Hostage of the Software: Experiences Teaching Inferential Statistics to Undergraduate Human-Computer Interaction students and a Survey of the Literature

Frode E. Sandnes[1,2(✉)] and Evelyn Eika[1]

[1] Faculty of Technology, Art and Design, Oslo Metropolitan University, Oslo, Norway
[2] Westerdals Oslo School of Art, Communication and Technology, Oslo, Norway
{frodes, Evelyn.Eika}@oslomet.no

Abstract. Students' knowledge of inferential statistics is lacking in many Computer Science study programs. Yet, the needs for inferential statistic skills have emerged with new fields of study such as human-computer interaction involving observation of human activity. This paper presents experiences teaching inferential statistics to undergraduate computer science students with a focus on the actual goals of the investigations and not the mechanisms and mathematics of statistics. The teaching framework involves teaching statistics as a set of systematic black-box tools.

Keywords: Inferential statistics teaching, experimental design, human-computer interaction, statistics software

1 Introduction

Null hypothesis significance testing can be a powerful mechanism for analyzing empirical data, yet it can also be highly misleading when applied incorrectly [1]. Several voices echo the challenges of making statistics interesting and relevant to students [2, 3], especially when these students are non-specialists [4]. Several pedagogical strategies have been applied to the teaching of statistics such as small group collaboration [5], problem-based learning [6], the use of practical examples [7, 8], and distance and virtual education modes [9, 10].

Many study programs in engineering and especially computer science include some courses on statistics. These courses are often taught by mathematicians, statisticians, or physicists and often focus on the mathematical sides of statistics. The mathematical angle is indeed valuable and applicable to many areas of computer science. However, the more practical sides of statistics such as inferential statistics that concerns itself with hypothesis testing are also increasingly relevant. Traditionally, inferential statistics has been more visible in the curriculum of other fields of study such as agriculture, where one compares crops, medicine, and health sciences [11], where one may compare the effects of treatment, psychology, education [12, 13], educational policy [14, 15], language learning [16, 17], linguistics [18, 19, 20], and social science [21], to name a few.

Inferential statistics has been absent in the curriculum of many engineering, and computer science, study programs. This is because traditional computer science and engineering have utilized different research methods, which do not require inferential statistics. Typical tasks include measuring differences based on deterministic processes, for example, the timing of computer program execution, success rates such as detecting object in images [22] and in video [23]. Other measurement types include the accuracy of computation, such as that of geo-localization based on information in images [24], geo-location based on light intensities [25], shadows [26], or underwater light intensities [27]. Engineering research often also simply involves demonstrating the workings of a new engineering solution, for instance, demonstrating that reliable information transfer is possible via paper [28].

Our experience is that practical knowledge and appreciation of inferential statistics is generally low among staff in such programs, even among the statisticians. This may be because computer science has traditionally focused on systems and algorithm. When such systems or algorithms are measured under controlled conditions, they give very similar or even identical results each time. Inferential statistics has therefore not been considered a particularly useful or relevant methodology. However, with the emergence of more multidisciplinary topics in computer science and engineering, such as human-computer interaction, ICT in education, and the merging of ICT and health, the need of inferential statistics has emerged as it involves measuring human behavior, which is highly variable. We share our objectives with Peiris et al. [29] who promoted the introduction of effective statistical tools to students early during their undergraduate studies.

2 Human-Computer Interaction

Human-computer interaction (HCI) has emerged in the curriculum of many computer science study programs during the last decade. HCI can be considered the study of all phenomena related to the interaction between human and machine. HCI is a multidisciplinary field that relies on a range of research methodologies. We have focused on HCI generally and design for all specifically [30]. One may consider the research method a tool where one needs to choose the most suitable tool for a given problem. Examples from our own human-computer research lab include the use of traditional computer science techniques in HCI such as graph theory [31], heuristic evaluation [32, 33], qualitative research methods based on interviews [34] and text analysis [35, 36], visualization [37, 38, 39], as well as design and development. Design includes sketching in 2D [40] and 3D [41, 42], 3D modelling [43], design of concepts such as new interaction styles for self-service kiosks [44], collaborative work [45] and volunteering [46], tactile feedback for pedestrians [47], design of devices such as augmented reality displays [48], and the development of new design methods [49].

Examples of exploration through development include new interaction techniques such as wheel controls [50], human behavior monitoring based on touch dynamics [51], new color design tools that support human contrast perception [52, 53], physical navigation tools for blind users using radar [54], and virtual navigation in static panoramic views [55]. Common to these studies is that they allow a new idea to be tried by building working prototypes. The focus is often not on the testing of the final results, but rather on the discovery over various technical challenges on the way and how these can be solved.

Inferential statistics is indeed also a highly relevant methodology in human-computer interaction. However, the degree to which the focus is placed on qualitative or quantities methods seems to vary as many human-computer interaction courses are purely qualitative. We have taken a balanced approach introducing the students to a wide range of methods, including inferential statistics. Typical examples of quantitative problems studied by students and staff in our lab include comparative studies of dyslexia [56, 57, 58]. Such studies often compare two groups, namely, dyslexic participants and a control group, and therefore often rely on paired t-tests. T-tests are also used in other studies of cognitive aspects of interaction involving two groups [59, 60] and studies involving users with and without vision [61] or when comparing two keyboard layouts [62] or left-right interaction directions [63]. Text entry experiences such as those involving new interactions styles often rely on repeated measures ANOVAs as there are often more than two levels per factor, or more factors [64, 65]. Often text entry experiments require learning, such as chording [66, 67], and the learning effects are studied over time through various sessions [68]. ANOVA is thus often a suitable tool in such cases.

3 Challenges of learning statistics

3.1 Pedagogical strategies

There are different pedagogical approaches to teaching statistics ranging from the very mathematical and theoretical to the very practical. Theoretical approaches usually evolve around lectures, while the practical approaches focus on learning by doing through assignments and coursework. The mathematical approach is common as it is simple and justified by the argument that students should fully understand the underlying principles. There appears to be a belief that good mathematical skills are essential for learning statistics. However, Galagedera et al. [69] found that perceived mathematical abilities have little effect on students' performance in elementary statistics. Much of the literature seems to favor practical approaches over theoretical approaches where students learn through practice. Marson [70] collected empirical evidence to support that the three key elements that lead to successful teaching of statistics include repetition, immediate feedback, and the use of real data.

3.2 Teacher qualifications

Teachers are essential to the successful teaching of statistics [71]. Several studies have pointed to the fact that statistics often is taught by non-statisticians with a lack of basic statistics knowledge [72] or with misconceptions about statistics [73]. In our view, the teacher must have a good grasp of statistics, but even more important in the context of applied experimental design is that the teacher has practical working experience with empirical experiment and analysis, perhaps from their own research. It is our opinion that it is not enough for a teacher to have a sound understanding of theoretical statistics without experience from actual empirical research. The preference for more practical and simple procedures over mathematical elegance is also echoed by Wood [74, 75] and Khait [76], among others.

3.3 Learning resources

We have found that until recently there have been very few suitable textbooks and learning resources available. Most resources focus on the mathematical sides and few give practical advice that is relevant for empirical research. Gliner et al. [77] surveyed several statistics textbooks and found that none of them contributes to removing common misconceptions about null hypothesis significance testing. Fortunately, the situation is gradually changing with the emergence of relevant textbooks such as [78] and various online leaning resources.

3.4 Statistics software

Computer assisted instruction has been shown to have a positive effect in statistics teaching [79]. However, a key challenge has been the lack of suitable statistics software. There is a vast number of commercial software packages on sale such as SPSS, SAS, STATA, etc., with SPSS being one of the market leaders. One main challenge with SPSS is the high cost, making it financially unrealistic to acquire for many higher education institutions with limited budgets. SPSS and other commercial software have also been criticized for being undemocratic in the sense that the internal algorithms are not open for scrutiny. Open source software is hailed as giving the users an opportunity to investigate the correctness of the underlying statistical algorithms. From our experience, the complexity of SPSS is the largest challenge. It provides a huge amount of functionality and it can be very daunting to navigate the menus for a novice. SPSS is considered easy to use once one has learned its use and knows some statistics and which tests one needs. It is undoubtedly hard to use for beginners who in addition are insecure about which tests to apply for a given problem. The many YouTube instruction videos for basic operations are testaments to this. Simple functions are hidden behind obscure menus. For instance, to conduct a Friedman test for non-parametric repeated measures analysis of variance of three groups, one needs to go to the analysis menu (which is quite long), select non-parametric tests, select legacy tests, and then K-related samples. This path is easy to remember, but nearly impossible to discover for beginners. Each test is associated with several complex dialogues with the various options hidden behind various buttons (hidden functionality). Moreover, the results appear in a separate output window, and it can be hard for users to connect their actions with the displayed output.

The open source landscape is dominated by R-project [80]. R-project is a comprehensive and powerful statistics package which is relatively easy to use, despite its being command-line based. There are also several open source GUI-alternatives for R [81], and R is easily extended through scripting and is thus popular with programmers. The main problem with R-project is that repeated measure analysis of variance is quite inaccessible. It is possible to perform such tests but it is not straightforward to set up such tests without in-depth statistical knowledge.

For several years we used Excel for the introduction to hypothesis testing as well as course management [82]. The advantages of Excel are that it is commonly available, although it is not open source. We have used the analysis toolpack that contains t-tests, one-way and two-way ANOVA, and regression analysis. Our experience is that the most challenging aspects of using Excel are for students to correctly interpret the results as the output is verbose. One major drawback with Excel is that it does not provide repeated measures ANOVA, which is essential for human-computer interaction as it most often involves within-subject designs. Note that it is possible to perform a rudimentary one-way repeated measures analysis using the two-way ANOVA function with subjects as one factor. There are, however, several extensions available for Excel, such as Charles Zaiontz's comprehensive statistics tools for Excel [83]. Regrettably, the security policy of our university does not allow students and teachers to install third-party macro packages in Microsoft Office on university machines. Various versions of Excel have also been criticized during the past two decades for inaccurate computations, including Excel 97 [84], Excel 2003 [85], Excel 2007 [86, 87], and Excel 2010 [88].

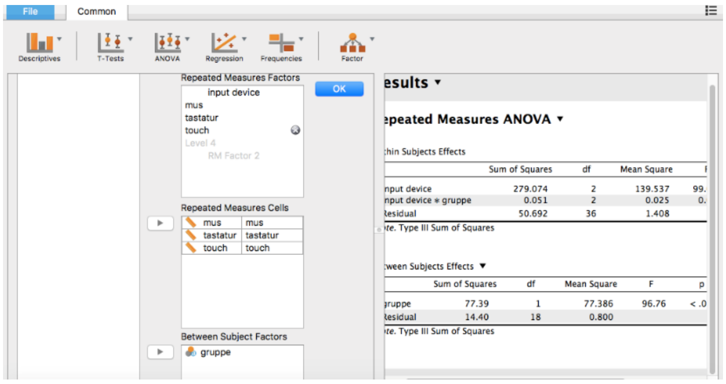

Fig. 1. JASP user interface (mixed ANOVA view).

In our teaching, we have started to use JASP (Jeffrey's Amazing Software Package) [89], a relatively young statistics software package developed at the University of Amsterdam (see Fig. 1 for an example screenshot). Note that JASP is different from the project of the same name (Java-based Statistics Processor [90] from two decades ago). JASP is based on R-project but presents the functionality through a simple and streamlined user interface that only exposes the most important functionality needed in introductions to inferential statistics, such as paired and independent t-tests, ANOVA, repeated measures ANOVA, correlation, and factor analysis. The ANOVA analysis functionality is especially useful as it supports multiple factors and mixed designs (within and between group factors) in addition to several post-hoc tests such as Tuckey, Sheffe, Bonferroni, and Holm-Bonferroni. Normality testing and other assumption tests are also available via the user interface. The number of options is also streamlined, making the perceived impression of simplicity. The output is also minimalistic, only displaying essential information. It changes dynamically as the users alter the configuration of the statistical tests. This overall software appears non-threatening and invites exploration. Moreover, its structure promotes correct use of statistical tests. The main drawback of JASP is the lack of non-parametric tests for more than two groups.

3.5 Statistical concepts

Students struggle with several issues when learning inferential statistics. Sotos et al. [91] gave a comprehensive review of common statistics misconceptions among students in various disciplines. Our experience is that the statistical notation appears cryptic and it is hard to understand the meaning of the various values listed, that is, the statistics for a given test, degrees of freedom, and the p-value. Students' conceptions and misconceptions of the p-value have been studied in detail by Reaburn [92], Wagenmakers [93], and others.

It is also challenging to connect the shorthand notation in scientific papers with the values that appear in the statistics software. Further, many students are very uncertain about how many observations are needed. Normal distribution is another issue. Normality is often one of the core assumption of the parametric tests. Another issue students struggle to grasp is the necessity of using an ANOVA test on all levels of the factor under investigation instead of just running a t-test on combination of pairs of levels. This challenge is also reported for papers published in medical journals [94, 95]. Students also struggle with understanding the need to use repeated measures ANOVA instead of an ordinary ANOVA when dealing with within-subject designs. In human-computer interaction, within-subjects designs are probably the most common; it is easier to execute as fewer participants are needed. In agriculture, on the other hand, within-subjects designs are usually not possible, and most studies are employing between-subject designs relying on basic ANOVA.

One of the largest challenges is selecting the correct statistical test given a specific problem. Many different tests were named after various people, which could be daunting for a beginner, yet quite recognizable for someone with some experience with empirical experimentation statistics. Examples include Wilcoxon, Mann-Whitney, Friedman, Kruskal-Wallis, etc. The connection of applying tests with strange names under certain circumstances may seem to be a bit of black magic to students. Unless one is using a full statistical package such as SPSS, or R-project, students may not actually have access to all the tests and therefore may choose a t-test or ANOVA as these are more easily available.

One recent textbook on experimental design [78] avoids t-tests altogether by analyzing two samples with an ANOVA test or a repeated measures ANOVA test. Indeed, the t-test can be replaced by an ANOVA test and students will then not use t-tests incorrectly by doing pairwise comparisons, a problem found in scientific papers as well [94, 95]. However, it is our opinion that when reporting an experiment with a t-test, the use of the t-test gives vital information to the reader about the experimental design. The use of t-tests is also an experimental convention when comparing two groups. We have opted for teaching the t-tests despite the risk of its being used incorrectly.

Our experience is also that students find it challenging to differentiate between when to use non-parametric tests and parametric tests. The assumption of normality is well known, but there are also other assumptions for various tests, such as homogeneity and sphericity that are less obvious. Moreover, the simple notion of considering the data type of the dependent variables is often ignored. It is recommended that interval data are used with parametric tests, and ordinal, categorical, and dichotomous (binary) data are used with non-parametric tests.

When the data suggest a non-parametric test, it may seem confusing and frustrating to students when there are actually no obvious standard tests available, e.g., a mixed-multi-factor designs. The many questions posted on various discussion groups are testaments to this challenge. It has also been found that many scientific papers incorrectly report parametric tests when the data suggest non-parametric tests [96].

3.6 Experimental design

Some students struggle with practical experimental issues that affect the statistical analyses. These difficulties include ensuring that the presentation order is varied in within-subject designs, recruiting enough participants, having sufficiently long session to get reliable measurements, and running a pilot to ensure that experimental setup is working as expected.

Based on our experiences with teaching statistics to undergraduate students over several years, we have developed a simple pedagogical framework with the specific goal of improving the quality and validity of the statistical analyses carried out by the students. Our framework is discussed in the subsequent sections.

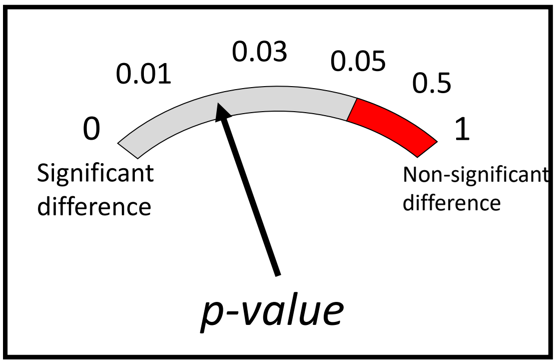

Fig. 2. Using a needle-instrument metaphor for helping students to build a mental model for how to interpret the p-value.

4 A toolbox approach to inferential statistics

The human-computer interaction course is offered to second year undergraduate students (three-year study programs). It is attended by students with a comprehensive mathematical background (computer engineering students) and those with a minimal amount of mathematical background (applied computer science students). It is also attended by students from other fields including library and information science, archival science, etc. The inferential statistics constitutes approximately 1/6 of the total syllabus in terms of lectures (approximately two weeks), yet it takes up 1/3 of the students' work (1 out of 3 graded project works). Inferential statistics was introduced into the curriculum for about 10 years and it was mostly based on Excel until the emergence of JASP. Quite some time is spent motivating the students for learning inferential statistics with arguments for why it is important, and when it is applicable.

When the students were first introduced to statistics, they had to relate to an overwhelming number of issues at once, such as various statistical tests with odd names, various constraints and assumptions, obscure notation, complex software, and unfamiliar terminology. The key to our pedagogical approach is to treat the statistical tests as measurement tools from a toolbox where the data are the input and the p-value is the output value of importance. We have used a needle-instrument metaphor to help students built a mental model of how to interpret p-values (see Fig. 2). The needle is a universal symbol of quantity and limit. When the needle is on the left-hand side of the red bar, there is significance (usually difference); when it is on the right red bar, there is non-significance. The left side of the red arc marks the significance level, which is usually 0.05 unless some correction is used such as Bonferroni or Holm-Bonferroni.

The focus is on the use of the tools and not how they work. The internal mathematical and algorithmic workings are omitted completely. It is an explicit goal not to include any mathematical expressions at all in the course material, besides the p-value inequalities.

symbol(value) = value, ***p*** = value

example	Test
t(38) = 2.428, ***p*** = .020	t-test
Z = -1.807, ***p*** = .071	Wilcoxon
F(2,27) = 4.467, ***p*** = .021	ANOVA
χ2(2) = 7.600, ***p*** = .022	Chi-squared
r(15) = -.918, ***p*** = .001	Pearson's
r_s(15) = -1.0, ***p*** = .001	Spearmann
H = 14.338, ***p*** < .01	Kruskal-Wallis
U = 67.5, ***p*** = .034	Mann–Whitney U-test

Fig. 3. Notation and notation pattern reference sheet for common tests.

A central part of the framework is also to train statistical literacy in the sense of being able to read and comprehend the terminology and notation found in various scientific papers. Extracts from scientific papers are hence used in the teaching. Students are also encouraged to search for and read literature for their assignments. Within the area of human-computer interaction, a great number of research papers can be read by undergraduate students as these are relevant to phenomena of user interfaces that the students are already familiar with. Good sources include proceedings from ACM SIGCHI conferences, ACM ASSETS, etc. The goal is to reduce anxiety associated with the unfamiliar coding of the standard notation and build students' confidence in interpreting the notation. Students who can decode the notation are probably also more likely to correctly encode the notation. Next, experiences from reading research papers are intended to help illustrate the purpose and use of the notations in practice. To help students, we employ simple summary sheets such as the one shown in Fig. 3.

		parametric	Non-parametric		
Category	Experiment type	Interval data	Ordinal and interval data	Nominal data	Dichotomous data
Independent measurements	Two groups	t-test	Mann–Whitney U-test	χ2-test for 2 × C table	χ2-test for 2 × 2 table (Fisher´s exact test (N < 20)
	Three or more groups	One-way ANOVA	Kruskal–Wallis one-way ANOVA	χ2 -Test for R × C table	N/A
	Three or more groups, multiple factors	Two-way, three-way, … ANOVA	None	None	None
Repeated measurements	Two groups	Paired t-test	Wilcoxon signed rank test	McNemar's test	McNemar's test
	Three or more groups	Repeated measures ANOVA	Friedman's test	Cochran's Q	None
	Three or more groups, multiple factors	Multi-factor repeated measures ANOVA	None	None	None
Mixed design	Two or more groups	Mixed design ANOVA	None	None	None
	Two or more groups, multiple factors	Multifactor mixed design ANOVA	None	None	None
Association	Correlation	Pearson's	Spearman' s rank	N/A	N/A

Fig. 4. Map of statistical tests.

The framework also relies on a map of statistical tests (see Fig. 4) that gives an overview of the tests covered inspired by an overview presented by McGrum-Gardner [97]. The horizontal dimension signals the data type of the dependent variable, and the vertical dimension signals the organization of the dependent variables and the experimental design. Clearly, the diversity of statistical tests and special cases are too large to be captured by a simple sheet of paper, and we thus focus on the most commonly needed cases.

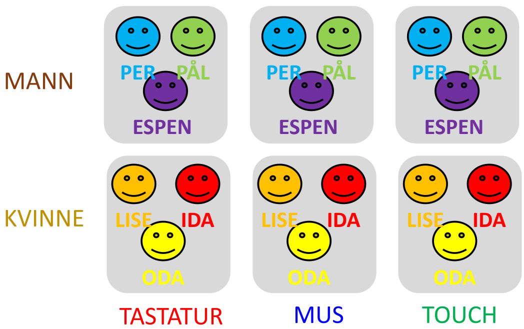

Fig. 5. Visualizing experimental design (in Norwegian)

Visualization is used extensively to illustrate the various concepts such as experimental designs. For example, Fig. 5 illustrates an example of a mixed design with one within-subjects factor (input device) with three levels (keyboard, mouse, and touch) and one between-subjects factor with two levels (male and female). The essence of the diagram is that each participant is exposed to all the within-group factors while different participants belong to only one between-group factor. Since the same participants occur in several (within) groups, a repeated measures ANOVA is needed.

We use the JASP software to show examples in-class and recommend students to use JASP for their assignments. However, it is not a requirement and students are free to employ the statistical tools of their choice.

Typical experiments that have been given to the students include finding empirical evidence of what gives the best performance of keyboards with alphabetical and qwerty layout, digit input with numeric keypad versus the number keys on normal qwerty keyboards, and what type of date input technique works best on web pages. For more common phenomena, the students must design the test environments; for more specialized cases such as scanning keyboards, the students are provided with basic code which they can tailor to their particular needs. More recently we have also experimented with free projects where the students themselves must propose a phenomenon they want to explore by conducting an experiment where they collect data that are analyzed using inferential statistics.

In addition to the challenges discussed, we have found that some students are uncertain about whether to include all raw observations in the test or whether to use a representative aggregated value for each participant/session (such as a mean performance score). Most students seem to grasp the idea of measuring performance. However, measuring error rates appear to be difficult in practice. In particular, how does one define what is meant by an error for a given problem? Also, it is quite common for students to make errors in the experimental setup which they discover after completing the project. Such errors, nevertheless, may provide learning opportunities.

5 Conclusions

This paper reviewed some of the literature on teaching inferential statistics together with our own experiences and observations from the classroom. We also provided examples of how we changed our inferential statistics teaching with the aim to make students perform inferential statistics more correctly. For a long time, the statistics teaching has been hindered by the limited availability of suitable statistics software. As known, the way the statistics is presented in software such as Excel leads students and researches to perform statistics in certain way, and sometimes incorrectly. Although software packages (e.g., JASP) are making a huge leap in making inferential statistics available to students, there is still room for improvement in terms of the potential for software support for good statistical practices.

References

1. Nickerson, R. S.: Null hypothesis significance testing: a review of an old and continuing controversy. Psychological methods 5, 241-301 (2000).
2. Phua, K.: How to Make the Learning of Statistics Interesting, Fun and Personally Relevant: Using Progressive Material as Examples for In-class Analysis and to Raise Social Awareness. Radical Statistics 95, 4 (2007)

3. Gordon, S.: Understanding students' experiences of statistics in a service course. Statistics Education Research Journal 3, 40-59 (2004)
4. Yilmaz, M. R.: The challenge of teaching statistics to non-specialists. Journal of statistics education 4, 1-9 (1996)
5. Garfield, J.: Teaching statistics using small-group cooperative learning. Journal of Statistics Education 1, 1-9 (1993)
6. Bland, J. M.: Teaching statistics to medical students using problem-based learning: the Australian experience. BMC Medical Education 4, 31 (2004)
7. Chermak, S., Weiss, A.: Activity-based learning of statistics: Using practical applications to improve students' learning. Journal of Criminal Justice Education 10, 361-372 (1999)
8. Smith, A. E., Martinez-Moyano, I. J.: Techniques in teaching statistics: Linking research production and research use. Journal of Public Affairs Education 18, 107-136 (2012)
9. Gemmell, I., Sandars, J., Taylor, S., Reed, K.: Teaching science and technology via online distance learning: the experience of teaching biostatistics in an online Master of Public Health programme. Open Learning 26, 165-171 (2011)
10. López, A. J., Pérez, R.: Learning statistics in a shared virtual campus. summarizing a five-year experience. Instructional Technology and Distance Learning 2, 29-40 (2005)
11. Howlett, B., Phelps, P.: Actively and Formatively Teaching Statistics To Physician Assistant Students. The Journal of Physician Assistant Education 17, 48-52 (2006)
12. Jian, H.-L., Sandnes, F.E., Huang, Y.-P., Cai, L., Law, K.: On students' strategy-preferences for managing difficult course work. IEEE Transactions on Education 51, 157-165 (2008)
13. Jian, H. L., Sandnes, F. E., Huang, Y. P., Huang, Y. M., Hagen, S.: Studies or Leisure?: a Cross-cultural Comparison of Taiwanese and Norwegian Engineering Students' Preferences for University Life. International journal of engineering education 26, 227-235 (2010)
14. Jian, H.-L., Sandnes, F.E., Huang, Y.-P., Huang, Y.-M., Hagen, S.: Towards harmonious East-West educational partnerships: A study of cultural differences between Taiwanese and Norwegian engineering students. Asia Pacific Education Review 11, 585-595 (2010)
15. Jian, H. L., Sandnes, F. E., Huang, Y. P., Huang, Y. M.: Cultural factors influencing Eastern and Western engineering students' choice of university. European Journal of Engineering Education 35, 147-160 (2010)
16. Brown, J. D.: Teaching statistics in language testing courses. Language Assessment Quarterly, 10, 351-369 (2013)
17. Jian, H.-L., Sandnes, F.E., Huang, Y.-P., Law, K., Huang, Y.-M.: The role of electronic pocket dictionaries as an English learning tool among Chinese students. Journal of Computer Assisted Learning 25, 503-514 (2009)
18. Eika, E., Hsieh, Y.: On Taiwanese pupils' ability to differentiate between English/l/and/r: A study of L1/L2 cross-language effects. First Language 0142723717709106, (2017)
19. Jian, H.-L.: On English speakers' ability to communicate emotion in Mandarin. Canadian Modern Language Review 71, 78-106 (2015)
20. Jian, H.-L.: Prosodic challenges faced by L1 English speakers reading Mandarin. ACTA Linguistica Hungarica 62, 35-62 (2015)
21. Cronin, A., Carroll, P.: Engaging business students in quantitative skills development. e-Journal of Business Education and Scholarship Teaching 9, 1-14 (2015)
22. Huang, Y.-P., Chang, T.-W., Chen, J.-R., Sandnes, F.E.: A back propagation based real-time license plate recognition system. Int. J. Pattern Recogn. Artif. Intell. 22, 233–251 (2008)
23. Huang, Y.P., Hsu, L.W., Sandnes, F.E.: An intelligent subtitle detection model for locating television commercials. IEEE Trans. Man Cybern. B 37, 485–492 (2007)
24. Sandnes, F. E.: Sorting holiday photos without a GPS: What can we expect from contents-based geo-spatial image tagging? In: Pacific-Rim Conference on Multimedia, LNCS, vol. 5879, pp. 256-267. Springer, Berlin, Heidelberg. (2009).
25. Sandnes, F.E.: Where was that photo taken? Deriving geographical information from image collections based on temporal exposure attributes. Multimedia Syst. 16, 309–318 (2010)

26. Sandnes, F.E.: Determining the geographical location of image scenes based on object shadow lengths. J. Sig. Process. Syst. 65, 35–47 (2011)
27. Gómez, J. V., Sandnes, F. E., Fernández, B.: Sunlight intensity based global positioning system for near-surface underwater sensors. Sensors 12, 1930-1949 (2012)
28. Huang, Y. P., Chang, Y. T., Sandnes, F. E.: Ubiquitous information transfer across different platforms by QR codes. Journal of Mobile Multimedia 6, 3-13 (2010)
29. Peiris, S., Peseta, T.: Learning statistics in first year by active participating students. In Proceedings of The Australian Conference on Science and Mathematics Education, (2012)
30. Whitney, G., Keith, S., Bühler, C., Hewer, S., Lhotska, L., Miesenberger, K., Sandnes, F.E., Stephanidis, C., Velasco, C.A.: Twenty five years of training and education in ICT Design for All and Assistive Technology. Technology and Disability 3, 163-170 (2011)
31. Sandnes, F. E.: Evaluating mobile text entry strategies with finite state automata. In: Proceedings of the 7th international conference on Human computer interaction with mobile devices & services, pp. 115-121. ACM (2005)
32. Sandnes, F.E., Jian, H.L., Huang, Y.P., Huang, Y.M.: User interface design for public kiosks: an evaluation of the Taiwan high speed rail ticket vending machine. J. Inf. Sci. Eng. 26, 307–321 (2010)
33. Berget, G., Herstad, J., Sandnes, F.E.: Search, read and write: an inquiry into web accessibility for dyslexics. Stud. Health Technol. Inform. 229, 450–460 (2016)
34. Sandnes, F.E.: What do low-vision users really want from smart glasses? Faces, text and perhaps no glasses at all. In: Miesenberger, K., Bühler, C., Penaz, P. (eds.) ICCHP 2016. LNCS, vol. 9758, pp. 187–194. Springer, Cham (2016). doi:10.1007/978-3-319-41264-1_25
35. Eika, E.: Universally designed text on the web: towards readability criteria based on antipatterns. Stud. Health Technol. Inform. 229, 461–470 (2016)
36. Eika, E., Sandnes, F.E.: Assessing the reading level of web texts for WCAG2.0 compliance—can it be done automatically? In: Di Bucchianico, G., Kercher, P. (eds.) Advances in Design for Inclusion. Advances in Intelligent Systems and Computing, vol. 500, pp. 361–371. Springer, Cham (2016)
37. Eika, E., Sandnes, F.E.: Authoring WCAG2.0-compliant texts for the web through text readability visualization. In: Antona, M., Stephanidis, C. (eds.) UAHCI 2016. LNCS, vol. 9737, pp. 49–58. Springer, Cham (2016). doi:10.1007/978-3-319-40250-5_5
38. Sandnes, F.E.: Understanding WCAG2.0 color contrast requirements through 3D color space visualization. Stud. Health Technol. Inform. 229, 366–375 (2016)
39. Sandnes, F.E.: On-screen colour contrast for visually impaired readers: Selecting and exploring the limits of WCAG2.0 colours. In: Black, A., Lund, O., Walker, S. (eds.) Information design: research and practice, pp. 405–416, Routledge (2016)
40. Sandnes, F.E., Jian, H.L.: Sketching with Chinese calligraphy. Interactions 19, 62–66 (2012)
41. Sandnes, F.E.: Communicating panoramic 360 degree immersed experiences: a simple technique for sketching in 3D. In: Antona, M., Stephanidis, C. (eds.) UAHCI 2016. LNCS, vol. 9738, pp. 338–346. Springer, Cham (2016). doi:10.1007/978-3-319-40244-4_33
42. Sandnes, F.E.: PanoramaGrid: a graph paper tracing framework for sketching 360-degree immersed experiences. In: Proceedings of the International Working Conference on Advanced Visual Interfaces, pp. 342–343. ACM (2016)
43. Sandnes, F.E.: Sketching 3D immersed experiences rapidly by hand through 2D cross sections. In: Proceedings of REV 2017. LNCS (in press)
44. Hagen, S., Sandnes, F.E.: Toward accessible self-service kiosks through intelligent user interfaces. Pers. Ubiquit. Comput. 14, 715–721 (2010)
45. Hagen, S., Sandnes, F. E.: Visual scoping and personal space on shared tabletop surfaces. Journal of Ambient Intelligence and Humanized Computing 3, 95-102 (2012)
46. Chen, W. C., Cheng, Y. M., Sandnes, F. E., Lee, C. L.: Finding suitable candidates: the design of a mobile volunteering matching system. In: International Conference on Human-Computer Interaction. LNCS, vol. 6763, pp. 21-29. Springer, Berlin, Heidelberg (2011)

47. Lin, M.W., Cheng, Y.M., Yu, W., Sandnes, F.E.: Investigation into the feasibility of using tactons to provide navigation cues in pedestrian situations. In: Proceedings of the 20th Australasian Conference on Computer-Human Interaction: Designing for Habitus and Habitat, pp. 299-302. ACM (2008)
48. Sandnes F.E., Eika, E.: Head-Mounted Augmented Reality Displays on the Cheap: A DIY Approach to Sketching and Prototyping Low-Vision Assistive Technologies, In: Antona, M., Stephanidis, C. (eds.) Universal Access in Human-Computer Interaction. Designing Novel Interactions Practices, LNCS, vol. 10278, pp. 168-186, Springer, (2017)
49. Sandnes, F.E.: Designing GUIs for low vision by simulating reduced visual acuity: reduced resolution versus shrinking. Stud. Health Technol. Inform. 217, 274–281 (2015)
50. Sandnes, F.E. Huang, Y.P.: From smart light dimmers to the IPOD: text-input with circular gestures on wheel-controlled devices, International Journal of Smart Home 1, 97-108 (2007)
51. Sandnes, F. E., Zhang, X.: User identification based on touch dynamics. In 9th International Conference on Ubiquitous Intelligence & Computing and 9th International Conference on Autonomic & Trusted Computing (UIC/ATC), pp. 256-263. IEEE (2012)
52. Sandnes, F.E., Zhao, A.: A contrast colour selection scheme for WCAG2. 0-compliant web designs based on HSV-half-planes. In: Proceedings of SMC2015, pp. 1233–1237. IEEE (2015)
53. Sandnes, F.E., Zhao, A.: An interactive color picker that ensures WCAG2.0 compliant color contrast levels. Procedia-Comput. Sci. 67, 87–94 (2015)
54. Gomez, J.V., Sandnes, F.E.: RoboGuideDog: guiding blind users through physical environments with laser range scanners. Procedia Comput. Sci. 14, 218–225 (2012)
55. Sandnes, F.E., Huang, Y.P.: Translating the viewing position in single equirectangular panoramic images. In: Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 389–394. IEEE (2016)
56. Berget, G., Mulvey, F., Sandnes, F.E.: Is visual content in textual search interfaces beneficial to dyslexic users? Int. J. Hum.-Comput. Stud. 92–93, 17–29 (2016)
57. Berget, G., Sandnes, F.E.: Do autocomplete functions reduce the impact of dyslexia on information searching behaviour? A case of Google. J. Am. Soc. Inf. Sci. Technol. 67, 2320–2328 (2016)
58. Berget, G., Sandnes, F.E.: Searching databases without query-building aids: implications for dyslexic users. Inf. Res. 20 (2015)
59. Sandnes, F.E., Lundh, M.V.: Calendars for Individuals with Cognitive Disabilities: A Comparison of Table View and List View. In: Brewster, S., Dunlop, M. (eds.) Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility, ACM, pp. 329-330 (2015)
60. Sandnes, F.E., Jian, H.L., Pair-wise variability index: Evaluating the cognitive difficulty of using mobile text entry systems. In: International Conference on Mobile Human-Computer Interaction. LNCS, vol. 3160, pp. 347-350. Springer Berlin Heidelberg (2004)
61. Sandnes, F.E., Tan, T.B., Johansen, A., Sulic, E., Vesterhus, E., Iversen, E.R.: Making touch-based kiosks accessible to blind users through simple gestures. Universal Access in the Information Society 11, 421-431 (2012)
62. Sandnes, F. E.: Effects of common keyboard layouts on physical effort: Implications for kiosks and Internet banking. In: Sandnes, F.E., Lunde, M. Tollefsen, M., Hauge, A.M., Øverby, E., Brynn, R. (eds.) The proceedings of Unitech2010: International Conference on Universal Technologies, pp. 91-100. Tapir Academic Publishers (2010)
63. Sandnes, F.E., Thorkildssen, H.W., Arvei, A., Buverad, J.O.: Techniques for fast and easy mobile text-entry with three-keys. In: Proceedings of the 37th Annual Hawaii International Conference on System Sciences. IEEE (2004)
64. Sandnes, F.E.: Directional bias in scrolling tasks: a study of users' scrolling behaviour using a mobile text-entry strategy. Behav. Inf. Technol. 27, 387–393 (2008)

65. Sandnes, F.E., Aubert, A.: Bimanual text entry using game controllers: relying on users' spatial familiarity with QWERTY. Interact. Comput. 19, 140–150 (2007)
66. Sandnes, F. E., Huang, Y. P.: Chord level error correction for portable Braille devices. Electronics Letters 42, 82-83 (2006)
67. Sandnes, F. E., Huang, Y. P.: Chording with spatial mnemonics: automatic error correction for eyes-free text entry. Journal of information science and engineering 22, 1015-1031 (2006)
68. Sandnes, F.E.: Can spatial mnemonics accelerate the learning of text input chords? In: Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 245–249. ACM (2006)
69. Galagedera, D., Woodward, G., Degamboda, S.: An investigation of how perceptions of mathematics ability can affect elementary statistics performance. International Journal of Mathematical Education in Science and Technology 31, 679-689 (2000)
70. Marson, S. M.: Three empirical strategies for teaching statistics. Journal of Teaching in Social Work 27, 199-213 (2007)
71. Petocz, P., Gordon, S., Reid, A.: Recognising and developing good statistics teachers. In Proceedings of the Seventh International Conference on Teaching Statistics, ICOTS7, Salvador, Brazil (2006).
72. Dabos, M.: Two-year college mathematics instructors' conceptions of variation. In The Teaching and Learning of Statistics (pp. 175-176). Springer International Publishing (2016)
73. Haller, H., Krauss, S.: Misinterpretations of significance: A problem students share with their teachers. Methods of Psychological Research 7, 1-20 (2002)
74. Wood, M.: The case for crunchy methods in practical mathematics. Philosophy of Mathematics Education Journal 14, (2001)
75. Wood, M.: Maths Should Not Be Hard: The Case for Making Academic Knowledge More Palatable. Higher Education Review 34, 3-19 (2002)
76. Khait, A.: Intelligent guesses and numerical experiments as legitimate tools for secondary school algebra. Teaching Mathematics and its Applications 23, 33-40 (2004)
77. Gliner, J. A., Leech, N. L., Morgan, G. A.: Problems with null hypothesis significance testing (NHST): what do the textbooks say?. The Journal of Experimental Education 71, 83-92 (2002)
78. Mackenzie, I. S.: Human-Computer Interaction. An Empirical Perspective, Morgan Kaufman (2013).
79. Basturk, R.: The effectiveness of computer-assisted instruction in teaching introductory statistics. Journal of Educational Technology & Society 8, (2005)
80. Crawley, M. J.: The R book. John Wiley & Sons (2012).
81. Snellenburg, J., Laptenok, S., Seger, R., Mullen, K., Van Stokkum, I.: Glotaran: a Java-based graphical user interface for the R package TIMP. Journal of Statistical Software 49 (2012)
82. Sandnes, F. E., Eika, E.: A Simple MVC-Framework for Local Management of Online Course Material. In: Uskov V. Howlett R., Jain L. (eds.) International Conference on Smart Education and Smart E-Learning. Smart Innovation, Systems and Technologies, vol. 75, pp. 143-153. Springer, Cham. (2017).
83. Zaiontz, C.: Real statistics using Excel. Retrieved January 21 (2017). http://www.real-statistics.com/
84. McCullough, B. D., Wilson, B.: On the accuracy of statistical procedures in Microsoft Excel 97. Computational Statistics & Data Analysis 31, 27-37 (1999)
85. McCullough, B. D., Wilson, B.: On the accuracy of statistical procedures in Microsoft Excel 2003. Computational Statistics & Data Analysis 49, 1244-1252 (2005)
86. McCullough, B. D., Heiser, D. A.: On the accuracy of statistical procedures in Microsoft Excel 2007. Computational Statistics & Data Analysis 52, 4570-4578 (2008)

87. Yalta, A. T.: The accuracy of statistical distributions in Microsoft® Excel 2007. Computational Statistics & Data Analysis 52, 4579-4586 (2008)
88. Mélard, G.: On the accuracy of statistical procedures in Microsoft Excel 2010. Computational statistics 29, 1095-1128 (2014)
89. Marsman, M., Wagenmakers, E. J.: Bayesian benefits with JASP. European Journal of Developmental Psychology 14, 545-555 (2017)
90. Nakano, J., Yamamoto, Y., Kobayashi, I., Fujiwara, T.: Using a statistical system Jasp for basic statistical education. Spring Conference of Korea Statistical Society, Jeonju, Korea (2014)
91. Sotos, A. E. C., Vanhoof, S., Van den Noortgate, W., Onghena, P.: Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. Educational Research Review 2, 98-113 (2007)
92. Reaburn, R.: Introductory statistics course tertiary students' understanding of p-values. Statistics Education Research Journal 13, (2014)
93. Wagenmakers, E. J.: A practical solution to the pervasive problems of p values. Psychonomic bulletin & review 14, 779-804 (2007)
94. Wu, S., Jin, Z., Wei, X., Gao, Q., Lu, J., Ma, X., Wu, C., He, Q., Wu, M., Wang, R., Xu, J.: Misuse of statistical methods in 10 leading Chinese medical journals in 1998 and 2008. The Scientific World Journal 11, 2106-2114 (2011)
95. Skaik, Y.: The bread and butter of statistical analysis "t-test": Uses and misuses. Pakistan Journal of Medical Sciences 31, 1558-1559 (2015)
96. Yim, K. H., Nahm, F. S., Han, K. A., Park, S. Y.: Analysis of statistical methods and errors in the articles published in the Korean journal of pain. The Korean journal of pain 23, 35-41 (2010)
97. McCrum-Gardner, E.: Which is the correct statistical test to use?. British Journal of Oral and Maxillofacial Surgery 46, 38-41 (2008)