StatHand: An Interactive Decision Tree Mobile Application to Guide Students’ Statistical Decision Making

Final report 2017

Lead Institution: Curtin University

Partner Institutions: The Australian National University, University of New England and Griffith University.

Project Leader: Peter Allen

Team Members: Dr Frank Baughman, Associate Professor Lynne Roberts, Dr Dirk Van Rooy, Dr Adam Rock and Dr Natalie Loxton.

www.stathand.net
Support for the production of this report has been provided by the Australian Government Department of Education and Training. The views expressed in this report do not necessarily reflect the views of the Australian Government Department of Education and Training.

With the exception of the Commonwealth Coat of Arms, and where otherwise noted, all material presented in this document is provided under Creative Commons Attribution-ShareAlike 4.0 International License creativecommons/4.0/license

The details of the relevant licence conditions are available on the Creative Commons website (accessible using the links provided) as is the full legal code for the Creative Commons Attribution-ShareAlike 4.0 International License creativecommons/4.0/legalcode

Requests and inquiries concerning these rights should be addressed to:
Learning and Teaching Support
Student Information and Learning Branch
Higher Education Group
Department of Education and Training

GPO Box 9880
Location code C50MA7
CANBERRA ACT 2601

<learningandteaching@education.gov.au>

2017

ISBN 978-1-76051-139-5 [PRINT]
ISBN 978-1-76051-140-1 [PDF]
ISBN 978-1-76051-141-8 [DOCX]
Acknowledgements

The project team would like to acknowledge the following individuals and groups for their contributions to the success of this project.

Mortaza Rezae, Xavier Begue and Ivan Dwiputera, who coded the StatHand iOS and web applications under the mentorship of Dr Aneesh Krishna at Curtin University.

Amanda Lourenco, James Finlay, Hannah Uren, Andrew Johnson and Dr Kate Dorozenko, who assisted with the collection, analysis and reporting of data summarised herein and with the development of content published in StatHand.

Parts of this report have been reproduced or adapted from the following publications, distributed under the terms of the Creative Commons Attribution License (CC-BY):


List of acronyms and abbreviations used

APA: American Psychological Association

APAC: Australian Psychology Accreditation Council
Executive summary

Project context and aim

Quantitative research methods underpin psychological literacy and are critical for the development of professional competence in psychology. They feature prominently in the undergraduate psychology curriculum in Australia and internationally, and are reflected in the course learning outcomes and graduate attributes specified by psychology course accrediting agencies worldwide. However, they are also an area of weakness for many students. Students are known to particularly struggle with the task of selecting appropriate statistical tests and procedures for different types of research questions, hypotheses and data types—an ability referred to as ‘selection skills’. Although not often practised in class, selection skills can be developed with training, and they appear to be underpinned by ‘structural awareness’—that is, an ability to disregard the surface features of a research scenario and to focus on its structural features and the relations between them.

Various aids and resources have been developed to facilitate the process of statistical decision making and to promote structural awareness. Those that have gained most traction are based around the concept of a ‘decision tree’, which prompts users to engage with each structural feature of their research design as well as with the hierarchical and horizontal relationships between them. The use of decision trees to aid statistical decision making is supported empirically by research that demonstrates both their objective efficacy and their subjective appeal.

However, despite their popularity, traditional (paper-based) statistical decision trees have limitations. Many of these limitations can be overcome through the use of mobile learning technologies. It is within this context that the StatHand project was proposed and funded. The aims of the project were to:

- Harness mobile technologies to build students’ statistical decision making capacities.
- Develop, and embed within our teaching, a cross-platform mobile application and companion website that actively engage students with the key decisions needed to select and deploy appropriate statistics for common data and hypotheses.
- Freely disseminate and evaluate the efficacy of the application and website.

Project approach and outputs

StatHand is a cross-platform application that helps users quickly identify appropriate statistics for a wide range of research questions, hypotheses and data types. Once the user has identified an appropriate statistic, StatHand further advises on its computation (including assumption testing and effect sizes where applicable), interpretation and reporting, and provides links to further resources and examples. StatHand is freely available in the iOS App Store for both iPhone and iPad and as a fully mobile-compatible web application at www.stathand.net.
The development of StatHand began with the identification of five key data analysis objectives and the statistics associated with each. A series of maps were then developed, linking the objectives with the statistics via a sequence of annotated decision points. As this content was being drafted and refined, technical specifications for the iOS and web applications were being prepared and quotes sought from local and international application/web design companies. The technical work was contracted to a team of final-year computer science students, who employed agile development methodologies complemented by the software development lifecycle framework to ‘build’ Version 1 of StatHand. The iOS applications were implemented using Xcode and Objective-C; the web application was developed using Visual Studio and the ASP.NET framework.

Following the release of StatHand Version 1, a second round of content development was initiated, this time focused on guiding users through the computation, interpretation and reporting of the statistics that StatHand helps to identify. This content was then incorporated into StatHand Version 2 in a manner similar to that described for StatHand Version 1.

Supplemental to the application is a package of instructors’ resources and two peer-reviewed publications. The instructors’ resources include a plain-language rationale for the application as a learning and teaching tool, instructions for using the application, tips for integrating the application into undergraduate research methods classes, and adaptable active learning activities. The first of the two peer-reviewed publications establishes the need for StatHand via a qualitative analysis of interviews conducted with psychology students, researchers and instructors. The second describes the rationale behind the development of StatHand and its features, and provides recommendations for its integration into the research methods curriculum.

**Project evaluation**

In the first of two evaluation studies, students and instructors were engaged in structured discussions around the merits and limitations of StatHand and were asked to provide suggestions for its future development. The student participants preferred StatHand to a paper decision tree. They liked that it was interactive and easily accessible, and considered it faster and more reliable than familiar alternatives. They said they would recommend StatHand to others but felt it would be of greater value if it also provided advice on the computation, interpretation and reporting of the statistics it helped to identify. These features have since been added to the application. The instructors saw StatHand as an aesthetically pleasing, clean, simple and user-friendly application of benefit to students because it encouraged self-efficacy and active learning. The instructors hoped the application would be incorporated into undergraduate psychology units, and they provided many suggestions for its expansion, including the addition of instructions for computing the statistics that StatHand helps users identify. These instructions are now part of the core functionality of StatHand.

In the second evaluation study, students were randomised to one of four aids for decision making (StatHand, a paper decision tree, a familiar textbook, and a familiar textbook plus a paper decision tree) and asked to identify a suitable statistic for each of five research scenarios. After controlling for prior performance in research methods and statistics, the
StatHand group demonstrated higher performance accuracy than the other three groups. The StatHand group also reported significantly lower cognitive load, higher confidence and greater satisfaction than at least one other group. However, it took the StatHand group longer to complete the task. The results of this evaluation indicated that StatHand can be considered relatively instructionally efficient when compared with the other three aids.

**Project impact and dissemination**

The project has been successful in advancing the team members’ understanding of the challenges that students face when selecting statistics, and this improved understanding has informed their teaching and supervision practices. It has also aided career and study progression for the technical development team and the student researchers involved with the project. The project has had an impact on the team members’ immediate students, with StatHand now embedded in the undergraduate psychology curricula at four Australian universities. The word has spread via various dissemination activities, including publications, presentations and professional networking. There has been narrow and broad opportunistic adoption, but systemic adoption is an ongoing process.
Table of contents

Acknowledgements........................................................................................................... ii
List of acronyms and abbreviations used ........................................................................ iii
Executive summary ............................................................................................................... iv
Table of contents ................................................................................................................ vii
Tables and figures .............................................................................................................. ix
Tables ................................................................................................................................. ix
Figures ................................................................................................................................. ix
Chapter 1: Context and aims .............................................................................................. 1
  Introduction ........................................................................................................................ 1
  Selecting appropriate statistics ......................................................................................... 2
  Aids for statistical decision making ............................................................................... 4
  Mobile learning technologies ......................................................................................... 6
  Summary and aims .......................................................................................................... 6
Chapter 2: Approach and outputs ...................................................................................... 8
  StatHand application ....................................................................................................... 8
  Instructors’ resources .................................................................................................... 13
  Publications .................................................................................................................... 13
Chapter 3: Evaluation .......................................................................................................... 15
  Qualitative evaluation .................................................................................................. 15
    Method ......................................................................................................................... 15
    Findings from student focus groups ........................................................................ 16
    Findings from instructor interviews ......................................................................... 19
  Experimental evaluation ............................................................................................. 22
Chapter 4: Impact and dissemination ............................................................................. 23
  Impact ............................................................................................................................. 23
  Dissemination ............................................................................................................... 24
    Publications ............................................................................................................... 24
    Presentations .............................................................................................................. 24
References ......................................................................................................................... 26
Appendix A ......................................................................................................................... 32
Appendix B ......................................................................................................................... 33
Tables and figures

Tables

Table 1. The Statistics, Tests and Procedures Described in StatHand, Grouped by Data Analysis Objective

Figures

Figure 1. An illustrative path through Version 1 of the StatHand iOS application on an iPhone 6.

Figure 2. Screen 9 in Version 2 of the StatHand application in landscape mode on an iPad Air 2 and Screen 10 showing the Index in the StatHand web application, running in Microsoft Edge on a Surface Pro 3.

Figure 3. Examples of scenarios included in the StatHand instructors’ resources package.
Chapter 1: Context and aims

Introduction

Quantitative research methods underpin psychological literacy (Cranney, Morris & Botwood, 2015; McGovern et al., 2010; Roberts, Heritage & Gasson, 2015) and are critical for the development of professional competence in psychology. They have featured in undergraduate psychology curricula since the discipline’s inception (Lipp et al., 2007; Norcross et al., 2016; Perlman & McCann, 1999a, 1999b; Wilson & Provost, 2006) and are reflected in the course learning outcomes and graduate attributes specified by psychology course accrediting organisations worldwide.

Of the six graduate attributes for four-year Australian undergraduate psychology courses, as listed in the Australian Psychology Accreditation Council (APAC, 2010) Accreditation Standards for Psychology Courses, at least two necessitate a solid understanding of quantitative research methods. The first, ‘research methods in psychology’, concerns the ability to ‘describe, apply and evaluate different research methods used by psychologists’ and to ‘design and conduct basic studies to address psychological questions’ (APAC, 2010, p. 40). The second, ‘critical thinking skills’, refers to the capacity to ‘apply knowledge of the scientific method in thinking about problems related to behaviour and mental processes’ (APAC, 2010, p. 41). At the time of writing, the APAC standards are under review, although the current consultation draft (dated June 2016) suggests that research methods will continue to play as prominent a position in the future of undergraduate psychology training in Australia as they have in the past. For example, the draft standards state that, among other competencies, graduates of an Australian four-year undergraduate psychology degree will be able to ‘analyse and critique theory and research in the discipline of psychology and communicate these in written and oral formats’, ‘demonstrate self-directed pursuit of scholarly inquiry in psychology’ and ‘undertake research to investigate [questions] relevant to the discipline of psychology’ (APAC, 2016, pp. 14–15).

In the United States, the second of five learning goals in an undergraduate psychology course detailed by the American Psychological Association (APA) Board of Educational Affairs Task Force on Psychology Major Competencies (2013) is ‘scientific inquiry and critical thinking’, which requires ‘the development of scientific reasoning and problem solving, including effective research methods’, ‘applying research design principles to drawing conclusions about psychological phenomena’ and ‘designing and executing research plans’ (p. 15). Similarly, in the United Kingdom, the British Psychological Society (2016) states that an undergraduate degree in psychology should equip graduates with the capacity to ‘generate and explore hypotheses and research questions’, ‘carry out empirical studies involving a variety of methods of data collection’, ‘analyse, present and evaluate quantitative … data and evaluate research findings’, and ‘employ evidence-based reasoning’ (pp. 19–20).

Collectively, these goals and standards reflect a widely held understanding that an ability to source, read, understand and critically evaluate research literature is a necessary precursor to evidence-based practice in psychology (APA Presidential Task Force on Evidence-Based
Practice, 2006). The vast majority of this literature is based on quantitative research methods (Kidd, 2002; Munley et al., 2002). It is also widely held that some of the most effective ways of teaching these skills involve regularly engaging students in all aspects of the research process—from the conception of meaningful research questions through to design, analysis, interpretation and reporting (Earley, 2014; Marek, Christopher & Walker, 2004; Stoloff, Good, Smith & Brewster, 2015; Wagner, Garner & Kawulich, 2011). Therefore, nearly all psychology departments provide undergraduate students with multiple opportunities to conduct original empirical research, either individually or in collaboration with other students or faculty (Kierniesky, 2005; Perlman & McCann, 2005).

**Selecting appropriate statistics**

Despite their prominence throughout psychology curricula, and their importance, research methods and (particularly) statistics are recognised areas of weakness for many students (Garfield & Ahlgren, 1988; Garfield & Ben-Zvi, 2007; Murtonen & Lehtinen, 2003; Murtonen, Olkinuora, Tynjala & Lehtinen, 2008). Students are known to struggle in particular with selecting appropriate statistical tests and procedures for different types of research questions, hypotheses and data types—an ability referred to as ‘selection skill’ (Ware & Chastain, 1989).

To illustrate this point, Gardner and Hudson (1999) presented 21 brief research scenarios to a sample of 23 students and asked them to recall appropriate statistical procedures for as many scenarios as possible within a 45-minute period. The scenarios reflected statistical concepts typically found in introductory statistics textbooks and widely used in behavioural science research. Despite most students having completed at least six research methods and statistics units, the overwhelming majority of them found the task difficult and performed poorly. On average, the students managed to read 10.9 scenarios within the allocated time and answered 25.3 per cent of them correctly. An additional 15.7 per cent of answers were coded as ‘partially correct’. When Gardner and Hudson (1999) asked the students how they made their decisions, several explanations for the poor performance emerged. These included students misinterpreting the research scenarios, knowing appropriate statistics but not being able to name them, misidentifying the measurement levels of variables (e.g., nominal, ordinal, continuous) and seizing on misleading keywords and data presentation formats.

When students have been asked to recognise (rather than recall) appropriate statistics, their performance has been similarly underwhelming. For example, Ware and Chastain (1989) developed an eight-item multiple-choice selection skill test, which they believed contained ‘problems that students should be able to solve after completing [an] introductory statistics course’ (p. 225). When they administered the test at the conclusion of such a course, the students answered fewer than 45 per cent of the items correctly. Ware and Chastain attributed this poor performance, at least partially, to a curriculum that taught statistical techniques ‘one at a time’ (p. 226) and gave little consideration to the development of selection skills. Some researchers have recognised that having had few opportunities to practise selection skills could account for the difficulties students experience when they must decide which statistic to use (e.g., Lovett & Greenhouse, 2000; Quilici & Mayer, 1996, 2002; Yan & Lavigne, 2014).
Although few research methods and statistics courses appear to do so, it is possible to train selection skills. For example, when Ware and Chastain (1991) restructured their introductory statistics course and placed more emphasis on when to use various statistics and less emphasis on computational procedures, they observed a significant improvement on their multiple-choice selection skills test. In a more controlled context, Quilici and Mayer (2002) trained students to focus on the structural (e.g., the nature of the independent and dependent variables, and the relationship between them) rather than on the surface-level (e.g., topic) features of basic research scenarios. In doing so, they improved the students’ abilities to categorise scenarios according to how they would be analysed. After training, the students were also better able to generate new scenarios that could be analysed using the same statistical procedures as existing scenarios. More recently, Yan and Lavigne (2014) reported similar findings. They observed that when students were provided with worked examples that emphasised the structural features of simple research scenarios, the students’ performance on subsequent categorisation tasks improved, as did their ability to identify the structural features defining each category.

Together, these findings suggest that selection skills are underpinned by ‘structural awareness’ (Quilici & Mayer, 2002) - that is, the ability to disregard the surface features of a research scenario and focus on its structural features and the relations between them. Consider the following scenario presented to participants by Allen, Dorozenko and Roberts (2016, p. 2):

You work at a university library, and have been tasked with finding out which students accrue the largest ‘overdue fines’. The head librarian has provided you with a data file that gives you the total amount of fines (in dollars) accrued by each borrower during the previous 12 months, along with a range of additional information (e.g., each borrower’s course of study, age, gender, number of items borrowed etc.).

Identifying an appropriate statistical technique for this scenario requires disregarding its ‘cover story’, or surface-level features, and focusing on identifying its structural features and the relationships between them. It requires first recognising that the broad intent is prediction (rather than, for example, a comparison between means) and identifying the independent and dependent variables. Here, there are several independent variables of varying types (i.e., dichotomous, nominal and continuous), and one continuous dependent variable. Second, it requires constructing a generic conceptual model representing the relationships between structural features. In this instance, the researcher’s intent is to use a combination of several independent variables to predict scores on a continuous dependent variable. Last, it requires integrating the conceptual model with existing knowledge to find possible solutions. For many research scenarios, a range of statistical techniques could be used to analyse the data, and this requires the researcher to compare techniques to determine the most appropriate statistical technique for the particular set of circumstances. Sometimes there may be two or more equally suitable techniques, but here the most obvious solution is multiple linear regression, which would provide coefficients useful for addressing the head librarian’s question, although additional considerations (e.g., the likely distribution of the dependent variable) may suggest other possibilities. An iterative process may be required between statistical technique selection and the testing of assumptions in order to make a final decision.
Without assistance, students find the process described above very challenging. However, ‘experts’ do not (Allen, Dorozenko & Roberts, 2016). Although the exact point of transition from novice to expert in this specific context is not known, the transition appears to necessitate a substantial amount of experience. For example, Rabinowitz and Hogan (2008) recruited graduate students enrolled in master’s and PhD courses at a university with ‘a very well established psychometrics program’ (p. 401) to complete a series of triad judgement tasks. In these tasks, participants were required to identify which of two statistics scenarios ‘goes best’ with a specified target scenario. When faced with the option of selecting a scenario that shared structural but not surface characteristics with the target, or the reverse (i.e., surface but not structural characteristics), even those participants with the greatest amount of experience (i.e., those who had completed between four and eight statistics units) did not reliably choose on the basis of structure. Those with the least experience made their choice based on surface characteristics. Indeed, it was not until the choice was between a scenario that was similar on structural characteristics only and one that was dissimilar on both structure and surface that the ‘experienced’ participants reliably made their choice based on the structural features of the scenarios. Furthermore, in the Gardner and Hudson (1999) study described earlier, even the most experienced members of their sample (students admitted entry into fourth year, master’s and PhD courses in psychology and education) rarely answered more than 50 per cent of the scenarios correctly.

Aids for statistical decision making

The above discussion suggests several points. First, even experienced students are not able to autonomously select appropriate statistics in a reliable way. Second, students are often required to make such decisions relatively early in their courses but are not always explicitly taught how to make them. Third, making incorrect decisions can carry substantial negative consequences. At a pragmatic level, basing a research report on the results of the ‘wrong’ statistical test will lead to incorrect interpretations and likely poor grades. At a deeper level, it reveals deficits in statistical reasoning and thinking (Bradstreet, 1996; Chance, 2002). Collectively, these points indicate a need for aids or resources that students can rely on to facilitate the process of statistical decision making and perhaps speed their transition from novice to autonomous expert.

Numerous such aids have been developed, including tip sheets that sort statistical tests according to their defining characteristics (e.g., Twycross & Shields, 2004), and charts that link common research goals to corresponding statistics (e.g., Beitz, 1998). However, the aids that have gained most traction are based around the idea of a ‘decision tree’. Such resources facilitate the process of statistical decision making by prompting the user to engage with each structural feature of their research design as well as the hierarchical and horizontal relationships between them (Schau & Mattern, 1997). In the short term, this ensures that the user considers all relevant aspects of the design before deciding on a statistical test, thus increasing the likelihood of a correct decision. In the longer term, decision trees help users to integrate their knowledge of statistical concepts into coherent and organised schemata, which can be quickly and effectively activated when required (Yin, 2012).

Decision trees have been used to guide statistical decision making for at least half a century (e.g., Fok, Angelidis, Ibrahim & Fok, 1995; Mock, 1972), and they are now commonly
included in statistics textbooks (e.g., Allen, Bennett & Heritage, 2014; Field, 2013; Tabachnick & Fidell, 2013). Their inclusion in such books is supported empirically by research illustrating both their objective efficacy and their subjective appeal. For example, Carlson, Protsman and Tomaka (2005) and Protsman and Carlson (2008) demonstrated that decision trees could facilitate statistical decision making that was significantly faster and more accurate (by a multiple of three) than the more traditional methods of statistical test selection (e.g., searching through a familiar textbook). The decision tree method was also significantly more popular than the textbook method among students (Carlson et al., 2005; Protsman & Carlson, 2008).

Despite their popularity, traditional statistical decision trees have limitations. First, they are usually limited in scope by the requirement to fit them on a single sheet of paper or within the pages of a textbook. Consequently, definitions and other information that would facilitate traversing the tree are either spatially separated from the tree or completely absent (Blankenship & Dansereau, 2000; Koch & Gobell, 1999). Second, when given to students without accompanying resources (e.g., a textbook), decision trees do not provide sufficient detail to interpret and execute the statistics that they help to identify. Third, although the complexity and nonlinearity of a statistical decision tree may be helpful to experienced users, new users may experience difficulty in fully processing the tree (sometimes referred to as ‘map shock’) and consequently lose the motivation to use it (Blankenship & Dansereau, 2000; Nesbit & Adesope, 2011).

To overcome these limitations, some researchers and educators have adapted the traditional decision tree for digital media. These hypertext systems typically comprise a series of interconnected pages or nodes (Unz & Hesse, 1999). Hypertext systems do not have the space constraints associated with paper decision trees, and they allow for links to external resources that aid learning (Koch & Gobell, 1999). Map shock can be eliminated because the user is shown only a small section of the tree at any given time, reducing its complexity and hence its ability to overwhelm (Blankenship & Dansereau, 2000). However, a hypertext system can provide a disjointed experience, where users become disoriented and lose track of their location within the system. This phenomenon, sometimes referred to as being ‘lost in hyperspace’ (Otter & Johnson, 2000), can constrain the novice user’s ability to develop an understanding of how concepts are connected. Despite this limitation, meta-analytic findings support the overall efficacy of hypertext systems when compared with textual interfaces. In particular, graphical map interfaces are associated with more effective (medium to large effect sizes) and efficient (small to medium effect sizes) performance than textual interfaces (Chen & Rada, 1996).

Koch and Gobell (1999) adapted paper decision trees for delivery on the World Wide Web and, in doing so, were able to provide users with definitions, links to online resources, and information about how to enter and analyse data in commonly used statistical software. Like Carlson et al. (2005) and Protsman and Carlson (2008), Koch and Gobell (1999) found that students using their online decision tree were better able to identify appropriate statistical tests than students in a comparison condition. Unfortunately, Koch and Gobell’s (1999) website is no longer active. A current example of an online statistical test selection tool is provided by the University of California, Los Angeles (UCLA) Institute for Digital Research and Education at http://www.ats.ucla.edu/stat/mult_pkg/whatstat/default.htm. The site provides a table of statistical tests based on the number and nature of dependent
and independent variables, with ‘how to’ links for a range of statistical software. However, the large size of the table (and the use of a table rather than a decision tree) combined with the limited information provided may contribute to map shock for inexperienced users.

A range of software for selecting statistical techniques has also been developed. Some currently available software applications (e.g., Subramanian, 2014; Wacharamanotham, Subramanian, Völkel & Borchers, 2015) automatically select the statistical test for the user rather than guide the user through the steps in order to make their own decision, greatly reducing the pedagogic potential. STTestMAP (Eng, Hua, Kheng, Ying & Yee Hui, 2011) is a visual tool that guides students through a systematic process to select a statistical test, but it does not appear to be publicly available. Despite their potential benefits, hypertext decision trees and currently available software require the user to have a live internet connection.

Mobile learning technologies

Unlike websites and web applications, mobile learning applications can be developed to maintain all (or most) of their functionality in the absence of an internet connection (Kretser et al., 2015). Mobile learning can be defined as ‘the use of mobile or wireless devices for the purpose of learning while on the move’ (Park as cited in Yu, Lee & Ewing, 2014, p. 2126). In recent years, the use of mobile learning technologies such as smart devices and mobile applications has increased rapidly, and among Western higher education students their use is near ubiquitous (Chen, Sellhamer, Bennett & Bauer, 2015; Dahlstrom & Bichsel, 2014). Their broad appeal relates to many factors, including portability, accessibility to information and resources anywhere and at any time (Jeng, Wu, Huang, Tan & Yang, 2010) and utility. Increasingly, students are expressing preferences for mobile learning technologies for various educational purposes (Bowen & Pistilli, 2012; Farley et al., 2015), and mobile learning applications have been identified as one of the technologies expected to have the biggest impact on education this decade (Johnson, Adams & Cummins, 2012; Martin et al. 2011).

Summary and aims

Quantitative research methods are essential for the development of professional competence in psychology. However, they are also an area of weakness for many students. In particular, students are known to struggle with the skill of selecting appropriate statistics for common research questions, hypotheses and data types. Decision trees are known to facilitate this selection process, but extant trees have a range of limitations that constrain their efficacy. Recent developments in mobile learning technologies offer the potential to overcome many of these limitations. Within this context, the StatHand (formerly StatTree) project was proposed to the Australian Government Office for Learning and Teaching. The aims of the project, as articulated in the project proposal were to:

- Harness mobile technologies to build students’ statistical decision-making capacities.
- Develop, and embed within our teaching, a cross-platform mobile application and companion website that actively engage students with the key decisions needed to select and deploy appropriate statistics for common data and hypotheses.
• Freely disseminate and evaluate the efficacy of the application and website.

The following chapters of this report describe how, and to what extent, these aims were achieved.
Chapter 2: Approach and outputs

The primary outputs associated with this project are the StatHand application and the associated resources. These resources include instructors’ resources and two peer-reviewed publications. The outputs were largely developed in tandem. They are outlined in the subsections that follow.

StatHand application

StatHand is a cross-platform application that helps users quickly identify appropriate statistics, tests and analytic procedures (hereafter referred to as ‘statistics’) for a wide range of research questions, hypotheses and data types. Once the user has identified an appropriate statistic, StatHand further advises on its computation, interpretation and reporting. StatHand is freely available in the iOS App Store for both iPhone and iPad and is accessible as a fully mobile-compatible web application at www.stathand.net.

Version 1 of StatHand focused on the selection of appropriate statistics. Content development began with the identification of five key data analysis objectives and the statistics associated with each. These are listed in Table 1. A map was then developed, which identified the key decision points between each objective and the relevant statistics. One ‘branch’ of this map (for the ‘compare samples’ objective) is shown in Appendix B. Explanatory text to accompany each decision was then drafted. As shown in Appendix C, this text is brief but provides sufficient detail for a novice user to make an informed decision without the immediate need to consult additional sources (e.g., the internet, textbooks).
Table 1. The Statistics, Tests and Procedures Described in StatHand, Grouped by Data Analysis Objective

<table>
<thead>
<tr>
<th>Objective</th>
<th>Statistics, tests and procedures described in StatHand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Describe a sample</td>
<td>Bar graph; category count; histogram; interquartile range; mean; median; mode; pie chart; range; standard deviation; stem-and-leaf plot</td>
</tr>
<tr>
<td>Compare samples</td>
<td>ANCOVA (independent samples and mixed; one-way and factorial); ANOVA (independent samples, repeated measures and mixed; one-way and factorial); chi-square (goodness of fit and contingencies); Cochran’s Q test; Friedman two-way ANOVA; Kruskal-Wallis one-way ANOVA; Mann-Whitney U test; McNemar test of change; t-test (one sample, independent samples and paired samples); Wilcoxon signed-rank test (one sample and paired samples)</td>
</tr>
<tr>
<td>Analyse relationships or associations between variables</td>
<td>Chi-square test of contingencies (with Phi or Cramer’s V); correlation coefficients (point-biserial, rank-biserial, Spearman’s and Pearson’s); eta; linear regression (bivariate and multiple, standard and hierarchical); logistic regression (binary and multinomial, standard and hierarchical); ordinal regression (standard and hierarchical)</td>
</tr>
<tr>
<td>Examine the underlying structure of a measure</td>
<td>Confirmatory factor analysis; exploratory factor analysis; principal components analysis</td>
</tr>
<tr>
<td>Examine the reliability of a measuring instrument</td>
<td>Cohen’s kappa; Cronbach’s alpha; intraclass correlation; Kuder-Richardson 20; Weighted kappa</td>
</tr>
</tbody>
</table>


While the Version 1 content was being drafted, technical specifications for the iOS and web applications were also under development, and quotes were being sought from multiple application/web design companies. Several Australian companies were ruled out because they tended to bundle technical work with an extensive suite of additional services that were not required (e.g., intensive workshopping and conceptual development), provided quotes that far exceeded the available budget (e.g., $100,000+) or sought excessive fees simply to prepare proposals and quotes (e.g., $3,000+). Overseas companies based in emerging markets were able to provide more competitive quotes; however, communication difficulties presented an unacceptable level of risk. Finally, Mortaza Rezae, a final-year computer science student at Curtin University, was recommended to the project team. At the time, Mortaza was a volunteer iOS application developer for another research project in the School of Psychology and Speech Pathology at Curtin University, and he came with the...
recommendation of one of his academic supervisors, who had a clear understanding of Mortaza's abilities and the demands of the StatHand project. Mortaza assembled a team with two colleagues, Xavier Begue and Ivan Dwiputera (also final-year computer science students at the time), and they collectively prepared a detailed proposal and quote for the development of the StatHand iOS and web applications. Based on their proposal and quote, Mortaza, Xavier and Ivan were selected as the technical development team for the StatHand project.

The team primarily employed agile software development methodology for the technical development of the StatHand iOS and web applications. Agile is a project management framework that facilitates the management and control of iterative and incremental projects (McLaughlin, 2016). In complement to the agile methodology, the software development lifecycle framework was also utilised (Davis, Bersoff & Comer, 1988). The iOS apps were exclusively implemented using Xcode and Objective-C. This is the de facto standard for developing iOS applications. The web application was developed using Visual Studio and the ASP.NET framework. Although the team members were experienced with these tools and frameworks, they sought the technical assistance of their mentor, Dr Aneesh Krishna, at Curtin University when encountering certain problems. Dr Krishna helped the team to consider these problems from different perspectives, which allowed them to architect new and efficient solutions and to successfully deliver Version 1 of StatHand on time and within budget.

The result of the content and technical development described above is shown in Figure 1. When the StatHand application is launched (Screen 1), the user is presented with the first of several annotated questions: What do you want to do? There are five options available for this first question: (1) describe a sample, (2) compare samples, (3) analyse relationships or associations between variables, (4) examine the underlying structure of a measure, and (5) examine the reliability of a measuring instrument. Let’s imagine we are planning a simple study to examine whether caffeine affects response time. Response time data will be collected for two groups of adults who will drink either coffee or water immediately before testing. The most appropriate option on Screen 1 would be the second option, ‘compare samples’, because we wish to compare the performance of the coffee drinkers with that of the water drinkers. After making our selection for the first annotated question, we are presented with a second question, which asks us to identify the number of dependent variables in the study (Screen 2). A user uncertain about what is meant by ‘dependent variables’ can consult the brief explanatory text below the question; more-experienced users can simply make their selection. We indicate that we have ‘one’ dependent variable (Screen 2), which is measured on an ‘interval or ratio’ scale (Screen 3). Next, we are required to consider the number and nature of our independent variable(s). As shown in Screens 4 and 6 in Figure 1, each option can be expanded for context-specific definitions and examples by tapping the relevant Information buttons. Finally, we are asked to indicate whether we have any control variables (Screen 7), which, in the current example, we do not. Having now engaged with each relevant structural feature of our research scenario, we are presented with an appropriate analytic choice (Screen 8). In this case, an independent samples t-test.
Figure 1. An illustrative path through Version 1 of the StatHand iOS application on an iPhone 6. Screens 1 to 7 depict the decision points that a user would encounter when determining an appropriate statistical test for comparing two independent samples on a continuous dependent measure. Screen 8 depicts the recommended test based on the sequence of decisions made by the user. Reproduced from ‘Introducing StatHand: A Cross-Platform Mobile Application to Support Students’ Statistical Decision Making’ by P. J. Allen, L. D. Roberts, F. D. Baughman, N. J. Loxton, D. Van Rooy, A. J. Rock, and J. Finlay, 2016, Frontiers in Psychology, 7, Article 288, p. 6.

At any point during the decision making, a user can review their previous choices by using the History tool, as shown in Screen 9 in Figure 2. This feature allows the user to retrace their steps and to draw stronger connections between their choices and the solutions they reach. Selecting any entry on the History screen returns the user to the corresponding decision point. Users can also navigate through StatHand using the Back button and Forward button, or jump directly to a statistic from the searchable Index (shown in Screen 10). Also shown in Screen 9 in Figure 2 is the Notes tool, with which the user can pin their annotations to specific pages within the application or retrieve notes they have added on.
other pages. Finally, tapping the Share icon in the toolbar at the bottom of the screen reveals options to print, email or save the annotated sequence of decisions that have led to the current page (including any notes associated with those decisions). These features work in similar ways in the web version of StatHand at www.stathand.net, which has been designed for compatibility with any device capable of running a modern web browser.

Figure 2. Screen 9 in Version 2 of the StatHand application in landscape mode on an iPad Air 2 and Screen 10 showing the Index in the StatHand web application, running in Microsoft Edge on a Surface Pro 3. In Screen 9, the sequence of decisions that has led to an independent samples t-test is shown in the History tool on the left. Also shown in Screen 9 is the Notes tool, which is accessed from any screen by tapping the icon in the upper right corner of the screen. Adapted from ‘Introducing StatHand: A Cross-Platform Mobile Application to Support Students’ Statistical Decision Making’ by P. J. Allen, L. D. Roberts, F. D. Baughman, N. J. Loxton, D. Van Rooy, A. J. Rock, and J. Finlay, 2016, Frontiers in Psychology, 7, Article 288, p. 6.

In Version 2 of StatHand, additional content was incorporated into the iOS and web applications. Developed and implemented in the applications in much the same way as the Version 1 content, the new content in Version 2 guides users through the computation, interpretation and reporting of each statistic that StatHand helps to identify (listed in Table 1). The Version 2 content provides advice on assumption testing and the calculation of effect sizes where appropriate, offers links to additional reputable online and offline information about each statistic and provides examples illustrating the use of each statistic in open-access (and typically Creative Commons licensed) peer-reviewed published research. A partial example of the Version 2 content is visible in Screen 9 (see Figure 2); the full example is available in Appendix D. In total, there are 65 pages of content similar to the content shown in Appendix D. Embedded within those 65 pages are 80 short explanatory
videos, generally developed according to evidence-based multimedia-learning-object design principles (e.g., Clark & Mayer, 2011) and hosted on YouTube.

Instructors’ resources

The package of instructors’ resources is available, on request, from www.stathand.net/home/ir/ to anyone who teaches research methods, statistics and related subjects at recognised higher education institutions. The package contains a brief rationale for the use of the application as a learning and teaching tool, instructions for using the application, tips for integrating StatHand into undergraduate research methods and statistics classes, and active learning activities that instructors can adapt for their teaching purposes. The active learning activities are based around scenarios that instructors can use to stimulate discussion and assess their students’ abilities to identify appropriate statistics under a variety of circumstances. Three examples are shown in Figure 3.

Scenario 1: A friend of yours is running an experiment that involves measuring the self-esteem of 40 children, and then randomising them into two conditions. The children in the experimental condition are then praised after displaying good behaviour, whereas the children in the control condition are not. After a period of time, the self-esteem of each child is measured for a second time. Your friend wants to know if any changes in self-esteem observed between the pre- and post-tests are influenced by the experimental manipulation. What statistical analysis would you recommend?

Scenario 2: You are interested in whether Star Trek fans or Star Wars fans are more likely to be married or not married. Which statistical analysis would you use?

Scenario 3: The AFL commission is interested in whether a random sample of West Coast Eagles supporters and a random sample of Dockers supporters differ in the average amount of money they spend per season on AFL merchandise. Which statistical analysis would you recommend?

Figure 3. Examples of scenarios included in the StatHand instructors’ resources package. These can be used to stimulate discussion and assess students’ abilities to identify appropriate statistics under a variety of circumstances.

Publications

Early in the project, qualitative data were collected to (a) validate the assumption that students, but not ‘experts’ (i.e., researchers and research methods instructors) find selecting appropriate statistics for varied circumstances difficult; (b) better understand how students and experts identify and choose between statistics; and (c) canvas students’ and experts’ thoughts about the format and features of an aid that could facilitate these processes. These data informed the development of StatHand. They are reported in detail in ‘Difficult Decisions: A Qualitative Exploration of the Statistical Decision Making Process from the
Perspectives of Psychology Students and Academics’ (Allen, Dorozenko & Roberts, 2016), published in Frontiers in Psychology (2016 Impact Factor = 2.32), which has been accessed more than 5,200 times (as at July 2017) and has a high attention score (top 25 per cent) by Altmetric. It can be accessed at the publisher’s website: www.journal.frontiersin.org/article/10.3389/fpsyg.2016.00188/full.

In summary, and as anticipated, the student participants found the process of identifying appropriate statistics challenging, and even those who had completed several courses in research methods struggled to articulate on more than a very superficial and intuitive level how they would identify appropriate statistics for common research scenarios. Although some students recognised that there is a systematic process of decision making, none could describe it clearly or completely. By way of contrast, the researchers and research methods instructors in the sample were able to describe a far more systematic, comprehensive, flexible and nuanced approach to statistical decision making, which begins early in the research process and considers multiple contextual factors. They were sensitive to the challenges that students experience when making statistical decisions, which they attributed partially to how research methods and statistics are commonly taught. When asked to consider the format and features of an aid that could facilitate the process of statistical decision making, both groups expressed a preference for an accessible, comprehensive and reputable resource that follows basic decision tree logics. The academics, in particular, felt this aid should function as a teaching tool that would engage the user with each choice-point in the decision making process, rather than simply provide an ‘answer’.

Following the release of Version 1 of StatHand, a second paper was prepared and published in Frontiers in Psychology. Titled ‘Introducing StatHand: A Cross-Platform Mobile Application to Support Students’ Statistical Decision Making’ (Allen et al., 2016) and available at http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00288/full, this paper gave an overview of the rationale behind the development of StatHand, described the features of the application and provided recommendations for its integration into the research methods curriculum. These recommendations, based on the unified theory of acceptance and use of technology (Venkatesh, Morris, Davis & Davis, 2003), included (a) demonstrating StatHand at the outset and throughout the course; (b) linking StatHand to existing teaching resources; (c) minimising competition from other sources of interaction while using StatHand; and (d) using StatHand regularly across the curriculum, in methods and non-methods classes. At the time of writing, this paper has been accessed more than 3,200 times and has been given a high attention score (top 25 per cent) by Altmetric. Along with Allen, Dorozenko and Roberts (2016), it has been recently republished in an eBook titled Research Methods Pedagogy: Engaging Psychology Students in Research Methods and Statistics (Roberts, 2016), which is available at http://journal.frontiersin.org/researchtopic/3622/research-methods-pedagogy-engaging-psychology-students-in-research-methods-and-statistics.
Chapter 3: Evaluation

StatHand has been subjected to two evaluation studies. The first, a qualitative evaluation, is reported in full in this chapter. The second, an experimental evaluation, is currently being prepared for publication. To avoid a duplicate publication of the findings, the experimental evaluation is only briefly summarised here.

Qualitative evaluation

The purpose of this evaluation study was to develop a qualitative understanding of psychology students’ and psychology instructors’ perceptions of the utility of StatHand. Its secondary purpose was to gather ideas that may inform the further development of the application. The data were collected, analysed and reported by Hannah Uren, PhD candidate and research assistant in the School of Psychology and Speech Pathology at Curtin University, under the supervision of Peter Allen and Lynne Roberts.

Method

This study was reviewed and approved by the Human Research Ethics Committee at Curtin University before commencement (Approval ID: RDHS-117-16).

Participants

Student participants: Twenty-five undergraduate psychology students (19 female) and two speech pathology students (both female), with a mean age of 22 (range 18–29), were recruited as participants in this study. Nine students were in their second year, 10 were in their third year, and eight were in their fourth year of study. Most were studying full-time, with only one student studying part-time. Twenty-two students were domestic, and five were international. Participants were recruited via posts on relevant unit learning management system sites and through short presentations preceding lectures. Focus groups ran for one hour, and participants were given $30 cash to compensate them for their time.

Instructor participants: Nine staff members from the School of Psychology and Speech Pathology at Curtin University were recruited via individual emails. They included tutors, unit coordinators and research supervisors. Six were PhD educated, and the remaining three were currently enrolled in a PhD program. Staff members were not compensated for their participation.

Materials and procedure

Student focus groups: After informed consent had been established, the lead facilitator demonstrated the use of both StatHand and a paper decision tree (from Allen et al., 2014), with reference to Scenario 3 (see Figure 3). Copies of StatHand (preloaded on an iPad) and the paper tree were then distributed to participants, who were encouraged to use them (either individually or in pairs) to identify appropriate statistics for another six scenarios (all featured in the StatHand instructors’ resources). During this time, the lead facilitator and support facilitator mingled with participants and provided advice where appropriate. The lead facilitator then guided group discussion around the relative merits and limitations of
each aid, and prompted participants to consider how each could be improved and the circumstances in which they could see themselves using StatHand in the future.

Instructor interviews: After informed consent had been established, the interviewer ascertained the interviewee’s level of previous experience with StatHand and provided a brief demonstration if necessary. A semi-structured interview followed, in which the interviewee was prompted to consider the relative merits and limitations of StatHand, its utility in the context of teaching and supervision, and possible directions for its ongoing development.

Data preparation and analysis

Student focus groups and instructor interviews were transcribed verbatim and then imported into NVivo 11 for analysis. Each transcript was read and re-read, and initial interpretations noted down. Data were then systematically coded, and similar codes were collated to form themes.

Findings from student focus groups

In this section, the students’ perspectives on the strengths of StatHand are explored, along with their suggestions for the application’s future development.

Strengths of StatHand

Improved confidence in decision making: Participants described feeling more self-confident when using StatHand than when using a paper decision tree. This was attributed to the application’s step-by-step navigation, its clarity and ease of use, and its provision of examples and definitions.

Step-by-step navigation: Student participants described becoming easily lost and confused when using a paper decision tree to identify an appropriate statistic. Participants from all three year groups reflected on how the ‘step-by-step’ structure of StatHand forced them to think about each aspect of their research design. Being unable to ‘skip over’ a step in StatHand when it was perceived as ‘too difficult’ led to increased confidence in statistic selection. As one participant commented:

I find it a lot easier to go step-by-step instead of having to look at all the different [statistical tests] all at once and go from there. You don’t really second guess yourself with StatHand. (second-year student)

Student participants explained that the amount of information presented in a paper decision tree or text book often leads them to feel overwhelmed: ‘there is just so much going on ... whereas on the app, it’s based on what you actually choose’ (third-year student); ‘I don’t really like it [the paper decision tree] .... I look at it and I’m like, “Oh god, that looks really confusing” ’ (fourth-year student).

In comparison, the step-by-step design of StatHand was comforting and reassuring. ‘I think [StatHand is] not overwhelming. With the book, sometimes it’s just so much to look at. Whereas with [StatHand], it’s like, “I’ll hold your hand! ” ’ (third-year student).
Suggestions made by StatHand: Students reported as confidence boosting the way in which StatHand offers suggestions when the combination of research design features selected do not lead to a viable statistic: ‘I think that’s pretty handy because it tells you you’re wrong somewhere. Whereas if you’re using the [paper] decision tree, you don’t know if you’re right or wrong’ (third-year student). Another student commented:

I like that [StatHand] tells you if you’ve done something wrong. It suggests going to this, or looking at…. This one’s telling me to re-look at my criterion variable. It just tells you, and you can go back and do it properly. Whereas that one [the paper decision tree], if you get stuck, you really don’t know. You could be wrong or right. (fourth-year student)

Cross check with examples: Confidence appeared to be further boosted by the ability to check one’s own study design against a typical ‘example’: ‘the example at the end allows you to put it in context’ (third-year student).

Student participants suggested that StatHand could be improved with multiple examples for each type of statistical test. In addition, they indicated they would like the ability to type in each variable of their study design as they worked through StatHand and to then have a table at the end where they could compare their design characteristics with those of an example.

You’re given a scenario, for example, then you’ve worked through this, but it asks you to put in your variables, and put in your covariates. You get that reassurance knowing that if it’s asked you to put two in, you can put your two in. It can reaffirm that you’re more right. It makes you physically think about what it is, as opposed to thinking, ‘That’s maybe two, but I’m not really sure. I’m just going to press two’. (fourth-year student)

Accessibility and ease of use: The attendees at the student focus group saw StatHand as simple to understand and easy to use. They considered StatHand especially useful for beginners because of the application’s clear and simple interface. For beginners, the application was considered faster than using a book, because definitions for unfamiliar terms could be found quickly by tapping the relevant Information buttons. However, for those with more experience, using the paper decision tree was considered faster. There was some concern that the Index, Information and History tools were not obvious and could be overlooked. Even those participants who had previously used StatHand discovered they were not aware of its full functionality. Student participants said they would like to see a ‘How to Use StatHand’ tutorial when the application first opens, although they admitted they would probably not use it. Student participants also saw the application format of StatHand as a key strength. A digital resource reduced the need to carry textbooks. StatHand was also considered both faster and easier to navigate than a textbook. One student noted, ‘Everything’s at your fingertips. You don’t have to go and rummage through chapter notes or lecture slides or a secondary source’ (fourth-year student). StatHand was also considered more equitable than books as it is free to download, and most students have ready access to computers and mobile devices, both at university and at home.
Easy to access definitions: Even upper-year students said they struggled to remember basic statistical concepts (e.g., the differences between nominal, ordinal, interval and ratio measurement scales; the differences between independent and dependent variables), and they liked being able to use StatHand to find definitions of terms via the Information buttons. For example, ‘It’s a quicker way of getting to the point. If I looked at the decision tree, I’d be stumped on the first one, because I wouldn’t know whether it was scale or nominal [data]’ (fourth-year student). But these buttons were not immediately obvious to all, as illustrated by one participant, who said, ‘I didn’t even think to click on those info buttons. That made it so much easier’ (fourth-year student).

Ability to work forwards and backwards: Student participants explained that a key strength of the paper decision tree was the ability to easily work both forwards (from left to right) and backwards (from right to left). In particular, this allowed students who felt they had an idea about which statistical test was appropriate for a research design to check the accuracy of their intuition quickly. StatHand has similar utility available via the Index and History tools, yet many student participants who had used StatHand previously were unaware of this. When participants did learn about the utility of the Index and History, they explained that it was not intuitive that StatHand could be used in reverse. This suggests that a short tutorial might be needed to explain the full functionality of StatHand to students. Despite the Index and History tools not being intuitive for some students, they liked them and saw value in the ability to save a decision thread and send it to themselves or other students. Students said they could see themselves sharing StatHand output as part of exam preparation or group assignments.

Improvements suggested by students

More visual learning aids: Student participants saw potential for increased aids for visual learners. A copy of the full network of pages within StatHand (partially shown in Appendix B) was considered desirable by some students because it would allow them to see the structural relationships between statistical tests and to visualise each statistic in the context of other similar (yet distinct) possibilities. Another suggestion was to incorporate the SPSS symbols for nominal, ordinal and scale data into the StatHand interface.

Increased interactivity: Student participants suggested that StatHand had the potential to evolve from ‘a calculator to help you find which test to use’ into a ‘comprehensive learning tool’ (third-year student). Suggestions included quizzes for exam preparation and links to YouTube videos, book chapters and useful websites. One student suggested that StatHand could be turned into a game: ‘I could see a game whereby it popped up with a test and you had to match which variables and stuff went into it. Good for playing on the bus’ (second-year student).

Including alternative terminology: There were a few instances of students being confused by terminology. These mostly occurred when a term used in StatHand differed from that used in SPSS or textbooks. For example, ‘association’ versus ‘relationships’, ‘interval or ratio data’ versus ‘scale data’, and ‘Pearson’s product-moment correlation coefficient’ versus ‘Pearson’s r’. Overall, students explained that they would like language to be consistent across resources or for alternative terms to be provided, when applicable, within the StatHand application.
The next steps: Students wanted StatHand to provide information about ‘what to do next’ once they had reached an appropriate statistic via the application. Examples of such information included an illustration of what the dataset would look like for each statistic, advice about how to calculate the statistic and test its assumptions (where applicable) using SPSS, and information about reporting the statistic in assignments and reports. One student suggested interactive SPSS output tables, where interpretive details could be accessed by tapping elements. Other suggestions for resources for inclusion in StatHand included APA formatting rules, qualitative methodologies and, as mentioned previously, a StatHand instructional tutorial when opening the application for the first time.

Summary of focus group findings

Overall, students preferred StatHand over the paper decision tree. The student participants liked StatHand’s interactivity and accessibility, and they considered it faster and more reliable than a paper decision tree, especially for students unfamiliar with key statistical terms. The students said they would recommend StatHand to others but felt that the application would be of greater value if it also provided advice on the computation, interpretation and reporting of the statistics it helped to identify. Some students reported that not all StatHand features were intuitively obvious, and offered various suggestions for the future development of the application. Since these data were collected, Version 2 of StatHand has been developed. As described in Chapter 2 of this report, Version 2 incorporates many of the student participants’ suggestions.

Findings from instructor interviews

In this section, the instructors’ perspectives on the strengths of StatHand are explored, along with their suggestions for its future development.

Strengths of StatHand

An active learning tool: From the perspective of the instructors, a key strength of StatHand was that students would be forced to ‘slow down’ and consider all the characteristics of their research designs, thus avoiding the temptation to jump to conclusions. Instructors liked how StatHand offered students ‘edible chunks of information’, which the instructors believed would reduce students’ propensity to feel overwhelmed when faced with statistical decision making. Instructors described a tendency for students to select statistical tests based on familiarity rather than on appropriateness. It was suggested that StatHand would remove this tendency by forcing students to work through each aspect of their design. Instructors also assumed that, when using StatHand, students would be more likely to remember the steps associated with choosing an appropriate statistical test, because the application forces the user to engage actively with the full decision making process.

I think having it in this format would be more intuitive or engaging for students. I think I also like how... I don’t know how to explain it, but because you’re actively interacting with the app to generate the decisions, I think it’s more engaging, of course, but I think it maybe is more rewarding when you get the result.
This was seen as a strength because students would be more likely to retain the information learnt through StatHand than information learnt through the paper decision tree.

_Students would try and race to the end of the decision tree and be like, ‘Oh, I know that test so it must be this’. Whereas this way [using StatHand] ... it explains everything and every step._

**Provision of examples:** Instructors also appreciated the research design examples provided alongside the suggested statistics. Instructors reported that students often repeat names of familiar statistical tests without completely understanding what the tests entail. An instructor suggested that the examples in StatHand may provide deeper learning opportunities than those available when using a paper decision tree.

_I like the fact that it [StatHand] gives the example, because a lot of the time you can just regurgitate the name but the student doesn’t really understand what a working example of it would be._

**Student self-efficacy:** Instructors believed that using StatHand would give students increased confidence in their ability to make statistical decisions because the application would provide students with greater autonomy. Instructors liked the interactivity of the application, which they felt would guide two particularly challenging groups of students: those who are disinterested and those who are overwhelmed by research methods. As one instructor said:

_I think there’s a sense that you’re collaborating with this tool to come up with that answer, rather than just consuming information. I think there’s also a level of certainty in putting the information in and then getting the stuff out, whereas if you’re interpreting a decision tree, how do you know you haven’t read the tree wrong?_  

Instructors saw StatHand as a way to build students’ confidence and empower them to make their own statistical decisions.

**Equitable:** Staff members liked that StatHand was accessible to all students on a variety of digital platforms. As StatHand is a freely available resource, all students would have equal access, which is not necessarily the case with textbooks. Furthermore, instructors noted that StatHand would lessen the need for students to carry heavy books, and exercises in statistical decision making could be run easily in supervision meetings (e.g., with honours students) without the need for physical resources.

_I think the app idea is really good, because these days most people are sitting there with their phone in front of them._

**Improvements suggested by instructors**

**Links to resources:** Instructors recommended the inclusion of links to websites, videos and books. Easy to absorb resources, such as videos and graphics, were particularly recommended and seen as preferable to textbooks, which can be overly rich in detail. Instructors also wanted to see more than one example included for each statistic. One
suggestion was to have one theme throughout the application, showing how one research topic or question could be addressed via a variety of research designs and statistics.

**Embedded within a variety of psychology courses:** Instructors saw it as essential to the uptake and use of StatHand that its use be promoted to students from the first day of the psychology degree and throughout various units, not only those focused on research methods and statistics. Instructors recommended that the full functionality of the application be explained, and suggested that a brief video tutorial appear upon opening the application for the first time. They also noted that it would be important to set aside some time to fully explain the use of the application to students, as some useful functionality may not be intuitively obvious to all students.

*They can be aware it’s a resource; but I think they really have to have practice and experience with it to see the value of using it. I would embed it into my teaching practices if I wanted students to use it.*

**A one-stop shop:** Instructors reported they would like StatHand to become a ‘one-stop shop’ for statistical needs, not only for undergraduate students but also for postgraduate students and staff. They noted that StatHand answers the ‘what’, but not the ‘how’. Instructors said they would like StatHand to include step-by-step instructions on how to calculate statistics using SPSS, and power and effect size calculators. Instructors also wanted to see qualitative designs and advanced quantitative analyses incorporated into StatHand but understood the challenges associated with achieving this.

**More graphics:** As with the student participants, instructors reported value in incorporating figures and graphics into StatHand where possible and appropriate.

*There’s still students out there, when, if they were to read ‘to examine the association between two variables’, some students still would struggle and go, ‘What does that mean?’ So, if there was some sort of way of providing a small graph or diagram next to that...*

Staff suggested that graphical representations of factor analysis, in particular, would be useful for students. Another suggestion was to have a decision tree on the screen so that as the StatHand user moves through each step, the relevant path on the decision tree is highlighted.

**Uses for StatHand**

Instructors perceived a range of uses for StatHand. They suggested that StatHand could be used not only in research methods units but also in a variety of undergraduate psychology and health science units. When critiquing literature, students could use StatHand to assess whether authors had used the most appropriate statistics. StatHand was also considered an important tool for student self-study for assignments and exams. In addition, it could be used in fourth-year supervision to encourage students to suggest which statistics would be appropriate for their research design, rather than relying on their supervisors’ advice. The instructors in the sample believed that StatHand would allow students to feel ownership of their project by providing them with the tools to make informed statistical decisions, rather
than writing ‘because my supervisor told me’ in their research proposals. Finally, staff considered StatHand a useful refresher tool for those in their final years of study.

**Summary of instructor interviews**

The instructors considered StatHand an aesthetically pleasing, clean, simple and user-friendly application that was of benefit to students because it encouraged student self-efficacy and active learning. They hoped to see the application incorporated into a variety of undergraduate psychology units and suggested numerous expansions for the application, such as instructions for running the statistical tests and more graphics. As previously noted, many of these suggestions have been incorporated into Version 2 of StatHand, which was developed following the collection of these data.

**Experimental evaluation**

As noted previously, the results of the experimental evaluation of StatHand, which was conducted by James Finlay, honours student in the School of Psychology and Speech Pathology at Curtin University under the supervision of Peter Allen, are being prepared for publication, and are summarised only briefly in this report.

Over 200 students were randomised to one of four aids for statistical decision making (StatHand, a paper decision tree, a familiar textbook, and a familiar textbook plus the paper decision tree) and asked to identify a suitable statistic for each of five research scenarios. Results suggested that, after controlling for prior performance in research methods and statistics, StatHand users demonstrated higher performance accuracy than the other three groups. StatHand users also reported significantly lower cognitive load, higher confidence and greater satisfaction than at least one other group. However, it took StatHand users longer to complete the task. StatHand can be considered relatively instructionally efficient when compared with the other three aids.
Chapter 4: Impact and dissemination

Impact

Although a formal impact plan was not a required part of the initial StatHand proposal, in this section the project is briefly considered in terms of the Impact Management Planning and Evaluation Ladder framework (Hinton, 2014).

First, the project has been successful in advancing the project team members’ understanding of the challenges faced by students when they are required to select statistics appropriate to their research questions and designs. This understanding has informed their teaching and supervision practices. The publications and presentations resulting from the project have aided project team members’ career progression and increased their networks and esteem within the Scholarship of Teaching and Learning community. The project also provided the first major contract for the technical development team. The lead developer, Mortaza Rezae recently commented that ‘it is without a doubt that the team has acquired extensive skills and knowledge after completing this project …. these skills have allowed all of the [technical development] team members to advance in their respective fields’. The project has also provided honours students with the opportunity to co-author papers and presentations, thus increasing their prospects of securing postgraduate scholarships in the future.

Second, project team members’ immediate students have benefited from the project, as StatHand has been embedded into the undergraduate psychology curricula at four Australian universities. As noted in the previous chapter, the application appeals to students (who have self-reported on its perceived utility) and has been found to promote greater accuracy in decision making relative to other decision making aids.

Third, spreading the word has occurred via the dissemination activities described below. The first two publications associated with the project have been accessed over 8,400 times. Both have been given high attention scores by Altmetric, and have been shared across multiple social networks. The iPad and iPhone iOS applications have been downloaded over 20,000 times, and there have been 22,000 visits (and 73,000 ‘hits’) to the StatHand website in the first six months of 2017.

Narrow opportunistic adoption is evident in the instructor interviews cited in Chapter 3, as well as in anecdotal reports of colleagues recommending StatHand to students and making use of the application within their teaching and supervision. Broad opportunistic adoption is suggested by the access and download figures reported above, although this cannot be confirmed.

To date, there is little evidence of narrow or broad systemic adoption; however, dissemination is ongoing, and Version 2 of StatHand has only recently been published.
Dissemination

StatHand has been the subject of two journal articles (with a third forthcoming) and 12 conference presentations. Further details are provided below. Furthermore, opportunistic dissemination has occurred via training workshops, networking with colleagues, research group meetings, Twitter postings (and re-tweets), and the distribution of postcards, magnets and business cards (see examples in Appendix E).

Publications


A third paper, describing an experimental evaluation of StatHand, is currently in preparation.

Presentations

Presentations are listed in chronological order. Note that StatHand was referred to as ‘StatTree’ in some of the earlier presentations.


References


Appendix A

Certification by Deputy Vice-Chancellor (or equivalent)

I certify that all parts of the final report for this OLT grant/fellowship (remove as appropriate) provide an accurate representation of the implementation, impact and findings of the project, and that the report is of publishable quality.

Name: .................................................. Date: 13/03/17
Appendix B

Draft schematic representation of the pages (decision points) contained within the ‘compare samples’ branch of the StatHand application
Appendix C

The explanatory text that accompanies the options available at the first decision point in StatHand, ‘What do you want to do?’ (indicated with an asterisk in Appendix B). Note that in the application itself, the explanatory text immediately below the title question is always visible, whereas the text associated with each option can be accessed ‘on demand’ with either a tap or mouse click.

**What do you want to do?**
Choose the option that best matches your objectives. For further information about an option, tap the icon beside it.

- Describe a sample.
A sample is a group of cases (humans, animals, objects etc.) selected from a population of interest.
A sample can be described statistically using measures of central tendency (e.g., mean, median and mode) and dispersion (e.g., standard deviation and range), or graphically (e.g., with a bar graph, histogram etc.).

- Compare samples.
A sample is a group of cases (humans, animals, objects etc.) selected from a population of interest.
Using different statistics we can (a) compare one sample to another (e.g., comparing the average amount of alcohol consumed by a sample of male students with the average amount consumed by a sample of female students); (b) compare a sample to itself, under a different set of conditions (e.g., comparing the average reaction times of a sample of students before and after drinking alcohol); or (c) compare a sample to a known population characteristic or other predetermined value (e.g., comparing the average amount of alcohol consumed by a sample of students with the amount recommended by a national health authority).

- Analyse relationships or associations between variables.
In research, a variable is any characteristic that can be measured or manipulated and that can take on at least two values.
A range of different statistics are available for analysing relationships or associations between variables. In choosing an appropriate statistic, the purpose of the analysis as well as both the number and nature of variables being analysed need to be taken into consideration. In all instances, all cases (humans, animals, objects etc.) require a score on every variable.

- Examine the underlying structure of a measure.
The measuring instruments developed and used in research often comprise multiple items (i.e., statements or questions), which are assumed to capture aspects of a smaller number of underlying characteristics. Techniques for examining the structure of these measures are covered in this section of StatHand.

- Examine the reliability of a measuring instrument.
Reliability refers to the consistency or dependability of a measuring instrument over time, items or observers. Different statistics are suitable for these different types of reliability.
Appendix D

Example of the content integrated into Version 2 of StatHand

RESULT: Independent samples t-test
Tests for a statistically significant difference between two independent sample means.

For example, is there a difference between the average weekly alcohol consumption levels of male and female engineering students? In this example, the IV (gender) has two levels (male and female), while the DV (weekly alcohol consumption) is a ratio variable.

Assumptions
Before running and interpreting an independent samples t-test, there are a number of assumptions that should be met. Some are a matter of research design and addressed before and during data collection. Others can be assessed statistically after data collection.

Design:

Scale of measurement: The dependent variable should be measured on a continuous (i.e., interval or ratio) scale. Examples of suitable DVs for a t-test include psychological measures of intelligence or personality and physical characteristics like age, height or reaction time.

Independence of observations: There should be just one score for each case. Cases should be independent (i.e., unable to influence each other).

Analysis:

Normality: Each group of scores should come from a normally distributed population. We can argue that this assumption has been met if each group of scores is approximately normally distributed. It can be assessed graphically or statistically. In the video below, the following methods are illustrated: skewness and kurtosis, the Shapiro-Wilk test, histograms, stem-and-leaf plots, box plots, normal Q-Q plots and detrended normal Q-Q plots. Although it is not necessary to examine all of these statistics and graphs, it is recommended that you check at least one of each. Remember, before running a t-test, you need to check that each group of scores is normally distributed.

The t-test is said to be ‘robust’ against moderate violations of the normality assumption, provided you have a reasonable number of cases and your group sizes are relatively equal. When faced with severe violations, consider running a Mann-Whitney U test instead.

Homogeneity of variance: Both groups of scores should come from populations with the same variance. We can argue that this assumption has been met if both groups of scores have roughly equivalent variances. This can be assessed using Levene’s test, which is produced alongside the t-test in SPSS. When Levene’s test is statistically significant, the assumption has been violated. If homogeneity of variance cannot be assumed, an alternative version of the t-test (usually called a t-test assuming unequal variances, or Welch’s t-test) can be used instead. SPSS produces both versions of the test by default.

Run and Interpret
Allen, Bennett and Heritage (2014) describe a study in which a teacher wants to determine whether or not there is a difference between the mean IQ levels of her male and female students. The videos below illustrate how to answer this question using SPSS.

Continued overleaf
Quick guide:
- Select Analyze, Compare Means, and then Independent-Samples T Test.
- Move the independent variable into the Grouping Variable field, and click Define Groups.
- Enter the values used to represent each level of the independent variable in the Define Groups sub-dialogue, then click Continue.
- Move the dependent variable into the Test Variable(s) field.
- Click OK.

Effect Size
Cohen's $d$ is a measure of the difference between two group means, expressed in standard deviations. It can be calculated here: http://www.psy.edu.hk/mm/effectsizefaqs/calculator/calculator.html

At this website, you can also calculate:
- Glass’s delta (recommended if you have unequal variances).
- Hedges’ g (recommended if you have unequal group sizes).

As very general rules-of-thumb, which should not be followed too literally, $d = .2$ can be characterised as ‘small’, whereas $d = .5$ and $d = .8$ can be characterised as ‘medium’ and ‘large’ respectively.

Further Resources
Textbook resources:
Further information about the independent samples t-test can be found in virtually any introductory research methods or statistics textbook. Below are a few examples:

Online resources:
GraphPad t-test Calculator: http://graphpad.com/quickcalc/ttest1/. Select Unpaired t-test if you’ve met the homogeneity of variances assumption. Otherwise, select Welch’s unpaired t-test.
Vassarstats Concepts and Applications of Inferential Statistics: http://vassarstats.net/textbook/ch11pt1.html. Chapter 11 of this popular online statistics textbook outlines the purpose and logic of the independent samples t-test, as well as its calculation and interpretation.

In the published literature:
Garratt-Reed, Roberts and Heritage (2016) compared students enrolled in a traditional face-to-face version of an introductory psychology course with students enrolled in an online version of that same course on a range of different dependent variables. For a number of these comparisons, they used independent samples t-tests. For example, they state: ‘students in the face-to-face unit ($M = 24.43, SD = 2.84$) scored significantly higher than students in the online unit ($M = 21.49, SD = 3.26$) on the group presentation assessment, $t(792) = 6.49, p < .001, 95\% BCa CI [2.05, 3.83]$. For your reference, BCa stands for ‘bias corrected and accelerated’ and is one of several methods available for calculating confidence intervals.
Appendix E

Examples of promotional materials

**Figure E3:** Business card (front).

**Figure E2:** Postcard (back).

**Figure E5:** Magnet (front).

**Figure E4:** Business card (back).