SUMS: A flexible approach to the teaching and learning of statistics

MAXINE V. SWINGLER[1] AND PAUL BISHOP
University of Glasgow, UK

KEVIN M. SWINGLER
University of Stirling, UK

Statistical Understanding Made Simple (SUMS) (http://www.gla.ac.uk/sums) is an online statistics tutorial generator which covers basic statistical concepts and focuses on applying these concepts to real data. Instructors can upload their own data to the site, and students choose the level at which they want to work from one of three levels. SUMS adopts a teaching approach where each concept is explained, the student explores the concept with an interactive game, and then applies their knowledge to a concrete example (using data provided by the instructor). SUMS was evaluated to determine its effectiveness as a stand-alone resource for psychology students and as a resource to support a psychology laboratory class, using a measure of statistics self-efficacy (based on Finney & Schraw, 2003) and a statistics comprehension test. Results of the evaluation showed that overall, SUMS had a positive impact on students' statistics comprehension and self-efficacy.

INTRODUCTION

Context

Online interactive tutorials are commonly used to support the teaching of introductory statistics (e.g., Berger, 2008; Dinov, 2002; Lane, 2008; West, 1996), and when evaluated, these resources improve students' performance in and understanding of statistics (Aberson, Berger, Healey, Kyle, & Romero, 2000; Aberson, Berger, Healey, & Romero, 2003; Dinov, Sanchez, & Christou, 2008; Utts, Sommer, Acredolo, Mahler, & Matthews, 2003). However, applying these resources to the teaching of statistics in introductory psychology raises two issues:

Firstly, students enter undergraduate psychology courses with varying levels of numeracy (Mulhern & Wylie, 2004, 2006; Tariq, 2002) and of anxiety about statistics (Bologlu & Zelhart, 2003; Onwuegbuzie & Wilson, 2003). The online statistics tutorials often require prior knowledge of basic mathematics and statistics (e.g., use of mathematical symbols and formula, interpretation of graphs), which students may or may not have. Secondly, it is considered important to illustrate statistical concepts using data from students' own subject of study (Ben-Zvi, 2000; Garfield, 1995). However, as these resources are often targeted at students of various subjects in the sciences and social sciences (including psychology) who undertake an introductory statistics course (Aberson et al., 2000, 2003) outside their major subject, this can often lead to psychology students learning to use statistics on unfamiliar data. There is a need for a flexible resource that makes allowances for students' prior knowledge and allows for the use of familiar data.

Aims of study

The aim of the present study was to address these issues by developing a flexible online statistics resource that enabled:

- instructors to upload data relevant to their course and familiar to their students, and
- students to choose the level at which they wished to learn about statistical concepts.

The evaluation approach focused on whether the online resource improved psychology students' confidence (self-efficacy) in statistics and their comprehension of basic statistical concepts.

Description of resource

SUMS (Statistical Understanding Made Simple) is an online resource for instructors and students of introductory level statistics (Swingler, 2008). The site, which is freely available for use by all, automatically constructs bespoke tutorials based on data uploaded by instructors. These tutorials are built using HTML, Java applets and Javascript to produce an online exercise based on the uploaded data. Each tutorial is hosted on the SUMS website (http://www.gla.ac.uk/sums) and is

[1]Correspondence should be addressed to the first author at: Department of Psychology, University of Glasgow G12 8QQ. Email: m.swingler@psy.gla.ac.uk

made publicly available. Instructors may also download a zipped file allowing them to host the tutorial elsewhere. The process is summarised in Figure 1.

Pedagogical approach

The pedagogical approach utilised in SUMS is outlined in Figure 2. Every topic has a similar structure and is covered at three progressive levels of learning. The idea behind the structure was to encourage students to actively engage in constructing their own knowledge of statistics by first exploring statistical concepts and then applying these concepts to data presented in a familiar context (Garfield, 1995). Furthermore, the student engages in this process at their chosen level and can progress through topics and levels at their own pace. The focus of each level is summarised in Figure 3. Level 1 focuses on the understanding of statistical concepts, Level 2 on the use of formulae, and Level 3 on the theory behind statistical concepts.

The interactive element of the tutorial was based on Java applets that allowed active manipulation of data, use of graphics to explore and visualise data, online calculators and consistent feedback. These techniques are widely used in current online statistics resources (e.g., Dinov et al., 2008; Utts et al., 2001) and in the teaching of statistics (Ben-Zvi, 2000; Garfield, 1995). The department of statistics at the University of Glasgow and established statistics textbooks (Dancey & Reidy, 2007; Miller, 1984) were consulted for guidance on statistical content of the tutorial.

Evaluation approach

The SUMS tutorial was evaluated using two samples of first-year undergraduate psychology students in two contexts. In Context 1, the SUMS tutorial was used as a stand-alone resource, and the tutorial was evaluated by a group of students who had expressed anxiety about the statistics element of their course. In Context 2, the online tutorial supported a practical laboratory and report in social psychology. Students were encouraged to use the tutorial but its use was not compulsory.

In both contexts, students were asked to complete Level 1 of the tutorial only (understanding of statistical concepts and interpreting results), because this corresponded to the level of statistics taught in first-year psychology. All students taking the course received basic instruction on how to use statistical software to analyse and interpret their data in their laboratory class, but received no additional tuition in statistics.

The resource was evaluated using two measures: a measure of students' self-efficacy in statistics and a multiple-choice test assessing comprehension of basic statistical concepts. Self-efficacy is defined as the confidence a student has in completing a specific task, and has been shown to be a good predictor of

Figure 1 Summary of the SUMS tutorial generator from a student's and an instructor's perspective

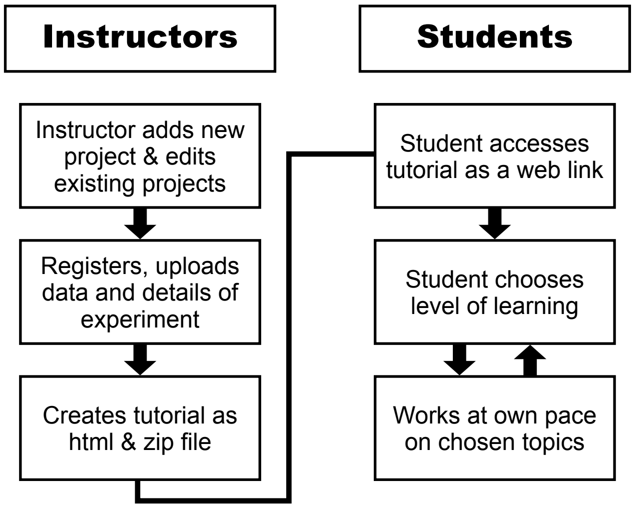


Figure 2 Summary of the pedagogical approach

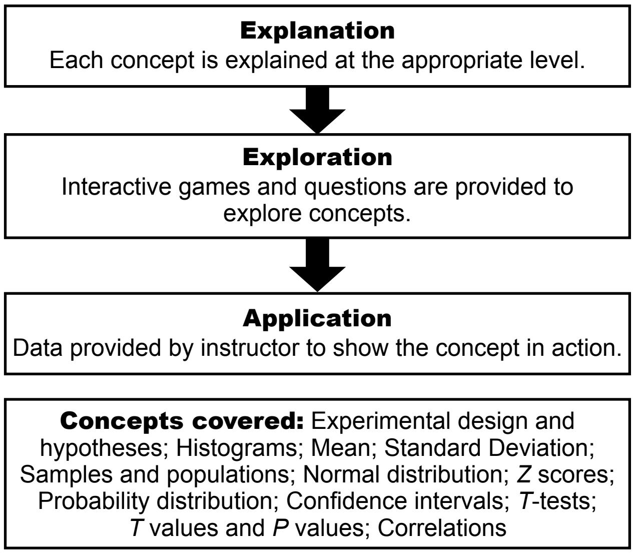
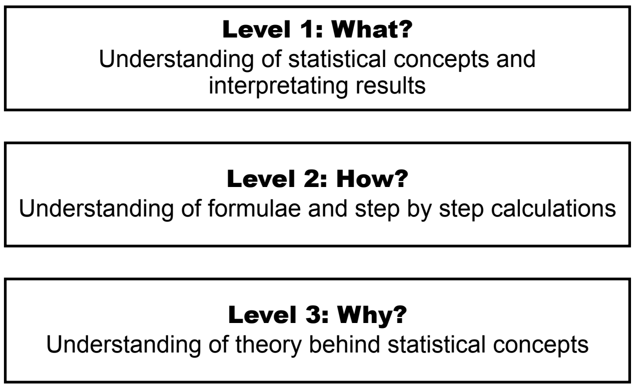


Figure 3 Summary of the focus of each level of learning



performance in a variety of academic contexts (Bandura, 1997; Pajares, 1996) including mathematics (Kranzler & Pajares, 1997; Pajares & Miller, 1994). Current self-efficacy in performing statistical tasks is positively related to statistics performance and negatively related to statistics anxiety, and students' self-efficacy is improved after completion of a statistics course (Finney & Schraw, 2003; Lane, Lane, & Hall, 2004). The present evaluation used self-efficacy statements similar to those used by Finney and Schraw (2003), but adapted the tasks to the learning outcomes of the online tutorial.

It was hypothesised that the use of the tutorial in either evaluation context would have a positive impact on self-efficacy and comprehension.

METHOD

Context 1: SUMS Used as a Stand-Alone Resource

Participants

Participants were 24 first-year psychology students at the University of Glasgow (23 female, 1 male), between the ages of 16 and 30, and recruited via a notice on the Psychology Department virtual learning environment (VLE) offering additional support to students struggling with statistics. Students received a course credit for participation.

Design

A pre- and postintervention design was used, with all participants completing the same tutorial topics and evaluation measures.

Measures

Statistics self-efficacy. The pre- and posttutorial self-efficacy questionnaire consisted of 10 self-efficacy statements on current ability to complete a number of statistical tasks related to 10 concepts covered by Level 1 of the tutorial. Participants rated each task using a Likert-type scale ranging from 1 (*no confidence at all*) to 6 (*complete confidence*) for each task (see Appendix A).

Statistical comprehension. The multiple-choice test consisted of 10 questions (based on the tutorial concepts) with four answer options. To avoid effects of question familiarity, two equivalent versions of the multiple-choice test were used. Order of presentation (pre- or posttutorial) of the two versions was counterbalanced. Students' comments on the tutorial and any technical problems encountered were recorded.

Procedure

The software generated a freely available online tutorial based on the famous Bobo Doll psychology experiment (Bandura, Ross, & Ross, 1961), which is familiar to most psychology undergraduates. Participants were tested as a group and the evaluation was run over one session in the university computer laboratories. Participants completed the pretutorial measures, worked at their own pace through 10 topics of the online tutorial at Level 1 (experimental design, hypotheses, histograms, measures of central tendency, standard deviation, samples and populations, normal distributions, choosing a *t* test, independent groups *t* test, *p* values and *t* values), then completed the posttutorial measures.

Context 2: SUMS Used to Support Laboratory Teaching and Coursework

Participants

The participants were Level 1 psychology students. Eighty students completed Questionnaires 1 (pretest) and 2 (posttest), and 181 students completed only Questionnaire 2 (posttest). Ages ranged from 17 to 38 years old. The low number of students who completed both Questionnaires 1 and 2 was due to the fact that a limited sample of laboratory classes received Questionnaire 1 (pretest), whereas a larger sample of students was available to complete Questionnaire 2 (posttest).

Design

The students were assessed on their self-efficacy at two points. The first was before the tutorial was available, during the laboratory class (pretest). The second point was immediately after the coursework related to the laboratory was handed in to the department for marking (posttest). Also, at the second point of testing the students were assessed on comprehension of several statistical concepts.

The students were divided into two self-selected groups, those who chose to complete the tutorial and those who did not. The decision to use the tutorial was a purely voluntary one and the students could use the online tutorial as often as they liked.

Measures and procedure

The class instructor uploaded the data and details of the experiment to SUMS, to generate a freely available online tutorial based on the laboratory class. The tutorial was then mounted on the department's VLE and its presence advertised to the students on the Level 1 psychology course.

Participants were asked to fill out two questionnaires, one at the pretest and one at the posttest. The first was a set of six self-efficacy tasks (see Appendix A) concerning statistical concepts related to learning outcomes of the laboratory class (see Table 2 in results for tasks). As in Context 1, students rated their confidence in completing each of the tasks on a six-point Likert scale.

The second questionnaire repeated the self-efficacy tasks and included a comprehension test (eight statements which the student had to rate as true or

Table 2
Means and standard deviations (in brackets) of the pre- and posttest self-efficacy task ratings, for users and nonusers of the online tutorial

Self-efficacy task	Users Pretest	Users Posttest	Nonusers Pretest	Nonusers Posttest
1	3.26 (1.14)	4.26 (1.14)	3.60 (1.07)	4.00 (1.15)
2	3.53 (1.44)	4.26 (1.37)	3.35 (1.19)	4.33 (0.94)
3	3.00 (1.35)	3.71 (1.17)	2.81 (1.12)	3.47 (1.24)
4	3.62 (1.81)	4.24 (1.16)	3.56 (1.29)	3.93 (1.44)
5	4.74 (1.24)	5.29 (0.87)	4.95 (0.95)	5.23 (1.02)
6	3.59 (1.84)	4.68 (1.12)	4.12 (1.12)	4.58 (1.09)
Overall	3.55 (0.69)	4.38 (0.76)	3.75 (0.68)	4.28 (0.69)

false) on the same concepts used in the self-efficacy questionnaire. Students indicated if they had used the online tutorial or not.

RESULTS

Context 1: SUMS Used as a Stand-Alone Resource

Self-efficacy

In order to investigate the effect of the tutorial on self-efficacy, a repeated measures MANOVA with one within-subjects factor of time administered (pre- and posttutorial) was conducted with scores on each of the 10 self-efficacy tasks as dependent variables. The MANOVA found a multivariate difference between pre- and posttutorial self-efficacy tasks ($F(1, 23) = 6.88$, $p = .001$, Wilks Lambda = .17). As the dependent variables were uncorrelated, separate univariate paired samples *t* tests were conducted on pre- and posttutorial self-efficacy tasks (the α was set at .005 following a Bonferroni adjustment for familywise error). As shown in Table 1, all self-efficacy tasks contributed to the multivariate difference between pre- and posttutorial self-efficacy tasks (all $p < .005$) apart from Task 4 ($p < .008$).

To investigate the effect of self-efficacy task independently of the tutorial, paired samples *t* tests compared self-efficacy ratings between each task in the pre- and posttutorial conditions (the α was set at .001 following a Bonferroni adjustment for familywise error). In the pretutorial condition, mean self-efficacy ratings were highest for Task 2, followed by Tasks 1 and 4, and lowest for Task 8 (all $p < .001$). In the posttutorial condition, self-efficacy ratings were highest for Task 1, followed by Tasks 7 and 2 (all $p < .001$).

Table 1
Means and standard deviations (in brackets) of self-efficacy ratings for each task before and after participation in the online tutorial

Self-efficacy task	Pretutorial	Posttutorial
1	3.58 (1.42)	5.04 (1.95)
2	3.75 (1.38)	4.92 (1.02)
3	2.62 (1.33)	4.38 (1.10)
4	3.70 (1.44)	4.71 (1.33)
5	2.16 (1.20)	4.25 (1.19)
6	2.70 (1.25)	4.00 (1.22)
7	2.37(1.60)	4.88 (1.36)
8	2.00 (1.21)	4.21 (1.44)
9	2.46 (1.47)	3.70 (1.51)
10	2.37 (1.38)	4.00 (1.38)
Overall	2.77 (1.47)	4.41 (1.31)

Comprehension

Mean correct responses (maximum = 10) for pre- and posttutorial multiple- choice tests were 4.46 ($SD = 1.76$) pretutorial and 7.48 ($SD = 1.48$) posttutorial. A paired samples *t* test comparing correct responses pre- and posttutorial showed a significant increase in correct responses posttutorial ($t(23) = 4.39$, $p = .0001$).

Comments

The students reported no technical problems. A selection of comments on the SUMS resource is included in Appendix B.

Context 2: SUMS Used to Support Laboratory Teaching and Coursework

Self-efficacy

Mixed design 2 X 2 ANOVAs with pretest and posttest as the within-subject factor and use of tutorial (user versus nonuser) as a between-subject factor was used to analyse each self-efficacy task separately.

The only two tasks that revealed findings relevant to the hypothesis were Task 1 (explaining a p value) and 6 (reporting a t test). As Table 2 shows, there was an interaction between testing time and use of tutorial for both tasks (Task 1, $F(1, 78) = 5.85$, $p = .01$; Task 2, $F(1, 78) = 7.60$, $p = .01$). Simple main effect analyses found that for both tasks the tutorial user group had lower self-efficacy at pretest (all $p < .01$), but there was no difference between the user and nonuser groups at posttest.

Comprehension

The scores on the comprehension test from the second questionnaire were compared using an independent groups t test. This revealed that the user group scored higher than the nonuser group ($t(180) = 2.28$, $p = .02$) (users $M = 4.18$, $SD = 1.08$; (nonusers $M = 3.67$, $SD = 1.08$).

The discrepancy in degrees of freedom between the analyses of the self-efficacy tasks and the comprehension test is due to more students completing the second questionnaire only.

Discussion

Use of the SUMS tutorial as a stand-alone resource (Context 1) improved students' self-efficacy in their ability to complete a variety of basic statistical tasks, and their performance on the comprehension test.

One exception was self-efficacy for the task of distinguishing between the information given by the mean, median and mode. Self-efficacy in this task was initially high ($M = 3.70$), allowing less room for improvement after using the resource. However, self-efficacy was also initially high for Tasks 1 and 2 (experimental design) and these tasks did show improvement after the tutorial. Level 1 of the tutorial covered measures of central tendency at the most basic level of interpretation, and students may already be familiar with central tendency from the GCSE/Standard Grade mathematics syllabus (Mulhern & Wylie, 2006). Thus, the central tendency topic at Level 1 may not have included enough new material to significantly improve students' self-efficacy.

Students who used the SUMS tutorial to support their laboratory and coursework showed better performance on a statistics comprehension test than those who did not use the SUMS tutorial. In terms of self-efficacy, the user group initially rated their self-efficacy as lower than the nonuser group on two tasks in the pretest (explaining a p value and reporting a t test). It appears that the tutorial appealed to those who lacked confidence in their abilities with these tasks. However, by the posttest both groups gained in self-efficacy and were now equivalent, showing that the user group had reached the same level of confidence as the nonusers. Overall, it appears that the tutorial had more impact on students motivated to seek help due to anxiety or lack of confidence in statistics (i.e., users in Context 1) and less impact as a resource to assist coursework (i.e., users in Context 2). Nevertheless, it is encouraging that the self-efficacy gains that have been found with face-to-face teaching in statistics (Finney & Schraw, 2003) can be partially replicated by online resources.

Although the results of the above evaluations are promising, there are limitations within their designs which limit their interpretation. There was no real control condition in the first evaluation, meaning there is some ambiguity as to whether any improvement is due to the tutorial, or whether similar improvements could have been gained with more traditional teaching approaches (e.g., textbooks, lectures).

In the second evaluation the self-selection of the tutorial by participants means that there was nonrandom allocation of participants to the two conditions, with the attendant issues in the generalisation of the results. In fact, the analyses above indicate that there were noticeable differences in self-efficacy between the intervention and nonintervention groups.

Another limitation of the present study is that the resource was tested at Level 1 only, and the impact of Levels 2 and 3 on improving self-efficacy is unknown. Further intervention using a sample of students with a wider range of abilities and from subsequent years of the course would be required to test the effectiveness of more challenging levels of the resource.

Conclusions

Overall, the SUMS online resource produced positive results in self-efficacy and comprehension as a stand-alone resource for first-year psychology students. The second context produced positive results in terms of comprehension, however the results in term of self-efficacy were complicated by the differences in confidence of users and nonusers. These evaluations only made use of Level 1 (the easiest level) of the resource as the interventions were focused on introductory classes. Future evaluations will include examining the usefulness of Levels 2 and 3 of the resource, and comparisons with more traditional teaching methods.

Acknowledgements

We would like to thank Dr Mitchum Bock for useful comments on the statistical content of SUMS, and Dr Lorna Morrow and Judith Stevenson, for comments on earlier versions of SUMS. The project was funded by the University of Glasgow Learning and Teaching Development Fund.

References

Aberson, C., Berger, D., Healy, M., Kyle, D., & Romero, V. (2000). Evaluation of an interactive tutorial for teaching the central limit theorem. *Teaching of Psychology, 27*, 289–291.

Aberson, C., Berger, D., Healy, M., & Romero, V. (2003). Evaluation of an interactive tutorial for teaching hypothesis testing concepts. *Teaching of Psychology, 30*, 76–79.

Baloglu, M., & Zelhart, P. (2003). Statistical anxiety: A detailed review of the literature. *Psychology and Education: An Interdisciplinary Journal, 40*, 27–37.

Bandura, A. (1997). *Self efficacy: The exercise of control.* New York: Freeman.

Bandura, A., Ross, D., & Ross, S. A. (1961). Transmission of aggression through imitation of aggressive models. *Journal of Abnormal and Social Psychology, 63*, 575–582.

Ben-Zvi, D. (2000). Towards understanding the role of technological tools in statistical learning. *Mathematical Thinking and Learning, 2*, 127–155.

Berger, D. E. (2008). *Web interface for statistics education (WISE)*. Retrieved January 29, 2009, from http://wise.cgu.edu/

Dancey, C., & Reidy, J. (2007). *Statistics without maths for psychology: Using SPSS for Windows* (4th ed.). Harlow: Prentice Hall.

Dinov, I. D. (2002). *SOCR: Statistics Online Computational Resource*. Retrieved January 29, 2009, from http://socr.ucla.edu/SOCR.html

Dinov, I. D., Sanchez, J., & Christou, N. (2008). Pedagogical utilization and assessment of the statistic online computational resource in introductory probability and statistics courses. *Computers and Education, 50*, 284–300.

Finney, S. J., & Schraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology, 28*, 161–186.

Garfield, J. (1995). How students learn statistics. *International Statistical Review, 63*, 25–34.

Kranzler, J. H., & Pajares, F. (1997). An exploratory factor analysis of the mathematics self-efficacy scale Revised MSES-R. *Measurement and Evaluation in Counseling and Development, 29*, 215–228.

Lane, M. L., Lane, J., & Hall, R. (2004). Self-efficacy and statistics performance among sport studies students. *Teaching in Higher Education, 9*, 435–448.

Lane, D. (2008). *Online statistics: An interactive multimedia course of study*. Retrieved January 29, 2009, from http://onlinestatbook.com/index.html

Miller, S. H. (1984). *Experimental design and statistics*. London: Routledge.

Mulhern, G., & Wylie, J. (2004). Changing levels of numeracy and other core mathematical skills among psychology undergraduates between 1992 and 2002. *British Journal of Psychology, 95*, 355–370.

Mulhern, G., & Wylie, J. (2006). Mathematical prerequisites for learning statistics in psychology: Assessing core skills of numeracy and mathematical reasoning among undergraduates. *Psychology Learning and Teaching, 5*, 119–132.

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*, 195–209.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*, 543–578.

Pajares, F., & Miller, D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology, 86*, 193–203.

Swingler, K. M. (2008). *Statistical Understanding Made Simple: An automatic resource generator for teaching statistics*. Retrieved January 29, 2009, from http://www.gla.ac.uk/sums

Tariq, V. N. (2002). A decline in the numeracy skills amongst bioscience undergraduates. *Journal of Biological Education, 83*, 76–83.

Utts, J., Sommer, B., Acredolo, C., Mahler, M. W., & Matthews, H. (2003). A study comparing traditional and hybrid internet-based instruction in introductory statistics classes. *Journal of Statistics Education, 11*, 1–10.

West, R. W. (1996). *Java applets for teaching*. Retrieved January 29, 2009, from http://www.stat.sc.edu/~west/javahtml/

Manuscript received 10 September 2008.
Revision accepted for publication 12 January 2009.

Appendix A

Self-efficacy statements: Context 1

1. Identify a within and a between subjects design.
2. Identify an experimental and null hypothesis and the direction of a hypothesis.
3. Interpret a frequency plot of any given data set.
4. Distinguish between the information given by the mean, median and mode.
5. Explain what the value of the standard deviation means in terms of the variable being measured.
6. Describe the effect of sample size on the sampling error.
7. Identify if a frequency histogram of a given data set falls into a normal distribution.
8. Choose the appropriate *t* test according to the experimental design and hypothesis.
9. Explain the purpose of an inferential test.
10. Interpret the *p* value from the results of a *t* test.

Self-efficacy statements: Context 2

1. Explain a *p* value.
2. Explain a standard deviation.
3. Understand the difference between two standard deviations.
4. Choose the right *t* test.
5. Explain the difference between a mean and median.
6. Report a *t* test.

Appendix B

Evaluation Context 1: Students' Comments on the Online Tutorial

The tutorial was easy to use and not too complicated, I found it very useful.

I found this very useful. I would just like to have more time to grasp it properly. If this was available on the portal it would be very helpful.

I liked the games a lot – they made sense of the explanations.

Too much information to take in at one time but is all made worthwhile because it's on the portal for further learning.

Games allowed me not to be scared of the info that was put in the explanation section.

It is time consuming to go through the whole tutorial, but the games had a clear structure and a clear structure makes learning statistics much more interesting.

Found it very informative and statistics seem a lot clearer now. Thank you!