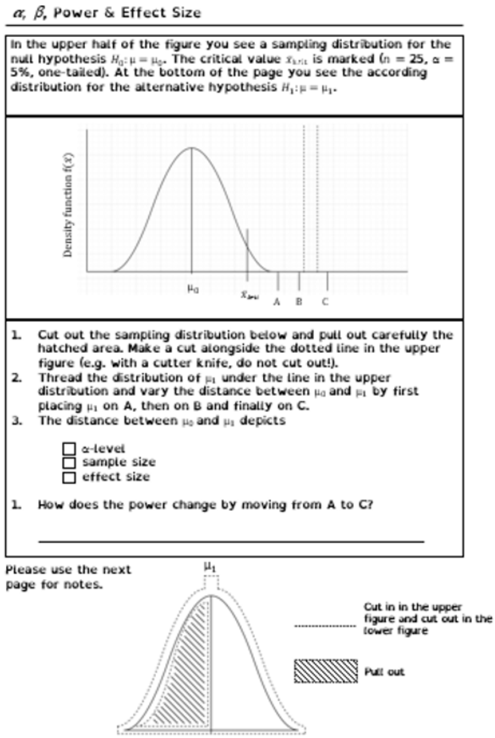


Figure 1C: Basics of inferential statistics: The relation between α, β and power.



To optimally meet competencies and enhance motivation and self-confidence, we constructed exercises which we considered challenging but not too demanding (Eccles & Wigfield, 2002; Herrmann, 2009). Furthermore, exercises were designed in accordance with requirements of later exam tasks. We included allocation exercises to help students consolidate the meaning and application of statistical symbols and expressions. Explanatory texts introduce each topic to point out its relevance for everyday life, for understanding statistics or for working as a psychologist and to help students integrate the concepts into their prior knowledge. Particularly, we used simple and vivid language to ensure an easy access to the content. Finally, to enhance motivation and reduce anxiety and negative attitudes, we included some humorous text elements (Field, 2014; Schacht & Stewart, 1990) and chose real-life and application-oriented tasks (Field, 2014).

In sum, the PPB offers a cohesive collection of exercises on topics relevant for the exam, instead of providing, for example, a mix of different worksheets on different topics. The PPB is provided and recommended by the lecturer and content can be looked up again later, so that students should get motivated to engage with the content further and in more depth.

The PPB was provided in the course roughly six weeks before the last statistics session in the summer term 2019 (i.e. two months before the exam). At the end of the summer term (two weeks before the exam), sample solutions for the exercises were provided online to enable students to check and revise their answers, to assess their performance and, if they had made mistakes, to learn from them.

Method

Psychology students at our university are usually asked to answer one paper-and-pencil questionnaire in their first statistics lecture and a second one after the final statistics exam at the very beginning of their third semester. In both surveys, students generate an individual code which enables us to combine the data of the two times of measurement. For the evaluation of the PPB, some questions relating to this offer were added to the second survey in the cohort 2018. Applicable ethical standards, specifically the APA standards, were followed in the conduct of the study. The rights of the participants in our research were protected.

First survey

In the first survey, students' gender, age, high school graduation year, mathematical skills, and the interest in and perceived relevance of statistics were assessed. It has been shown that basic mathematics tests assessing study relevant competencies are good predictors of future achievement in statistics (Austerschmidt & Bebermeier, 2019; Fonteye et al., 2015). Thus, mathematical skills were operationalized by an assessment of study relevant mathematical content (21 multiple-choice tasks: four on algebra, four on fractional arithmetic, four on percentage calculation, five on probability calculation and four on the interpretation of graphics and tables; internal consistency Cronbach's $\alpha = .63$). We measured the interest in statistics with the 6-point Likert item 'How interested are you in the study contents of methods and statistics?' (1 = *not at all*, 6 = *very*) and the perceived relevance of statistics with three 6-point Likert items ('How relevant do you think is sound knowledge in methods and statistics for a) psychologists in general b) successful graduation c) your later professional activity?', 1 = *not at all*, 6 = *very*, $\alpha = .74$).

Second survey

The second survey again assessed sociodemographic characteristics and perceived management of the study requirements in statistics with the single 6-point Likert item 'How well did you get along with the mathematical study content so far?' (1 = *not at all*, 6 = *very good*). In cohort 2018, also usage and rating of the PPB was assessed. Participants were asked to indicate if they had used the PPB (*yes / no*) and to rate the task difficulty (5-point Likert item: 1 = *much too easy*, 3 = *appropriate*, 5 = *much too difficult*) and amount of content covered (1 = *too little*, 2 = *appropriate*, 3 = *too much*). Further, students rated structure and design of the PPB on four 6-point Likert items (e.g. 'Instructions were clear to follow.', 1 = *does not apply*, 6 = *applies perfectly*, $\alpha = .82$) and the overall evaluation of the PPB on eight 6-point Likert items (e.g. 'I would recommend the PPB to other students', 'The PPB promotes understanding of the course content', 1 = *does not apply*, 6 = *applies perfectly*, $\alpha = .86$).

Exam grades

Additionally, students were asked to provide their individual code on their statistics exam sheet and thereby give consent to link their grade with the longitudinal data. Grades range from 1.0 (best grade) to 5.0 (fail).

Results

Out of 132 students addressed by the statistics course in cohort 2018, 117 (89%) answered the first survey, 90 (68%) answered the second survey, and 65 (49%) answered both. Information about grades are available for 92 students (70%), of which 61 also answered the items relevant to our analyses from both questionnaires[2]. To assess if PPB-users and non-users from the preceding cohort (H2) or the same cohort (H3), with comparable characteristics at study entry, differ in grade and course management, matched samples were generated using the MatchIt package (Ho et al., 2011) in R 3.5.2 (R Core Team, 2016). The matching process was based on math assessment, interest in and perceived relevance of statistics, gender and high school graduation year and used nearest neighbour matching. Due to unequal sample sizes (H1) and non-normality in the data (H1, H2 and H3), we consistently computed

[2] In cohort 2017, out of 137 students addressed by the statistics course, 106 (77%) answered the first survey, 81 (59%) answered the second survey and 48 (35%) answered both. Information about grades are available for 71 students (52%), of which 69 also answered the items relevant to our analyses from the first questionnaire.

Mann-Whitney-U-tests to assess differences in characteristics at study entry between PPB-users and non-users (H1), and differences in academic achievement between PPB-users and non-users from the preceding cohort (H2) or the same cohort (H3).

RQ1: Evaluation of the PPB

Most of the participants who answered the second survey in cohort 2018 indicated that they had worked with the PPB (72 out of 90; 80%). For these cases, descriptive statistics of the ratings will be reported. Regarding RQ1a and RQ1b, we found the overall rating of the PPB (*Mdn* = 4.6, *M* = 4.62, *SD* = 0.74) as well as the rating of structure and design (*Mdn* = 5.3, *M* = 5.15, *SD* = 0.70) to be very positive. Median and arithmetic mean for both measures are located clearly in the upper range of the scale ranging from 1 (lowest value) to 6 (highest value).

Relating to RQ1c, users mostly indicated the difficulty of exercises to be *'appropriate'* (40/72, 56%). 16 users each (22%) deemed the exercises either *'somewhat difficult'* or *'somewhat easy'* and no one rated them as *'much too easy'* or *'much too difficult'*. Further, nearly half of the users rated the amount of content covered by PPB to be appropriate (31/71, 44%), but even more (38/71, 54%) stated they would have appreciated more content to work on. No one chose the option *'too much'*. Thus, many participants were satisfied with the demands, but many participants were not fully satisfied with the coverage of thematic units.

H1: Characteristics of users and non-users

Characteristics at study entry (first survey) are available for 65 out of 90 (72%) of the students in cohort 2018 providing information about usage of the PPB in the second survey. We tested this sample for differences between PPB-users and non-users. We show descriptive statistics for both groups and results of two-tailed Mann-Whitney-U-tests for hypotheses 1a, b, and c in Table 2.

Table 2: Characteristics at study entry of users and non-users of cohort 2018 and results of two-tailed U-tests.

	Users			Non-users					
	n	*Mdn*	*M (SD)*	*n*	*Mdn*	*M (SD)*	*U*	*p*	*r*
Math assessment	54	12.5	12.37 (3.36)	11	10.0	9.64 (2.91)	162.5	.018	-.29
Interest	52	3.0	3.48 (0.98)	11	3.0	3.09 (1.04)	345.5	.259	-.14
Relevance	54	4.8	4.90 (0.67)	11	5.0	4.36 (1.23)	372.5	.183	-.17
Graduation[a]	54	1.0	2.53 (4.44)	11	4.0	4.73 (5.34)	306	.037	-.23
Age	54	20.0	21.22 (4.69)	11	23.0	24.8 (5.34)	142	.006	-.34

Note. *Mdn* = median, *r* = effect size.
[a]years since graduation.

Users solved more tasks in the initial mathematics assessment correctly than non-users, $U = 162.5$, $p = .018$, $r = -.29$, supporting H1a. Contrary to H1b, there are no significant differences for the interest in or the perceived relevance of statistics. However, arithmetic means indicate slightly higher values for users. Further, students who did not use the PPB are older than users, $U = 142$, $p = .006$, $r = -.34$. In line with this, non-users graduated a longer time ago than users, $U = 306$, $p = .037$, $r = -.23$. These findings support H1c.

H2: Achievement of users compared to non-users from the preceding cohort

Information on variables used for matching and outcome variables (grades or management) are available for 51 PPB-users in cohort 2018 and 69 cases from the preceding cohort (H2a, grade) or 43 PPB-users in cohort 2018 and 48 cases from the preceding cohort (H2b, management). For each hypothesis, PPB-users were matched to an equal number of non-users from the preceding cohort. The matching process aimed at an 1:1 ratio and yielded a sample of 102 (H2a) and 86 (H2b) participants (for sample statistics of matched samples see Appendix A), consisting of 51 (H2a) or 43 (H2b) users and an equal number of non-users from the preceding cohort. Matched samples were tested for group differences in grades and management (Table 3).

Table 3: Descriptive values of grade and management and results of one-tailed U-tests for matched samples of users and students from the preceding cohort.

	Users			Preceding cohort					
	n	*Mdn*	*M (SD)*	*n*	*Mdn*	*M (SD)*	*U*	*p*	*r*
Grade[a]	51	1.3	1.73 (0.90)	51	1.7	2.12 (1.04)	995	.019	–.21
Management	43	4	4.16 (0.95)	43	5	4.49 (0.86)	733	.960	.19

Note. *Mdn* = median, *r* = effect size. Users were matched to non-users with respect to math assessment, interest in and perceived relevance of statistics, gender, and high school graduation year.

[a]higher grades denote worse exam performance.

Mann-Whitney-U-tests revealed significant differences between matched PPB-users and non-users from the preceding cohort regarding grade (U = 995, p = .019, r = -.21) in the presumed direction, but not regarding management (U = 733, p = .960, r = .19). Therefore, H2 is partially confirmed. Users achieved better grades but did not report better management than non-users from the preceding cohort. Figures 2 and 3 show the distribution of values for grade and management for the matched groups.

Figure 2: Distribution of grade values for users and students (non-users) from the preceding cohort.

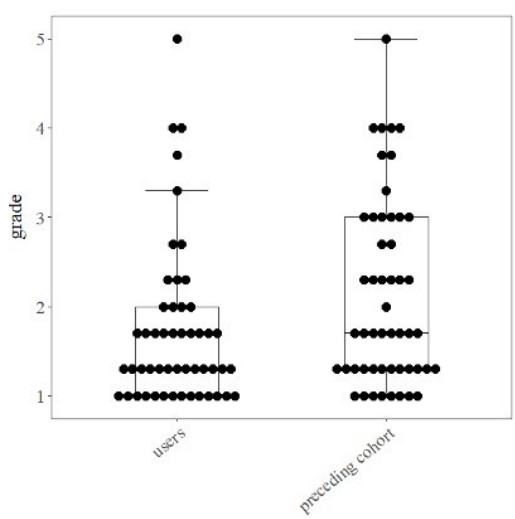
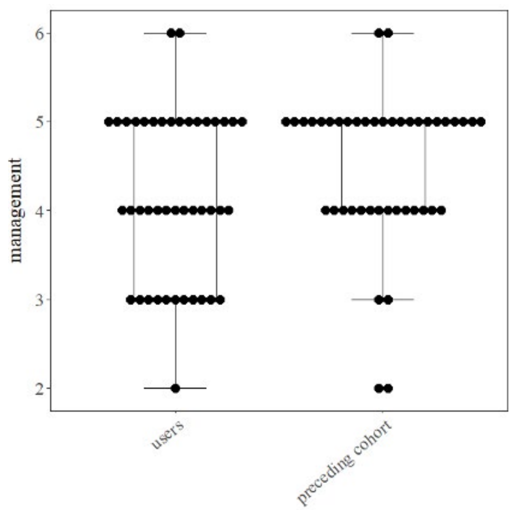


Figure 3: Distribution of management values for users and students (non-users) from the preceding cohort.



H3: Achievement of PPB-users compared to non-users from the same cohort

Information on variables used for matching and outcome variables (grades or management) are available for 51 PPB-users and 10 non-users (H3a, grades) and 43 PPB-users and 9 non-users (H3b, management). The matching process aimed at a 3:1 ratio of users to non-users and yielded a sample of 40 (H3a) and 36 (H3b) participants, consisting of 30 (or 27) users and 10 (or 9) non-users (for sample statistics of matched samples see Appendix B). Matched samples were tested for group differences in grades and management (Table 4).

Table 4: Descriptive values of grade and management and results of the one-tailed U-tests for matched samples of users and non-users from the same cohort (2018).

	Users			Non-users					
	n	*Mdn*	*M (SD)*	*n*	*Mdn*	*M (SD)*	*U*	*p*	*r*
Grade[a]	30	1.7	1.86 (1.00)	10	2.3	2.53 (1.11)	221.5	.012	–.36
Management	27	4.0	4.07 (0.96)	9	4.0	2.89 (1.36)	63.5	.014	–.37

Note. *Mdn* = median, *r* = effect size. Users were matched to non-users with respect to gender, high school graduation year, math assessment, interest in and perceived relevance of statistics.
[a]higher grades denote worse exam performance.

Mann-Whitney-U-tests revealed that matched users and non-users differ significantly regarding grade ($U = 221.5$, $p = .012$, $r = -.36$) and management ($U = 63.5$, $p = .014$, $r = -.37$). As hypothesized, users achieved lower (better) grades (H3a) and reported better management of requirements in statistics (H3b). Figures 4 and 5 show the distribution of values for grade and management for the matched groups.

Figure 4: Distribution of grade values for users and non-users from the same cohort (2018).

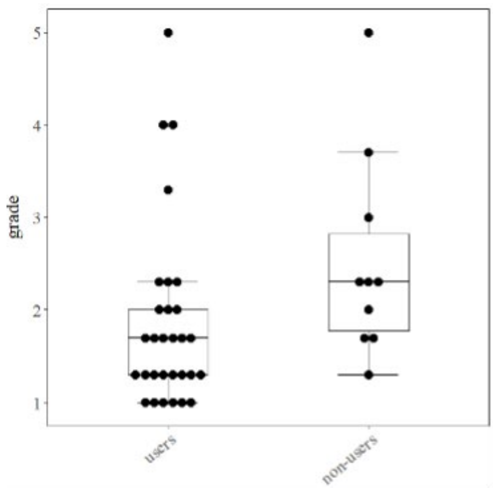
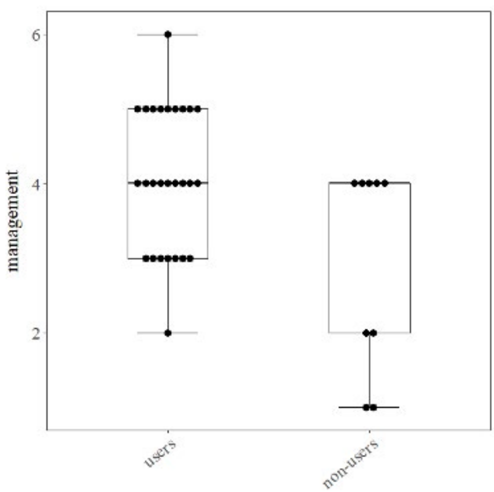


Figure 5: Distribution of management values for users and non-users from the same cohort (2018).



Discussion

Results indicate that the main goals of our study were reached. First, we found that most surveyed psychology students used our PPB, evaluated it positively and judged it as helpful (RQ1a). Also, we observed an excellent rating of structure and design (RQ1b). Task difficulty was perceived as adequate and nearly half of the users were also satisfied with the amount of content covered by the PPB, but even more would have liked additional content to be covered (RQ1c). These findings support the positive evaluation, too, since many students seem to get along well with the tasks and wish for more content to work on with the PPB.

Second, non-users differ from users regarding characteristics at study entry. Non-users are probably not as familiar with mathematical content as users are, as users show higher initial skills in mathematics (H1a), are younger and graduated from school more recently (H1c). However, despite a tendency to lower descriptive values in non-users, there were no significant group differences for interest in or relevance of statistics, and thus H1b must be rejected. These results go in line with findings of Bebermeier et al. (2019), showing that students with initially higher mathematical skills use voluntary learning tools in statistics more often.

Third and last, we revealed achievement differences between users and comparable non-users from the preceding (H2) and the same cohort (H3). PPB-users achieved better grades (H2a, H3a) and reported higher management of requirements compared to non-users from the same cohort (H3b), which confirms previous findings showing that using learning support can improve objective measures of academic achievement as well as the subjective improvement of academic success (Bebermeier et al., 2019; Matthews et al., 2013). Our results also go in line with findings of Austerschmidt and Bebermeier (2019), indicating that enhancing the support from one cohort to another can have diverging effects on objective or subjective learning outcomes.

We further conclude that most students accept our newly developed learning tool and work with it, indicating that further engagement with the content was achieved. The PPB seems to be suitable for students at the beginning of their statistics education and thus suitable for teaching basic knowledge without being too demanding (Cromley et al., 2020; Mayer et al., 2005), as shown by the results of RQ1. We showed benefits for students who used the PPB (H2 & H3) and conclude that the alternative representation and active engagement with the subject matter probably enhances learning and understanding (Ainsworth, 2014; Herrmann, 2009).

Even though the PPB is easily available to the students, time and effort must be invested in working with it. Results of H1 indicate that students who probably have particular problems with the course and its content did use the PPB less likely. However, the analyses based on matched, most comparable students from our samples, showed that both, high and low competent (operationalised by scores on the math assessment) students, can benefit from the PPB. Since it shall promote basic skills and enhance motivation, the latter group should therefore be particularly encouraged to use the offer or alternative learning support more suitable for them. Results show that non-users differ from users in key characteristics such as graduation year and solved math tasks, but they are not significantly less interested in statistics or consider it less relevant. Since the non-significant effects may also be a result of low statistical power due to the small number of non-users, further research should investigate the relationship between statistics affinity (i.e. interest in statistics) and use of learning tools more closely in order to derive recommendations. On the one hand, the differences in means of relevance and interest between users and non-users, and the small to medium effect sizes for the respective tests (see Table 2) indicate that non-users might show less affinity to statistics, but to a smaller extent than expected. If, on the other hand, non-users show comparable affinity for statistics as users, stressing the importance of statistics and the availability of additional learning tools might not be a determining factor in increasing the use of such support services. Instead, further research is necessary to investigate what might motivate students with such characteristics (i.e. older-than-average and/or less math skills) to use the PPB and comparable tools – or investigate what prevents them from doing so. A relevant factor might be the ability to manage the demands the statistics course poses, which is lower in non-users than in users. For example, some students might have less resources or be more easily overwhelmed by the high demands and therefore not use the PPB to reduce their workload.

But there are also some critical issues of our study to be addressed. Because of the small number of non-users, results should be interpreted with caution. Additionally, we observed a moderate dropout rate: not all students addressed by the course participated in our study and especially a significant number of students did not answer the second survey. For our analyses, data was available for two-thirds of the students. We conducted additional analyses, which showed that students who completed both surveys perform better in the math assessment, state a higher perceived relevance of statistics and perform better in the exam than students who only answered the first survey. Also, average grades achieved in our sample are slightly better than the average grade of the whole cohort, including non-participants. Thus, characteristics of dropouts probably correlate with measures from the survey, since more competent and more motivated students seem to attend the lecture more regularly, complete the course successfully, and therefore are more likely to answer the second survey. However, this phenomenon is commonly known in research investigating academic success (Larsen et al., 2013).

It is problematic to deprive a group of students of an offer for ethical reasons, and not in accordance with academic teaching in general either to force or to impede students to use an offer. In our study, this led to a small number of non-users which makes reliable inferences challenging. Matched samples of PPB-users and non-users from the same cohort still are not entirely comparable regarding matched variables and additionally could differ in further characteristics not considered in the present study. This may have at least partially caused differences in achievement. We therefore included the comparison with the preceding cohort before the introduction of the PPB for additional evidence. However, to figure out the incremental value of the PPB more precisely, more data of non-users (and users) from the following cohorts or samples of users and non-users from other universities should be investigated and could provide further insight.

Subsequent investigations should try to increase survey coverage rates and in this way reach (more) students with problems. Further, future studies should investigate differential effects of usage on achievement for low and high competent students, for the PPB but also for other support services. Thus, students can systematically be addressed by offers that suit them best and promise success. Further research is needed to figure out why especially low competent students rarely use the PPB and to further examine if it is as suitable and promising for them to enhance achievement. However, the small sample of non-users in the evaluation cohort may not be representative, and a very special group of students and thus it is difficult to draw conclusions based on this data.

For teachers, we recommend pointing out the relevance of statistics to enhance engagement and motivation, especially in low competent students. Recommending specific learning tools for certain levels of knowledge or specific goals to be achieved (e.g. repeating, deepening content or gaining basic skills) can help to meet students' needs. In this regard, it has been shown that within statistics education, assessment exercises with constructive feedback are effective for supporting students individually (Garfield & Ben-Zvi, 2007; Lovett & Greenhouse, 2000). Further, support services like the PPB should be embedded into the course, for example, by providing them within the lecture or covering exam relevant content. Probably, this approach contributed to the observed high usage and the positive evaluation. Additionally, it should be even more promising to provide learning materials step-by-step accompanying the lecture which we evaluate at the present while providing a revised version of the PPB. Finally, the PPB turns out to be a useful complement to lecture attendance and other learning tools which can further enhance learning success and is well accepted by students. As it is in the nature of academic teaching, a lot of learning opportunities offered are not obligatory. Thus, this evaluation provides ecologically valid empirical evidence. Students using the PPB may have spent additional time on learning, which naturally should lead to better grades, but this very fact was intended by the learning tool. Yet, it is essential to examine and (if possible) control potential confounding variables in future studies.

Authors

Kim L. Austerschmidt
Bielefeld University, Department of Psychology, Germany.

Denise Kerkhoff
Bielefeld University, Department of Psychology, Germany.
Konstanz University, Department of Psychology, Germany

Sarah Bebermeier
Leibniz University Hannover, Department of Psychology, Germany

Anne Hagemann
Society for Epilepsy Research, Bielefeld, Germany

References

Aberson, C.L., Berger, D.E., Healy, M.R. et al.. (2000). Evaluation of an interactive tutorial for teaching the central limit theorem. *Teaching of Psychology, 27,* 289-291. https://doi.org/10.1207/S15328023TOP2704_08

Ainsworth, S. (2014). The multiple representation principle in multimedia learning. In R.E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp.464–487). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369

American Psychological Association (2013). *APA guidelines for the undergraduate psychology major: Version 2.0.* American Psychological Association. http://www.apa.org/ed/precollege/undergrad/index.aspx

Austerschmidt, K.L. & Bebermeier, S. (2020). 'Richtig Einsteigen in Statistik' – Nutzung und Nutzen eines mathematischen Vorkurses im Psychologiestudium. ['Start properly in statistics' – Usage and usefulness of a preparatory course in mathematics for psychology students]. *Psychologie in Erziehung und Unterricht* [Psychology in Education and Teaching], *67*(1), 47–60. https://doi.org/10.2378/peu2020.art05d

Austerschmidt, K.L. & Bebermeier, S. (2019). Flexible Unterstützungsangebote in Statistik: Implementation und Effekte auf Studienerfolg [Implementation and Effects of Flexible Support Services on Student Achievements in Statistics]. *Zeitschrift für Hochschulentwicklung* [Journal of Higher Education Development], *14*(3), 137–155. https://doi.org/10.3217/zfhe-14-03/09

Bebermeier, S., Nussbeck, F.W., & Austerschmidt, K.L. (2019). The impact of students' skills on the use of learning support and effects on exam performance in a psychology students' statistics course. *Scholarship of Teaching and Learning Psychology.* https://doi.org/10.1037/stl0000170

Bebermeier, S., Nussbeck, F.W. & Ontrup, G. (2015). 'Dear Fresher' – How online questionnaires can improve learning and teaching statistics. *Psychology Learning & Teaching, 14,* 147-157. https://doi.org/10.1177/1475725715578563

Bobek, E. & Tversky, B. (2016). Creating visual explanations improves learning. *Cognitive Research: Principles and Implications, 27*(1). https://doi.org/10.1186/s41235-016-0031-6

British Psychological Society (2019). *Standards for the accreditation of undergraduate, conversion and integrated Masters programmes in psychology.* The British Psychological Society. https://www.sheffield.ac.uk/polopoly_fs/1.304874!/file/5_1_BPS_Accreditation_Guidance_2012_PrintPages1_43_2012.pdf

Budé, L., Van De Wiel, M.W. J., Imbos, T., et al. (2007). Students' achievements in a statistics course in relation to motivational aspects and study behaviour. *Statistics Education Research Journal, 6*(1), 5–21. http://www.stat.auckland.ac.nz/~iase/serj/SERJ6(1)_Bude.pdf

Carpenter, T.P. & Kirk, R.E. (2017). Are psychology students getting worse at math?: Trends in the math skills of psychology statistics students across 21 years. *Educational Studies, 43*(3), 282–295. https://doi.org/10.1080/03055698.2016.1277132

Chance, B., Ben-Zvi, D., Garfield, J. & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education Journal, 1*(1). https://escholarship.org/uc/item/8sd2t4rr

Chiesi, F., & Primi, C. (2010). Cognitive and non-cognitive factors related to students statistics achievement. *Statistics Education Research Journal, 9*(1), 6-26. www.researchgate.net/publication/290485990_Cognitive_and_non-cognitive_factors_related_to_students%27_statistics_achievement

Chew, P.K.H. & Dillon, D.B. (2014). Statistics anxiety update: Refining the construct and recommendations for a new research agenda. *Psychological Science, 9,* 196–208. https://doi.org/10.1177/1745691613518077

Cromley, J.G., Du, Y. & Dane, A.P. (2020). Drawing-to-Learn: Does meta-analysis show differences between technology-based drawing and paper-and-pencil drawing?. *Journal of Science Education and Technology, 29,* 216–229. https://doi.org/10.1007/s10956-019-09807-6

Dempster, M. & McCorry, N.K. (2009). The role of previous experience and attitudes toward statistics in statistics assessment outcomes among undergraduate psychology students. *Journal of Statistics Education, 17*(2), 1–7. https://doi.org/10.1080/10691898.2009.11889515

Doyle, D.A. (2017). Ugh… Statistics! College students' attitudes and perceptions toward statistics. *Honors in the Major Theses*, 165. University of Central Florida.

Eccles, J.S. & Wigfield A. (2002). Motivational beliefs, values, and goals. *Annual Review of Pychology, 53*, 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Field, A.P. (2014). Skills in mathematics and statistics in psychology and tackling transition. *The Higher Education Academy STEM project series.* https://www.heacademy.ac.uk/system/files/resources/tt_maths_psychology.pdf

Finney, S.J. & Schraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary educational psychology, 28*(2), 161–186. https://doi.org/10.1016/S0361-476X(02)00015-2

Fiorella, L. & Zhang, Q. (2018). Drawing boundary conditions for learning-by-drawing. *Educational Psychology Review, 30*, 1115–1137. https://doi.org/10.1007/s10648-018-9444-8

Fonteye, L., De Fruyt, F., Dewulf, N. et al. (2015). Basic mathematics test predicts statistics achievement and overall first year academic success. *European Journal of Psychology Education, 30*, 95–118. https://doi.org/10.1007/s10212-014-0230-9

GAISE College Report ASA Revision Committee (2016). *Guidelines for Assessment and Instruction in statistics education college report 2016.* www.amstat.org/education/gaise

Garfield, J. & Ben-Zvi, D. (2007). How students learn statistics revisited: a current review of research on teaching and learning statistics. *International Statistical Review, 75*, 372–396, https://doi.org/10.1111/j.1751-5823.2007.00029.x

Garfield, J. & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice.* Springer.

Garfield, J. & Ben-Zvi, D. (2009). Helping students develop statistical reasoning: Implementing a statistical reasoning learning environment. *Teaching Statistics, 31*(3), 72–77. https://doi.org/10.1111/j.1467-9639.2009.00363.x

Goode, C.T., Lamoreaux, M., Atchison, K.J. et al. (2018). Quantitative skills, critical thinking, and writing mechanics in blended versus face-to-face versions of a research methods and statistics course. *Teaching of Psychology, 45*, 124–131. https://doi.org/10.1177/0098628318762873

Herrmann, U. (Ed.) (2009). *Neurodidaktik: Grundlagen und Vorschläge für gehirngerechtes Lehren und Lernen* [Neurodidactics: Basics and suggestions for brain appropriate teaching and learning] (2nd Edn.). Beltz.

Heublein, U., Ebert, J., Hutzsch, C. et al. (2017). *Zwischen Studienerwartungen und Studienwirklichkeit. Ursachen des Studienabbruchs, beruflicher Verbleib der Studienabbrecherinnen und Studienabbrecher und Entwicklung der Studienabbruchquote an deutschen Hochschulen* [Between study expectations and study reality. Causes of study dropout, occupational fate of study dropouts and the development of the dropout rate at German universities]. Hannover: Deutsches Zentrum für Hochschul- und Wissenschaftsforschung. www.dzhw.eu/pdf/pub_fh/fh-201701.pdf

Heublein, U., Hutzsch, C., Schreiber, J., Sommer, D. & Besuch, G. (2010). *Ursachen des Studienabbruchs in Bachelor- und in herkömmlichen Studiengängen: Ergebnisse einer bundesweiten Befragung von Exmatrikulierten des Studienjahres 2007/08.* [Causes of dropping out in bachelor's and other study programs: Results of a nationwide survey of exmatriculated students in the academic year 2007/2008.] Hannover: Hochschul-Informations-System. http://ids.hof.uni-halle.de/documents/t1944.pdf

Ho, D.E., Imai, K., King, G. & Stuart, E.A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software, 42*(8), 1–28. www.jstatsoft.org/v42/i08/

Holmes, J.D. (2014). Undergraduate psychology's scientific identity dilemma: Student and instructor interests and attitudes. *Teaching of Psychology, 41*, 104–109. https://doi.org/0.1177/0098628314530339

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. https://iase-web.org/documents/SERJ/SERJ11(2)_Hood.pdf

Kennett, D., Young, A.M. & Catanzaro, M. (2009). Variables contributing to academic success in an intermediate statistics course: The importance of learned resourcefulness. *Educational Psychology, 29*, 815–830. https://doi.org/10.1080/01443410903305401

Larsen, M.S., Kornbeck, K.P., Kristensen, R. et al. (2013). *Dropout phenomena at universities: What is dropout? Why does dropout occur? What can be done by the universities to prevent or reduce it? A systematic review.* Copenhagen: Danish Clearinghouse for educational research, Department of Education, Aarhus University. https://edu.au.dk/fileadmin/edu/Udgivelser/Clearinghouse/Review/Evidence_on_dropout_from_universities_technical_report_May_2013.pdf

Leutner, D. & Schmeck, A. (2014). The generative drawing principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning (Cambridge Handbooks in Psychology)* (pp.433–448). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369

Leino, R., Doering, A., Gardner, M. & Cartwright, T. (2019). Psychology students' journey through the research methods curriculum: Insights from one qualitative and one longitudinal study. *ECER 2019, Network 22: Research in Higher Education.* https://eera-ecer.de/ecer-programmes/conference/24/contribution/48212/

Lovett, M. & Greenhouse, J. (2000). Applying cognitive theory to statistics instruction. *The American Statistician, 54*, 196–206. https://doi.org/10.1080/00031305.2000.10474545

Macher, D., Paechter, M., Papousek, I. & Ruggeri, K. (2012). Statistics anxiety, trait anxiety, learning behavior, and academic performance. *European Journal of Psychology of Education, 27*(4), 483–498. https://doi.org/10.1007/s10212-011-0090-5

Matthews, J., Croft, T., Lawson, D. & Waller, D. (2013). Evaluation of mathematics support centres: a literature review. *Teaching Mathematics and its Applications, 32*, 173–190. https://doi.org/10.1093/teamat/hrt013

Mayer, R.E., Hegarty, M., Mayer, S. & Campbell, J. (2005). When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction. *Journal of Experimental Psychology: Applied, 11*(4), 256–265. https://doi.org/10.1037/1076-898X.11.4.256

Meter, van P. & Garner, J. (2005). The promise and practice of learner-generated drawing: Literature review and synthesis. *Educational Psychology Review, 17*(4) 285–325. https://doi.org/10.1007/s10648-005-8136-3

Nückles, M. & Wittwer, J. (2014). Lernen und Wissenserwerb [Learning and knowledge acquisition]. In T. Seidl & A. Krapp (Eds.), *Pädagogische Psychologie* [Educational Psychology] (6th edn.) (pp.224–322). Beltz.

Onwuegbuzie, A.J. & Wilson, V.A. (2003). Statistics anxiety: Nature, etiology, antecedents, effects, and treatments – a comprehensive review of the literature. *Teaching in Higher Education, 8*(2), 195–209. https://doi.org/10.1080/1356251032000052447

R Core Team (2016). *R: A language and environment for statistical computing.* Vienna: R foundation for statistical computing. www.R-project.org/

Rajecki, D.W., Appleby, D., Williams, C.C. et al. (2005). Statistics can wait: Career plans activity and course preferences of American psychology undergraduates. *Psychology Learning and Teaching, 4*, 83–89. https://doi.org/10.2304/plat.2004.4.2.83

Reiß, S., Mildner, D., Nagler, H. & Schweizer, K. (2011). Teilnahmeinteresse an universitären Vorlesungen in Abhängigkeit vom Lehrveranstaltungsinhalt und Erfassungszeitpunkt [Participation interest in university lectures in relation to lecture content and recording time]. *Psychologiedidaktik und Evaluation VIII* [Psychology Didactics and Evaluation VIII], 151-161.

Ruggeri, K., Díaz, C., Kelley, K. et al. (2008). International issues in education. *Psychology Teaching Review, 14*, 65–74.

Schau, C. (2003). *Students' attitudes: The 'other' important outcome in statistics education.* www.semanticscholar.org/paper/STUDENTS'-ATTITUDES%3A-THE-%22OTHER%22-IMPORTANT-OUTCOME-Schau/114dc36e820be7fa827def035b71a41a68fe510d

Schacht, S. & Stewart, B.J. (1990). What's funny about statistics? A technique for reducing student anxiety. *Teaching Sociology, 18*, 52–56. https://doi.org/10.2307/1318231

Schwamborn, A., Mayer, R.E., Thillmann, H. et al. (2010). Drawing as a generative activity and drawing as a prognostic activity. *Journal of Educational Psychology, 102*, 872–879. https://doi.org/10.1037/a0019640

Stoloff, M., McCarthy, M., Keller, L. et al. (2009). The undergraduate psychology major: An examination of structure and sequence. *Teaching of Psychology, 37*, 4–15. https://doi.org/10.1080/00986280903426274

Sutherland, K.A. (2017). Constructions of success in academia: An early career perspective. *Studies in Higher Education, 42*(4), 743–759. www.tandfonline.com/doi/full/10.1080/03075079.2015.1072150

Tremblay, P.F., Gardner, R.C. & Heipel, G. (2000). A model of the relationships among measures of affect, aptitude, and performance in introductory statistics. *Canadian Journal of Behavioral Science, 32*, 40—48. https://doi.org/10.1037/h0087

Vanhoof, S., Castro Sotos, A.E., Onghena, P. et al. (2006). Attitudes toward statistics and their relationship with short- and long-term exam results. *Journal of Statistics Education, 14*(3). https://doi.org/10.1080/10691898.2006.11910588

Veilleux, J.C. & Chapman, K.M. (2017). Development of a research methods and statistics concept inventory. *Teaching of Psychology, 44*, 203–211. https://doi.org/0.1177/0098628317711287

Zieffler, J.G., Alt, S., Dupuis, D., Holleque, K. & Chang, B. (2008). What does research suggest about the teaching and learning of introductory statistics at the college level? A review of the literature. *Journal of Statistics Education, 16*(2). https://doi.org/10.1080/10691898.2008.11889566

Appendix A

Students' Characteristics of Matched Samples (H2).

	H2a (grade)				H2b (management)			
	Users (*N* = 51)		Preceding cohort (*N* = 51)		Users (*N* = 43)		Preceding cohort (*N* = 43)	
	Mdn	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*	*M (SD)*
Math assessment	12.0	12.10 (3.23)	13.0	12.29 (3.26)	13.0	12.28 (3.50)	12.0	11.98 (3.47)
Interest	3.0	3.49 (0.99)	3.0	3.29 (0.99)	3.0	3.42 (0.98)	3.0	3.33 (0.97)
Relevance	4.7	4.66 (0.68)	4.7	4.90 (0.83)	4.7	4.83 (0.64)	4.7	4.68 (0.75)
Graduation[a]	1.0	2.63 (4.55)	1.0	1.51 (2.27)	1.0	2.51 (4.88)	1.0	1.51 (2.35)
		f (%)		*f* (%)		*f* (%)		*f* (%)
Gender (female)		44 (84%)		43 (86%)		38 (88%)		26 (84%)

Note. *Mdn* = median, f = frequency.
[a]years since graduation.

Appendix B

Students' Characteristics of Matched Samples (H3).

	H3a (grade)				H3b (management)			
	Users (*N* = 30)		Non-users (*N* = 10)		Users (*N* = 27)		Non-users (*N* = 9)	
	Mdn	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*	*M (SD)*
Math assessment	11.0	10.47 (2.90)	10.0	9.60 (3.06)	11.0	10.52 (3.03)	10.0	9.89 (2.67)
Interest	3.0	3.13 (0.77)	3.0	3.10 (1.10)	3.0	3.22 (0.93)	3.0	3.00 (1.12)
Relevance	4.7	4.72 (0.68)	4.5	4.30 (1.28)	4.7	4.64 (0.61)	5.0	4.33 (1.35)
Graduation[a]	1.0	3.33 (6.06)	4.0	3.67 (2.00)	1.0	3.48 (5.98)	4.0	3.67 (2.00)
		f (%)		*f* (%)		*f* (%)		*f* (%)
Gender (female)		23 (77%)		8 (80%)		24 (89%)		8 (89%)

Note. *Mdn* = median, *f* = frequency.
[a]years since graduation.